DOCUMENT RESUME

ED 442 433 HE 033 031

AUTHOR	Chen, Sheying; Cheng, David X.
TITLE	Remedial Education and Grading: A Case Study Approach to Two Critical Issues in American Higher Education. A Research Report Submitted to the Research Foundation of the City University of New York.
SPONS AGENCY	City Coll. Research Foundation, New York, NY.
PUB DATE	1999-06-00
NOTE	59p.; Papers based on the studies were presented at the Annual Forum of the Association for Institutional Research (39th, Seattle, WA, May 30-June 2, 1999) and at the Anniversary Conference of North East Association for Institutional Research (25th, Philadelphia, PA, November 14-17, 1998).
CONTRACT	PSC-CUNY-RG-669282
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Academic Achievement; *Academic Persistence; College Students; Compensatory Education; *Dropout Research; *Educationally Disadvantaged; *English (Second Language); *Grade Inflation; High Risk Students; Higher Education; Language Skills; Non English Speaking; Remedial Mathematics; *Remedial Programs; Remedial Reading; School Holding Power; Student Attrition; Time to Degree
IDENTIFIERS	*City University of New York

ABSTRACT

This study, which focuses on one college of the City University of New York (CUNY), is in two parts. The first part is a paper titled "Open Admissions and CUNY in Crisis: A Comparison of Remedial and Non-Remedial Students"; the second paper is titled "Factors Affecting Grading Practices." The first study examined the difference between students who received remediation and those who did not need it, and assessed the differential impact of open and selective admissions. Sixteen data tables and 7 figures are included. The study found that students' native language played a major role in determining the need for remediation; that remedial students narrowed the academic performance gap as compared to nonremedial students during the normal 4-year period of college study; and that students with a greater need for remediation required a longer period of retention. The second paper discussed practical concerns of grade inflation. Findings revealed that adjunct faculty gave higher grades than full-time faculty; that faculty rank had only a marginal effect on grading; that grades in the humanities and social sciences were higher than in science and technology; and that the higher the course level, the higher the average grade. (Contains 56 references and 20 notes.) (CH)

REMEDIAL EDUCATION AND GRADING

A Case Study Approach to Two Critical Issues in American Higher Education

Sheying Chen, Ph.D.
Principal Investigator

David X. Cheng, Ed.D.
Co-Principal Investigator

A Research Report
Submitted to
The Research Foundation of
The City University of New York

PSC-CUNY Research Grant No. 669282

June 1999

Table of Contents

List of Tables

List of Figures

Preface

The last few years of the 20^{th} century have been witnessing renewed and intense debates about American higher education over a number of issues. Two topics have received serious attention and also produced widespread concern. One is remedial education or remediation, an uneasy position that the traditionally elite-oriented higher education has had to take to address the academic deficiencies of present-day students. The other is so-called grade inflation, a phenomenon that has been deemed as threatening the reputation of higher educational institutions. The two major issues have been combined to make many believe that the academic standards of American higher education are being compromised and undermined. All kinds of comments, criticisms, and arguments have been heard from various types of watch dogs, gate keepers, insiders, and outsiders. Politicians, joined by the media, have been pinpointing the problem and demanding a change. Employers and taxpayers, frustrated by the news and their own observations, have been wondering who should be held responsible for the problematic educational results. Educators and educational administrators, worried about the consequences, have been expressing their views to address the issue that is regarded not unique to any single institution.

What has not been accomplished, however, is sufficient evidence from empirical data that could have a final say in the arguments. It is fair to say that everyone seriously concerned about the issue has looked at some numbers related to his or her claim. Yet the discovery of the real facts and their implications demands more thorough and systematic analysis of the quantitative and qualitative information. Research, of course, takes more time and effort than what is needed for giving an opinion. But no opinion will be grounded unless and until the embedded facts are ferreted out. In this regard, credit should be given to the studies that have been done so far related to the two most important issues, particularly those that have made the City University of New York (CUNY) a significant case. CUNY has been singled out as a focal point of the debate for a number of known and unknown reasons; chief among them is probably its open admissions policy, which has become a target amid popular accusations of the problems related to remedial education and probably also "grade inflation." From a historical perspective, this situation can be very confusing as to why the University had the mandate of open admissions in the first place. We need to find out, from detailed case data, what purposes the open admissions policy has actually served, what outcome remedial education has really had, and what facts have been potentially related to grading. For both the pros and cons, this is the only basis for a meaningful conversation and discussion.

We are grateful to the Research Foundation of the City University of New York for recognizing the research need and placing confidence in us to carry out the project with a PSC-CUNY award. This report documents what we have done so far on the two different but related research topics, i.e., remediation and grading. In conducting the research, we carefully watched for our potential bias as educators, for instance, favoring (maybe) an elitist orientation or denying factors affecting grading other than objectivity. Despite that the results are quite shocking as they may have gone against some of our first impressions, as social scientists we embrace the findings with appreciation of the facts and the empirical testing they represent. As of this writing we have also had a chance to read the report of the New York City Mayor's Advisory Task Force on the City University of New York released on June 7, 1999. Although it seems just in time and in response to the Task Force's call for "objective measures" of remediation efforts, our report is part of a systematic pursuit of the understanding of underlying issues. We expect that the Task Force's conclusions will lead to certain changes while initiating a new round of debates. What caught many people's attention, however, was its recognition of the "critical importance" of this institution to New York and the nation, and "potentially a model of excellence and educational opportunity to public universities throughout the world" (Schmidt et al., 1999, p.5). Particularly, "Given the large scale and variety of its remediation efforts, CUNY ought to be the world's leading repository of knowledge" about: the cognitive needs of different types of remediation students; which instructional methods are most effective; which professors with what kind of training are most effective; and which institutions are best able to focus their energy and skills on remediation programs that work (ibid., p.31). These are exactly what have been on our research agenda, and the findings contained herein constitute our first step toward making more useful information about outcomes available and creating more reliable and valid measures as to what works in which situation.

While the accuracy and the impact of the Task Force's report are yet to be carefully assessed by professionals and the public, we are heartened by the Task Force's following comments: "CUNY's historic mission — to provide broad access to a range of higher education opportunities of quality suited to New York City's diverse population and to the City's needs — will be more important in the 21st century than every before" (ibid.). We would be more than gratified, therefore, if our case study and ongoing effort could contribute to "a concerted, long-term strategy to make CUNY the preeminent urban public university in the world" (ibid.). We hope you, the reader, will also join this effort and find our results fascinating and utilize them in future research design and policy dialog. We would appreciate your feedback to help us further inquire into the issues involved as American higher education enters the new millennium. Please send your comments and address correspondence to:

Sheying Chen, Ph.D.
Dept of Psychology, Sociology, Anthropology, and Social Work
CSI/CUNY 4S-223, 2800 Victory Blvd.
Staten Island, NY 10314-6600
E-mail: chen@postbox.csi.cuny.edu
Tel.: (718) 982-3766 Fax: (718) 982-3794

or

David X. Cheng, Ed.D.
Director of Institutional Research
CSI/CUNY 1A-304, 2800 Victory Blvd.
Staten Island, NY 10314-6600
E-mail: cheng@postbox.csi.cuny.edu
Tel.: (718) 982-2085 Fax: (718) 982-2578

Executive Summary

Focusing on one college of the City University of New York (CUNY) as a case study, this project contained two parts. The first part compared the students who received remedial education with those who did not in terms of their on-going performance. Potential factors contributing to the need for remediation and those affecting student retention and graduation were also explored. Using a longitudinal data set that covered six years, the project tracked a panel of 1,334 students who belonged to the cohort of Fall 1992 freshmen classes. The purpose was to illuminate the complicated implications of remediation, rather than a direct and straightforward program evaluation, in a particular context of the alleged crisis of CUNY resulted from its open admissions policy. The main findings of the study were as follows:

(1) Factors effecting students' need for remediation at entry: Students' native language played a major role in determining their need for remediation, that is, ESL (English as a second language) students had a greater need for remediation (except for math) than non-ESL students. This was on the contrary to the suspicion that the college studied was an exception to the impact of English as a second language (Lore & Murtha, 1997; Volpe, 1997). In addition, those students who were household heads had a greater need for remediation in math. ESL students possessed better math skills, while working full-time and older age were associated with better reading and writing skills.

(2) Difference between remedial and non-remedial students in their performance: In terms of ongoing academic achievement, the findings spoke positively for remedial students. Taken as a time series, the data clearly showed a trend for them to narrow their gap with non-remedial students in a normal period of college study (approximately 4 years) as indicated by their average GPA's. In fact, in their fifth year of study the difference was reversed, with the remedial group having a higher average GPA than the non-remedial group. Time series analysis further revealed that the "closing gap" trend was due to constant improvement of the remedial students on one hand and relatively unstable performance of the non-remedial students on the other. The present study was unable to determine the causes due to lack of means of control, although it was natural to assume that remediation had a positive effect. Whatever the reasons, the finding would lend some support to the open admissions policy. The chance for a remedial student to improve and catch up was great as long as he or she stayed on the path. It was amazing indeed to see from the results how dramatic improvement could be made even for those who flunked all three basic skills tests at entry.

(3) Comparison of students with different remediation needs and outcomes in terms of their retention and graduation patterns: First, we performed T-Test and ONEWAY analyses on potential group differences in the number of years staying out of any degree program ("stop-out"). Except for one variable (i.e., employment status) with somewhat mixed results, the findings were consistent and interestingly complementary to the above results from a longitudinal perspective. Specifically, data showed that a greater need for or a worse outcome of remediation resulted in a larger number of years staying out of any degree program. This finding suggests that if the institution wants to maintain its open admissions policy based on the above optimistic results, it should be prepared to allow the remedial students longer time before they can enter or return to study in degree programs (and thus possibly catch up with non-remedial students).

We then constructed two survival variables for examining the length of time between students' entry and two different end events, that is, retention/dropout and graduation. Consistent with the T-Test and ONEWAY analysis results on the "stop-out" variable, the findings clearly distinguished between the student groups with different remediation needs and outcomes. That is, a greater need for or a worse outcome of remediation resulted in a larger number of years needed to graduate, which also meant a longer period of retention. This might have an impact on graduation rates, though no calculation was feasible nor was any conclusion due to the censoring problem (particularly lack of transfer data). On the other hand, based on the promising performance and retention the remedial students have shown, the institutional effort should be placed on helping them move forward as quickly as possible and eventually complete their programs. Internal push through academic advisement etc. as well as external attraction with information on after-graduation job rewards may help to achieve this goal.

The second part of our study discussed the practical concern of grade inflation and argued that it is not a researchable question under the unstandardized condition. It instead explored the potential factors affecting faculty grading practice, which was considered a necessary basis for policy making and intervention (if ever deemed as desirable). Empirical data from the student information system at the case college were utilized, which contained 31,916 grade records. Findings revealed that (1) adjunct faculty gave higher grades than full-time faculty; (2) faculty rank had only marginal and mixed effects on grading; (3) grades were higher in the humanities and social sciences than in science and technology disciplines; and (4) the higher the course levels, the higher the average grades. Of all the variables examined, course level had the greatest impact while adjunct status ranked the next. Implications for policy intervention are discussed and methodological issues in grading research also indicated in the report.

Part I

Open Admissions and CUNY in Crisis:

A Comparison of Remedial and Non-Remedial Students[a]

[a] A paper based on this part of study has been presented at the 39th Annual Forum of the Association for Institutional Research (AIR), Seattle, May 30-June 2, 1999.

Background

The City University of New York (CUNY) is said to have entered a new era of crisis. The lasting debate over its open admissions policy adopted thirty years ago has now centered on its struggle with the need of entering classes for remedial education. In 1997, 87% of community college freshmen and 72% of senior college freshmen failed one or more of CUNY's remediation placement tests (math, reading and writing), and 55% of CUNY freshmen failed more than one (Schmidt et al., 1999). The fact that 80 percent of its 1996-97 freshman class of 1,800 did not pass all three basic skills placement tests has made one of CUNY's senior colleges, which has also offered associate degree programs, a focal point of the bashing and defending (e.g., Lore & Murtha, 1997; Editorial, 1997; Springer, 1997; Volpe, 1997). And the fact that the college had to schedule more than 70 remedial classes which accounted for 24 percent of the total freshman course work and cost over $1 million has raised widespread concern over the "drag" on the overall academics of the institution.

This is not just a local issue, however (Springer, 1997). Except for the 8 percent that are highly selective, colleges throughout this nation, private and public, are also struggling with the problem of teaching writing, reading, and mathematics to adults (Volpe, 1997). CUNY has been put at the spotlight of debate because of its open admissions policy. Proponents and opponents of that policy all point to some facts supportive of their arguments, though remediation is an issue CUNY and other universities have to deal with whether or not the institutions themselves should be responsible for the deficiencies. After rethinking open admissions and remediation, the New York City Mayor's Task Force "believes that remediation is still an appropriate and valuable endeavor for CUNY community colleges to undertake" (Schmidt et al., 1999, p.21). And for this reason, the Task Force members "salute CUNY's willingness to step into the breach for high school graduates whom the schools have failed, immigrants, and returning adults" (ibid.).[1]

The concern that open admissions may have lowered standards and CUNY failed to retain its students is at base a question about the outcome of remedial education. Although considerable resources have been spent on remediation in practice, existing research does not appear to have focused on this key issue. While previous studies have successfully proved the case that open admissions policy has helped thousands of students from working class families realize their dream of higher education (Lavin & Hyllegard, 1996), little is known about what a role remediation has played in making this happen. For both proponents and opponents of the open admissions policy, they may easily get confused as the discussion cries for more thorough empirical study and careful logical reasoning (Volpe, 1997). The fact that more than 40 percent of the open-admissions students never earned bachelor's degrees, for instance, may be taken as an indication of CUNY's failure in graduating its open-admissions students. Yet this may also serve as evidence of high standards and stiff requirements of CUNY's degree programs (Arenson, 1996).

It seems more pertinent to look at open-admissions and other students' performance on a comparative basis. If this is not feasible (the students may not be comparable, for example), then comparing remedial and non-remedial students might be a closer solution. The outcome of remedial education, therefore, has become the focal point in the debate. Two opinions are seen in sharp contrast to each other. One views remedial education as necessary for older students returning to school. "With a review course, the deficiency is eliminated. Mature and serious, they are ready for college level work" (Volpe, 1997). The other says, "The reality is that they drop out" (Badillo, cited from Arenson, 1996). Although researchers have made remarkable efforts in finding out the facts on retention and graduation including the influence of time and transfer (e.g., Lavin & Crook, 1990; Lavin & Hyllegard, 1996; Lavin et al., 1997; Retention Study Committee, n.d.), the outcome of remediation and possible intervention strategies need to be further explored.

In a sense, the issue of educational quality is firstly an issue of the outcome of remediation since it determines the preparedness of students who will eventually effect institutional quality standards. Especially, as CUNY has decided to limit remedial coursework for bachelor degree students and eventually phase it out in senior colleges, the issue needs to be studied more thoroughly. Since the issue also has national interdisciplinary significance, the participation of researchers from various social science disciplines is important to the accomplishment of the research tasks. There have been studies of the outcome of CUNY-wide Summer programs (Smodlaka, 1996). But the scope of the research needs to be expanded and the implications of the findings need to be further explored. Since each campus has its unique student body and institutional characteristics, research at the individual college level is especially needed.

Research Questions and Hypotheses

This study was intended to provide more adequate answers to the key issue through a detailed case study of the outcome of remedial education. Special emphasis was put on the difference between the students who received remediation and those who did not need it, which rendered an opportunity to assess the differential impact of open- and selective admissions when it was not feasible to make a direct comparison. Specific objectives were embodied in the following research questions:

(1) Who are the students who receive/do not receive college remediation? Or, what are the potential characteristics and factors contributing to the need for remedial education?

(2) What is the outcome of college remediation in terms of the rates of passing college basic skills tests after up to a year of remediation? How does college remediation impact on other aspects of academic performance?

(3) Are there any major differences between the students who have received remediation and those who have not in terms of retention and graduation patterns?

(4) What are the main factors, in addition to the need for and outcome of remediation, affecting student retention and graduation?

A review of relevant literature indicates that some of the common reasons for remediation, such as high percentages of immigrant students who know English as a second language, do not seem to have played a vital part in the case college (Lore & Murtha, 1997; Volpe, 1997). On the other hand, both literature and our observation suggest that factors such as older age, time interval at home or in the workplace before returning to school, and performance in high school might be associated with the need for remediation (ibid.). Our observation also suggests that family obligations might be another factor. Previous research points to employment, economic condition, and race/ethnicity as important facts related to retention patterns (Lavin & Hyllegard, 1996). A recent CUNY-wide study found that financial difficulties are the main reason students terminate or suspend their studies, supplemented by such institutional factors as deteriorating academic services (Gittell & Holdaway, 1996). Another recent study highlights the influence of time and transfer in affecting graduation rates (Lavin et al., 1997), though critics suspect that the students do not have the money to transfer (Arenson, 1996). Further looking into who are the leavers reveals that weak students are not the only ones to depart: a third of students in good academic standing also left CUNY (Lavin, Lerer, & Kovath, 1996). This finding spurs us to look for multiple factors instead of a sole determinant (i.e., unpreparedness at entry) in building an appropriate student attrition model. In addition, we intend to compare the performance of remedial students with that of non-remedial students since previous research has not provided enough factual information. Our hypotheses were:

(1) Factors effecting the need for remediation include students' performance in high school, aging (reflecting the impact of the length of interruption in schooling), language, employment status, and family obligations.

(2) Non-remedial students perform better academically than remedial students, although college remediation may have a positive impact on the achievement of the latter.

(3) Unpreparedness at entry is not the sole determinant of the patterns of student retention and graduation. Other factors include language, economic condition, employment, and full-time/part-time student status.

Data Sets and Analytic Strategies

Our research focused on one college as a case study. Empirical data were obtained from the campus-wide student information system. A working data set was constructed by extracting and combining data from different academic and administrative databases, which included 1,334 student cases. They belonged to the fall 1992 freshmen classes. The first part of the data was demographic and first term academic information. The second part was Fall to Fall enrollment information (degrees the students were pursuing and their GPA's) and degrees completed.

All the data were reported in aggregate forms with the confidentiality of the information on individual students being ensured. Data management and analysis were performed using SPSS for Windows V.8 and a few other computer programs. Data manipulation involved thoughtful creation of additional variables needed to tease out the full meaning of data. Selected univariate and bivariate analyses were first performed to explore the data sets. Logistic regression modeling and T-Test were employed as two major means for studying the differences in academic performance between the students who received college remediation and those who did not (Cabrera, 1994). One-way analysis of variance was performed to make more detailed comparisons by further subgrouping the remedial students. Time series analysis was used as a tool in the longitudinal study, with an overall time frame of 6 years (1992-1998). Finally, survival analysis/event history methods were employed to examine student retention and model different modes of student departure from college (e.g., degree attainment) (DesJardins, Ahlburg, & McCall, 1997; Xiao, 1997; Tamada & Inman, 1996).

It should be noted that these procedures were used for multiple purposes, not simply statistical inference. As a matter of fact, since we intended to include all the freshmen of Fall 1992 as a panel, we actually did not need to make any such inference. The inferential results would make sense when the data were supposed to constitute a random sample. In research practice, nonetheless, tests of significance were often used to analyze non-random data, and some might argue that significance at least points to the presence of a relatively considerable effect. The inferential results included in this article should only be interpreted in such a manner (i.e., for a hypothetical random sample of a larger population) (Chen, 1998).

The need for and the outcome of college remediation (as measured by performance in placement tests and other courses, such as number of placement tests passed, GPA, as well as student retention/graduation patterns) were of focal interest and served as key dependent variables in the study. College remediation (remedial courses including Summer immersion programs) was the principal independent variable in the study of student performance in credit courses, retention, and graduation. Variables such as age, performance in high school, race/ethnicity, economic or financial condition, employment, and family obligations were considered, as additional independent or control variables in both the remediation need and outcome studies.

Results

Facts Associated with Students' Need for Remediation

The study panel, i.e., the fall 1992 freshmen enrolled in academic programs at the college, included 1,334 students aged 21 on average. The oldest student in this panel was 58 years of age, and youngest 17 (median=18, with missing values for 5 cases). There were only slightly more female students than male students (713 vs. 621, or 53.4% vs. 46.6%). In terms of ethnic background, white students constituted 68.6% of the panel, blacks, 11.5%, Puerto Rican, 4.0%, Hispanic, 3.9%, Asian, 6.0%, Indian-Native American, 0.1%, and others, 5.8%. Among them, 48.7% were native English speakers, 32.4% were native speakers of other languages (compared to 16% in Schmidt et al., 1999, who were most comfortable with a language other than English), while 19.0% did not report their native language.

Table 1.1 shows the results of the three basic skills (reading, writing, and math) placement tests for this panel. It seemed that the students had difficulties mostly in the areas of math and writing, particularly the latter as nearly 60% of them failed the test and thus needed to take remedial courses. To explore the potential factors and characteristics associated with the need for remedial education, we first performed bivariate analyses to examine our hypotheses regarding the role of students' performance in high school and the effect of student age, which reflected the impact of the length of interruption in schooling. Then we used logistic regression modeling to integrate the results and take into consideration other potentially related facts.

Table 1.1 Results of Three CUNY Basic Skills Placement Tests

Test Type	Passed %	Failed %	Total N
Reading	83.7	16.3	1,329
Writing	40.2	59.8	1,322
Math	51.3	48.7	1,318

The panel had an average score of 64.36 (median=70.60) on the high school performance index (i.e., high school average, abbreviated variable name HSAVG), with a range of 11.0 to 95.0 (SD=21.11). T-Test analyses showed that the students who passed the three basic skills placement tests did have higher high school average than the students who failed the tests, with a difference of 6.20 for reading, 4.18 for writing, and 5.33 for math, respectively ($p < .001$ for all three hypothetical significance tests). On the other hand, T-Test analyses showed that student age had an impact on the test results, that is, those students who passed the three basic skills placement tests did have different mean ages than the students who failed the tests, though the results were mixed (i.e., a difference of 0.28 year older for reading, 0.59 year older for writing, and 1.12 year younger for math; $p < .001$ for all three hypothetical significance tests). It could be that age reflects not only the impact of the time interval at home or in the workplace before returning to school, which played a more important role in math skills, but also the effect of maturity, which prevailed in language skills.

Our logistic regression modeling incorporated the above exploratory results and also took into consideration the role of gender, ethnicity, native language, family income, employment status, and household status. High percentages of immigrant students who knew English as a second language were a common reason for remediation (Schmidt et al., 1999) although doubts had been cast on its role in this particular college (e.g., Lore & Murtha, 1997; Volpe, 1997). We wanted to reexamine the issue by including appropriate data. We did not choose such variables as "birth place," "birth country," "mother's birth place," and "father's birth place" because directly examining the language issue would have greater and more immediate relevance. Besides, substantial numbers of cases had missing values on those variables. For the same reason, we decided to use the indicator of "native language" rather than "language most comfortable with" or "other language spoken at home." For the income variable, Table 1.2 gives a profile of the financial situation of the student panel, which was rather consistent with the larger picture of the entire CUNY system (Schmidt et al., 1999). It is noticeable that over half of the student panel had a yearly family income below $24,000. Related to this fact, over half of the students were working full-time (35 hours or more per week) and part-time (fewer than 35 hours) (see Table 1.3), and this was related to the fact that many students were enrolled on a part-time basis (see Table 1.4). Of the 1,334 students included, 979 (73.4%) were enrolled on a full-time basis while 355 (26.6%) on a part-time basis. The variable of household status indicated the different family responsibilities of the students (see Table 1.5).

Table 1.2 Family Income of the Student Panel

	Frequency	Percent	Valid Percent	Cumulative Percent
<$4,000	102	7.6	9.1	9.1
4k-7,999	77	5.8	6.9	15.9
8k-11,999	77	5.8	6.9	22.8
12k-15,999	70	5.2	6.2	29.0
16k-19,999	70	5.2	6.2	35.3
20k-23999	173	13.0	15.4	50.7
24k+	554	41.5	49.3	100.0
Subtotal	1123	84.2	100.0	
Missing cases	211	15.8		
	------	-------		
Total	1334	100.0		

Table 1.3 Employment Status of the Student Panel

	Frequency	Percent	Valid Percent	Cumulative Percent
full-time	177	13.3	15.4	15.4
part-time	452	33.9	39.4	54.8
not employed, seeking	202	15.1	17.6	72.4
not employed	316	23.7	27.6	100.0
Subtotal	1147	86.0	100.0	
Missing cases	187	14.0		
	------	--------		
Total	1334	100.0		

Table 1.4 Employment Status by FT/PT Student Status

		Student Status		
		Full-time	Part-time	Total
Employment:	full-time	7.2%	38.2%	15.4%
	part-time	43.1%	29.3%	39.4%
	Not employed, seeking	18.9%	14.1%	17.6%
	not employed	30.8%	18.4%	27.6%
		---------	---------	----------
	Total	100.0%	100.0%	100.0%

Lambda= 0.181 (FT/PT as dependent variable), p< .001

Table 1.5 Household Status of the Student Panel

	Frequency	Percent	Valid Percent	Cumulative Percent
one of parents head of household	834	62.5	73.2	73.2
student or spouse head of household	305	22.9	26.8	100.0
Subtotal	1139	85.4	100.0	
Missing cases	195	14.6		
	------	-------		
Total	1334	100.0		

The preliminary findings of direct logistic regression analyses on the reading, writing, and math placement test results did not bode well for the role of gender, ethnicity, and family income. In numerous test runs of the analytical procedure they did not appear to be suitable predictors based on which a working model could be built (this might have important meaning for women, minority, and poor students against an often biased impression about them as inferior performers). Consequently, these variables were excluded from the models, which also helped to reduce complexity. The working model included 5 predictors: high school average (HSAVG), age, native language (NATVLANG), employment status (EMPLOYED), and household status (HOUSEHLD). Categorical variables were contrasted by the deviation method, with the effect for each category of an independent variable except one being compared to the overall effect (this was preferred since comparing to the last category might not make good sense for some of the variables). Tests of the model against a constant-only model indicated that the predictors, as a set, were reliable for predicting the results of the writing skills test (i.e., the need for remediation in writing skills, $X^2 = 23.40$, $p < .01$). For reading and math skills tests, the results were not statistically significant. However, since we were not dealing with a random sample we were more concerned with the discriminating ability of the models than with the issue of statistical inference. With the default cut point of 0.5, prediction success was a case of extremes for reading skills test, with 100% of those who passed and 0% of those who failed correctly predicted, for an overall success rate of 84.10%. For the writing test, prediction success presented a less contradictory case, with 20.96% of those who passed and 90.79% of those who failed correctly predicted, for an overall success rate of 62.23%. For the math test, prediction success was further balanced but less impressive on the whole, with 42.31% of those who passed and 72.57% of those who failed correctly predicted, for an overall success rate of 57.84%.

Table 1.6 displays the regression coefficients (B), standard errors (S.E.), Wald statistics, and odds ratios (Exp(B)) for each of the 5 predictors. According to the Wald criterion (less important in this case-population study) combined with the effect size (more meaningful, as measured by both the regression coefficient and the deviation of odds ratio from 1), native language would be a good predictor across three tests. This was on the contrary to the suspicion that the college was an exception to the impact of English as a second language (Lore & Murtha, 1997; Volpe, 1997). It should be noted that, though native speakers of English outperformed non-native speakers in language skills tests, they were surpassed by the latter in math tests. The ESL students seemed to be advantaged in math skills while English as a native language was associated with a greater need for remediation in this area (note in Table 1.6 the encoding of the math test result is contrary to that of the language test results). In addition, full-time (35 hours or more per week) employment status seemed to be linked with better language skills. Age of student, on the other hand, did not seem to be a consistently good predictor. As for the role of household status, it might have weakened the role of aging since the correlation matrix showed a relatively strong relationship with age. The inclusion of the interaction item, however, did not show any impact on the results. This variable played an important part in predicting the success in math placement tests.

The fact that high school average had only marginal effect sizes as compared with students' native language in predicting their needs for remediation suggested that the students might be able to make a fresh start even if their high school index was not that good. Although the former generally had better significance test results, in this study such results were only hypothetical while the effect sizes should count. The results could help to clarify the grounds of the open admissions plus remediation policy, particularly in view of the needs of the ESL students and those who were household heads.

Table 1.6 Facts Associated With the Need for Remediation: Logistic Regression

READING (Dependent Variable Encoding: P —> 0, F —>1):

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
AGE	-0.0067	0.0176	0.1453	1	0.7030	0.0000	0.9933
NATVLANG			15.4342	2	0.0004	0.1077	
English	-0.4726	0.1214	15.1578	1	0.0001	-0.1155	0.6234
Other	0.2485	0.1213	4.1964	1	0.0405	0.0472	1.2820
(Unknown)							
EMPLOYED			5.5769	3	0.1341	0.0000	
Full-Time	0.0094	0.1422	0.0043	1	0.9475	0.0000	1.0094
Part-Time	-0.3227	0.2142	2.2693	1	0.1320	-0.0165	0.7242
No, Seeking	-0.0003	0.1719	0.0000	1	0.9987	0.0000	0.9997
(Not Employed)							
HOUSEHLD							
Parent Head	0.0983	0.1255	0.6144	1	0.4331	0.0000	1.1033
(Student Head)							
HSAVG	-0.0107	0.0039	7.7665	1	0.0053	-0.0765	0.9893
Constant	-0.8873	0.5049	3.0878	1	0.0789		

	Chi-Square	df	Significance
Goodness-of-fit test (Hosmer and Lemeshow)	11.2122	8	.1900

WRITING (Dependent Variable Encoding: P —> 1, F —>0):

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
AGE	0.0268	0.0134	4.0341	1	0.0446	0.0366	1.0272
NATVLANG			15.9820	2	0.0003	0.0889	
English	0.2943	0.0870	11.4309	1	0.0007	0.0789	1.3421
Other	-0.3222	0.1003	10.3306	1	0.0013	-0.0741	0.7245
(Unknown)							
EMPLOYED			13.7734	3	0.0032	0.0716	
Full-Time	0.0443	0.1027	0.1862	1	0.6661	0.0000	1.0453
Part-Time	0.3680	0.1472	6.2486	1	0.0124	0.0530	1.4449
No, Seeking	-0.0416	0.1262	0.1084	1	0.7419	0.0000	0.9593
(Not Employed)							
HOUSEHLD							
Parent Head	0.0720	0.0972	0.5486	1	0.4589	0.0000	1.0746
(Student Head)							
HSAVG	0.0095	0.0035	7.2936	1	0.0069	0.0591	1.0096
Constant	-1.6185	0.4189	14.9320	1	0.0001		

	Chi-Square	df	Significance
Goodness-of-fit test (Hosmer and Lemeshow)	23.4041	8	.0029

Table 1.6 Facts Associated With the Need for Remediation: Logistic Regression (continued)

MATH (Dependent Variable Encoding: P —> 1, F —>0):

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
AGE	0.0047	0.0134	0.1228	1	0.7261	0.0000	1.0047
NATVLANG			15.6925	2	0.0004	0.0867	
English	-0.1271	0.0862	2.1740	1	0.1404	-0.0106	0.8806
Other	0.3885	0.0981	15.6897	1	0.0001	0.0938	1.4748
(Unknown)							
EMPLOYED			1.3086	3	0.7271	0.0000	
Full-Time	0.0146	0.1018	0.0206	1	0.8858	0.0000	1.0147
Part-Time	0.0341	0.1460	0.0545	1	0.8154	0.0000	1.0347
No, Seeking	0.0701	0.1243	0.3177	1	0.5730	0.0000	1.0726
(Not Employed)							
HOUSEHLD							
Parent Head	0.2999	0.0947	10.0299	1	0.0015	0.0719	1.3497
(Student Head)							
HSAVG	0.0138	0.0033	17.3535	1	0.0000	0.0994	1.0139
Constant	-1.0829	0.4022	7.2484	1	0.0071		

	Chi-Square	df	Significance
Goodness-of-fit test (Hosmer and Lemeshow)	11.4377	8	.1781

Outcome of Remedial Education

To answer the question how effective college remediation might be, we looked at the case college in terms of the rates of passing CUNY basic skills tests after up to a year of remediation. Here program evaluation in its full sense was not feasible since every student who failed a basic skills test was supposed to take remedial courses other than to be put in a control group without access to remediation. Table 1.7 shows that the pass rates of the subsequent reading, writing, and math tests (posttests), as compared to the first placement test (pretest) results contained in Table 1.1, had significant increases (9.7%, 39.2%, and 28.6%, respectively). Although math and writing were still two major areas of difficulty, they were also the major areas of improvement. The total impact of remediation, the completion of which was represented by all three placement tests passed, is shown by Table 1.8. It seemed that college remediation had a good turnout in student achievements. After up to a year remediation, for instance, 43.2% of the student panel passed all the basic skills tests, who constituted 56.3% of the students that failed to pass the same kind of tests at their first try.

Table 1.7 Results of Three CUNY Basic Skills Placement Tests One Year Later

Test Type	Passed %	Failed %	Total N
Reading	93.4	6.6	1,334
Writing	79.4	20.6	1,334
Math	79.9	20.1	1,334

Table 1.8 Remediation at the Case College: Overall Statistics (1992-93)

	Frequency	Percent	Valid Percent	Cumulative Percent
non-remedial	311	23.3	23.3	23.3
completed rem in 1 yr	576	43.2	43.2	66.5
not compld rem in 1 yr	447	33.5	33.5	100.0
	------	-------	-------	
Total	1334	100.0	100.0	

How does college remediation impact on other aspects of academic performance, as measured by results in other (credit) courses? Of particular interest is the students' cumulative academic index, i.e., the grade point average (GPA, excluding grades in remedial courses). Table 1.9 contains information about the academic performance of the student panel in terms of a few distribution parameters of their GPA's over the years up to 1998. To assess the outcome of remedial education, however, we must compare the students who had received college remediation with those who had not. Table 1.10 contains the results of such a comparison through T-Test analysis. Taken as a time series, the data clearly showed a trend for the remedial students to close their gap with non-remedial students in a normal period of college study (approximately 4 years). That is, the academic difference between the two groups in terms of their GPA's became insignificant after three years of study. In fact, in their fifth year of study the difference was actually reversed, with the remedial group having a higher average GPA than the non-remedial group. This was amazing indeed, which, naturally, would give some credit to remediation and lend support to the open admissions policy provided that there was no significant "grade inflation" (cf. Part II; Arenson, 1996; Adelman, 1995) particularly favoring the remedial students. While this is yet a hypothesis requiring more rigorous research designs, anyone who is unbiased should have no difficulty to conclude from Table 1.10 and Figure 1.1 that the remedial students did deserve the educational opportunity given to them.

Table 1.9 GPA Distribution of the Student Panel (1993-98)

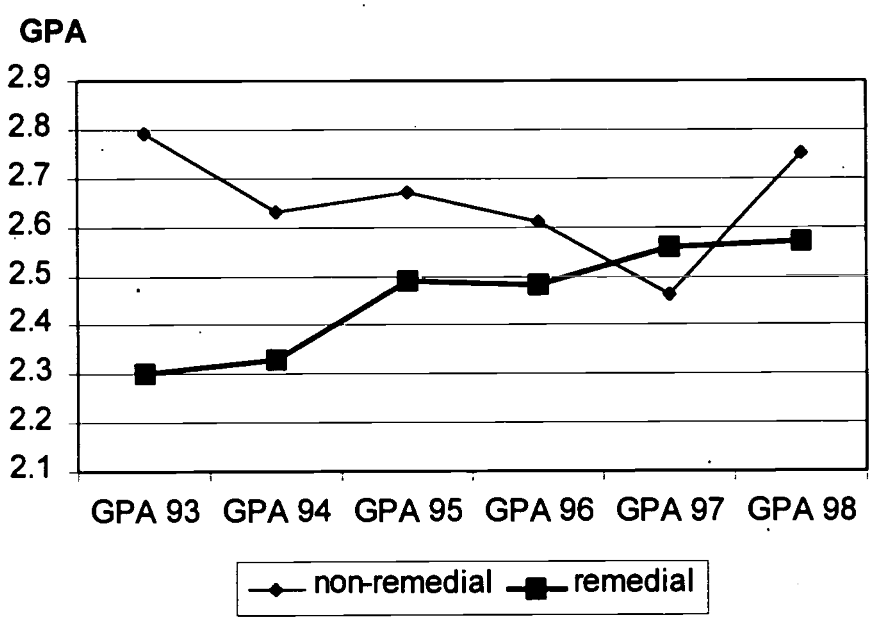
	GPA93	GPA94	GPA95	GPA96	GPA97	GPA98
Median	2.5000	2.5000	2.5800	2.5900	2.5750	2.7600
Mean	2.4152	2.4093	2.5409	2.5152	2.5358	2.6236
S.D.	.8811	.7809	.6406	.6556	.7101	.8588
N	882	615	455	323	198	140
Missing cases	452	719	879	1011	1136	1194

Table 1.10 Remedial Students and the Outcome of Remediation: Comparison of GPA's (1993-98)

		N	Mean	S.D.	Mean Difference
GPA93:	Non-remedial	213	2.7867	.6861	
	Remedial	669	2.2970	.9036	.4897**
GPA94:	Non-remedial	168	2.6308	.7130	
	Remedial	447	2.3260	.7898	.3048**
GPA95:	Non-remedial	131	2.6731	.6733	
	Remedial	324	2.4874	.6200	.1857*
GPA96:	Non-remedial	84	2.6089	.6228	
	Remedial	239	2.4823	.6648	.1267
GPA97:	Non-remedial	50	2.4626	.8686	
	Remedial	148	2.5605	.6495	-.0978
GPA98:	Non-remedial	42	2.7538	.9734	
	Remedial	98	2.5679	.8038	.1860

* p< .01 **p< .001

Figure 1.1 Comparison of the Progresses of Remedial and Non-Remedial Students



It should be noted that although our original research proposal and previous presentations have used the term "the impact of remediation," the focus of this study was not a formal evaluation research design because it was not feasible. An impact assessment would make sense only when other things were made equal through such a strategy as randomization. But randomized assignment of the students into the remedial and non-remedial groups was impossible and, therefore, too many unknown factors would interfere with the determination of the impact of remediation. Directly comparing the two groups with each other, however, would still address a number of practical questions: How did the remedial group perform in relation to the non-remedial group? What was the potential influence of remediation on retention and graduation? Was remedial education worthwhile? Should open admissions take any blame for the remedial outcome? Findings bearing on these questions should be important even if we were unable to determine the relative significance of remediation in relation to the role of other institutional and student characteristics that were likely incomparable. For example, if the students who did not do well in entrance tests *could* catch up and excel given necessary time and educational exposure despite an uncertain benefit of remedial instruction itself, should we simply end it and shut the college door, which would mean a denial or total loss of those students' potential achievements, or should we keep up the open admissions and remediation policies? This, of course, was a very political question. But the bottom line was for us to find out whether or not the remedial students *could* or *did* catch up however complicated the potential reason.

To give a closer examination of the remedial students' academic performance, we further analyzed the data using the ONEWAY procedure. The results in Table 1.11 reconfirmed the T-Test findings while providing more detailed information on the differences between three groups of students in the study panel, i.e., those who did not need remediation, those who completed remedial courses within a year, and those who did not complete the remedial courses within a year. Figure 1.2 demonstrates that the "closing gap" trend found in the T-Test analyses was due to a constant improvement of the remedial students (especially those whose remedial needs went beyond a year) on one hand and a relatively unstable performance of the non-remedial students on the other.

The remedial students were further regrouped into three categories: those who failed one basic skills test at entry, those who failed two, and those who flunked all three. This variable shows how much the difficulty/need was at entry while the preceding analysis was about how long the difficulty/need persisted. The ONEWAY procedure was utilized again to compare these three groups plus a group of those who did not fail any of the basic skills tests at entry. Table 1.12 and Figure 1.3 contain the results. The findings were similar to the above ONEWAY analysis, yet further suggested that no matter how poorly a student performed in the entry tests, the chance for him or her to improve and catch up was great if he or she stayed on the path. Space forbids and we cannot pursue an internal turnover analysis here to show the actual within-group process. But the information in the table and the figures is clear enough as to how dramatic improvement could be made even for those who flunked all three basic skills tests at entry.[2] And with high plausibility one might speculate that college remediation did have a positive impact on student achievements.

Table 1.11 Outcome of Remediation: One-Way Analysis of Variance on Three Groups

	Remediation Needs Over Time	N	Mean	S.D.	Std Error
GPA93:	non-remedial	213	2.7867**	.6861	4.701E-02
	completed remediation in 1 yr	440	2.4862	.7937	3.784E-02
	not completed remediation in 1 yr	229	1.9334	.9887	6.533E-02
	Total	882	2.4152	.8811	2.967E-02
GPA94:	non-remedial	168	2.6308**	.7130	5.501E-02
	completed remediation in 1 yr	303	2.4473	.7276	4.180E-02
	not completed remediation in 1 yr	144	2.0708	.8548	7.123E-02
	Total	615	2.4093	.7809	3.149E-02
GPA95:	non-remedial	131	2.6731*	.6733	5.883E-02
	completed remediation in 1 yr	230	2.5467	.6068	4.001E-02
	not completed remediation in 1 yr	94	2.3423	.6314	6.513E-02
	Total	455	2.5409	.6406	3.003E-02
GPA96:	non-remedial	84	2.6089*	.6228	6.796E-02
	completed remediation in 1 yr	169	2.5556	.6274	4.826E-02
	not completed remediation in 1 yr	70	2.3053	.7219	8.629E-02
	Total	323	2.5152	.6556	3.648E-02
GPA97:	non-remedial	50	2.4626	.8686	.1228
	completed remediation in 1 yr	99	2.6031	.5872	5.902E-02
	not completed remediation in 1 yr	49	2.4743	.7593	.1085
	Total	198	2.5358	.7101	5.047E-02
GPA98:	non-remedial	42	2.7538	.9734	.1502
	completed remediation in 1 yr	66	2.6047	.7724	9.507E-02
	not completed remediation in 1 yr	32	2.4919	.8727	.1543
	Total	140	2.6236	.8588	7.259E-02

* $p < .01$ ** $p < .001$

Figure 1.2 Comparison of Students with Different Remedial Needs (1)

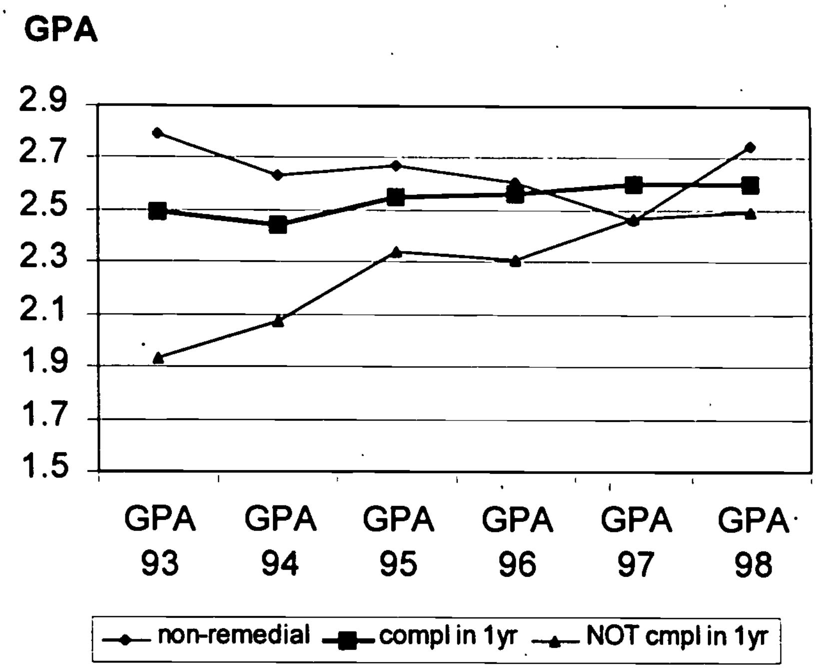
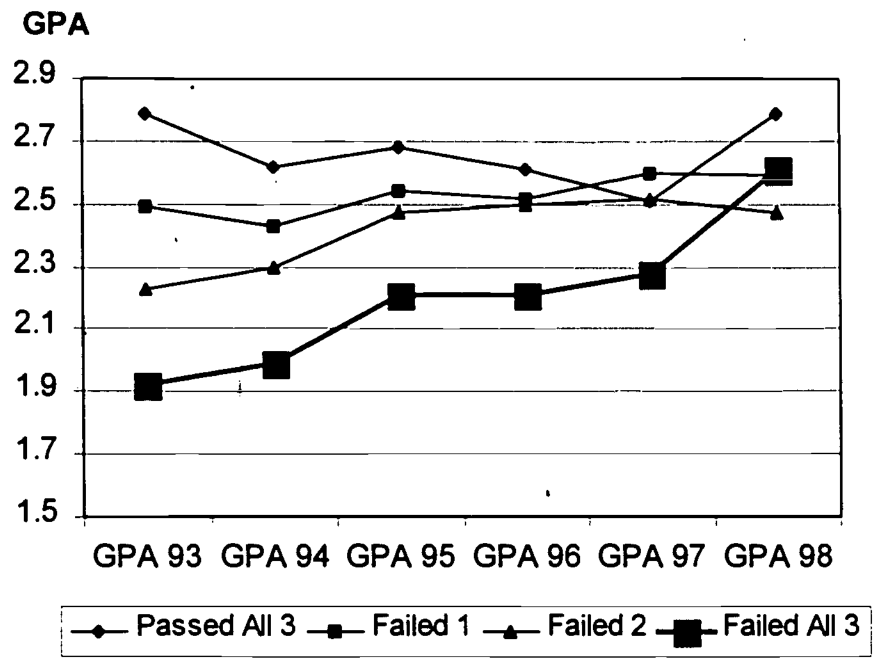


Table 1.12 Outcome of Remediation: One-Way Analysis of Variance on Four Groups

	Skill Test Results at Entry	N	Mean	S.D.	Std. Error
GPA93:	passed all 3	216	2.7946*	.6856	4.665E-02
	failed 1	348	2.4312	.8841	4.739E-02
	failed 2	222	2.2333	.8762	5.881E-02
	failed all 3	96	1.9245	.9199	9.389E-02
	Total	882	2.4152	.8811	2.967E-02
GPA94:	passed all 3	171	2.6298*	.7217	5.519E-02
	failed 1	232	2.4254	.7551	4.957E-02
	failed 2	155	2.2950	.8039	6.457E-02
	failed all 3	57	1.9926	.7830	.1037
	Total	615	2.4093	.7809	3.149E-02
GPA95:	passed all 3	133	2.6838*	.6756	5.858E-02
	failed 1	179	2.5446	.6523	4.876E-02
	failed 2	108	2.4674	.5634	5.421E-02
	failed all 3	35	2.2054	.5186	8.765E-02
	Total	455	2.5409	.6406	3.003E-02
GPA96:	passed all 3	84	2.6089	.6228	6.796E-02
	failed 1	129	2.5195	.6741	5.935E-02
	failed 2	86	2.5033	.6442	6.947E-02
	failed all 3	24	2.2067	.6501	.1327
	Total	323	2.5152	.6556	3.648E-02
GPA97:	passed all 3	53	2.5113	.8714	.1197
	failed 1	82	2.6023	.7460	8.239E-02
	failed 2	50	2.5202	.4735	6.696E-02
	failed all 3	13	2.2754	.4556	.1264
	Total	198	2.5358	.7101	5.047E-02
GPA98:	passed all 3	44	2.7861	.9641	.1453
	failed 1	57	2.5895	.7734	.1024
	failed 2	34	2.4721	.8512	.1460
	failed all 3	5	2.6140	.8871	.3967
	Total	140	2.6236	.8588	7.259E-02

* $p < .001$

Figure 1.3 Comparison of Students with Different Remedial Needs (2)



Difference in Retention and Graduation

Was there any major difference between the students who received remediation and those who did not in terms of their retention and graduation patterns? Were there other factors that might also be related to student retention and graduation? Traditional approach to these questions focuses on the calculation of graduation rates and testing of related hypotheses. Yet the calculation of rates is often problematic without taking into consideration time and alternative outcomes. A potential issue is that longitudinal data such as these are often "censored," meaning that the true value of the duration time for a subject may be unknown since the end event may have not occurred. A useful procedure called survival (duration) analysis or event history method has been developed to deal with the problem of censoring, which was applied in the present study.

To better address the above questions, we tried to examine the lengths of time between entry and different end events and make the group comparisons. It should be noted that ours was a panel study with the same entry year for all students included, which alleviated the censoring problem in one respect. It made sense, therefore, to first apply the ordinary techniques to the examination of some carefully selected and created variables. Table 1.13 displays the results of T-Test and ONEWAY analyses on the potential group differences in the number of years dropping and staying out of any degree program ("stop-out"). Except for the somewhat mixed results for one variable (employment status), the findings supported all of our hypotheses from a longitudinal perspective. Especially, data showed that a greater need for or a worse outcome of remediation resulted in a larger number of stop-out years. This tentative finding suggests that if the institution wants to maintain its open admissions policy based on the kind of optimistic results shown earlier, it should be prepared to allow the remedial students longer time before they can enter or return to study in degree programs. The same conclusion applies to ESL, low-income, and part-time students and those who were household heads. For the variable of employment status, it seemed that working part-time was not particularly detrimental to school study, though excessive workload as indicated by full-time employment status at entry did predict longer time of stop-out later.

Table 1.13 Comparison of Numbers of Years Staying Outside Any Degree Program

		N	Mean	S.D.
Family Income:	$24k+	554	3.90**	1.84
	low income	569	4.18	1.67
FT/PT Status:	ft	979	3.91****	1.78
	pt	355	4.45	1.74
Native Language:	English	649	3.91***	1.84
	other	432	4.28	1.68
Household Status:	parent household head	834	3.96*	1.78
	student/spouse househd head	305	4.24	1.71
Employment:	full-time	177	4.40*	1.63
	part-time	452	3.98	1.81
	not employed, seeking	202	4.02	1.71
	Not employed	316	3.95	1.78
	Total	1147	4.04	1.76
Skills Test Results:	passed all 3	322	3.84*	1.80
	failed 1	515	4.01	1.80
	failed 2	358	4.18	1.79
	failed all 3	139	4.35	1.59
	Total	1334	4.05	1.78
Remed. outcome:	non-remedial	311	3.80****	1.81
	cmpld rem in 1yr	576	3.74	1.81
	not cmpld rem in 1yr	447	4.62	1.58
	Total	1334	4.05	1.78

* $p < .05$ ** $p < .01$ *** $p < .005$ **** $p < .001$

Although the above results appear interesting, prediction based on entry characteristics was limited by the uncertainty whether or not they would continue to hold true over time, especially for such variables as employment status. Since our study focused on the comparison of remedial and non-remedial students, survival analysis would be a more powerful tool (Cheng,

1997) for the modeling of the "hazard" or "failure time" data, especially for dealing with the censoring problem. As a matter of fact, our data indicated that only one-fifth of our student panel graduated with a degree or certificate within the whole period of 6 years. To utilize the survival analysis technique to deal with this issue, we constructed two survival variables for examining the length of time between students' entry and two different end events, that is, retention/dropout and graduation (see Table 1.14). Other possible variables include the first or the longest period of time of being out of school for those students who ever dropped out but later returned, though we are not able to pursue those details here as both space and time forbid. The terminal event in this analysis was graduation, thus the time it took to graduate also served as an indicator of survival status with various values of the variable. According to our hypotheses, the results of the three basic skills tests at entry did not constitute the sole determinant of the patterns of student retention and graduation, as suggest by the above preliminary findings. Other variables included language, economic condition, employment, and full-time/part-time student status, which represented various factors that might bear on the occurrence of the end events studied. These additional variables, however, were used at a higher level of control in our modeling under which the comparisons of the groups with different remediation needs and outcomes were made. Since the output contained numerous tables and figures and the analysis went far beyond the scope of this report, we omit the results of higher level control and only report the basics in the following. And we focus on the outcome of remediation while the other dependent variable of remediation need is omitted from here.

Originated from its application in demographic, actuary and medical research, survival analysis is frequently carried out through the construction of "life tables." The most fundamental items on these tables are survival functions (distributions of the surviving as percentages of the total), hazard functions (distributions of those who died during the year as percentages of the remaining survivors, which yield such hazard rates as mortality rates on a diminishing yearly base), density functions, and censoring. Tables 1.15 and 1.16 contain the life tables for this study, based on which we can make subgroup and pairwise comparisons for the selected survival and control variables. Figures 1.4 through 1.7 show the plot output for the survival and hazard functions. The latter exhibit a generally positive time dependence, that is, hazard rate increases over time, which is pertinent to our particular research population and time frame. Consistent with the T-Test and ONEWAY analysis results on the "staying out" variable, the findings clearly distinguished between the student groups with different remediation needs and outcomes. That is, a greater need for or a worse outcome of remediation resulted in a larger number of years taken to graduate, which also meant a longer period of retention. Here we should note that retention means very differently for students than for faculty/staff. For the latter, long periods of retention mean long-time services and normally serve as a good indicator for an institution. For the former, being held in college by itself is a matter of investment, and it is usually desirable for the students to complete their studies as soon as possible to reduce that cost. In such a sense, retention will only have a positive meaning as opposed to (premature) dropout, or abortion rather than successful completion of a study plan.

Table 1.14 Distribution of the Survival Variables

Time Taken to Graduate With a Degree Awarded

Number of Years	Frequency	Percent	Valid Percent	Cumulative Percent
2	2	.1	.1	.1
3	27	2.0	2.0	2.2
4	62	4.6	4.6	6.8
5	119	8.9	8.9	15.7
6	67	5.0	5.0	20.8
not graduated yet after 6 years	1057	79.2	79.2	100.0
Total	1334	100.0	100.0	

*Retention in Degree Programs**

Number of Years	Frequency	Percent	Valid Percent	Cumulative Percent
0	365	27.4	27.4	27.4
1	262	19.6	19.6	47.0
2	161	12.1	12.1	59.1
3	155	11.6	11.6	70.7
4	151	11.3	11.3	82.0
5	100	7.5	7.5	89.5
6	140	10.5	10.5	100.0
Total	1334	100.0	100.0	

* Including years of stop-out for those who came back, but not including time in remedial and non-degree study.

Time Staying Outside Any Degree Program

Number of Years	Frequency	Percent	Valid Percent	Cumulative Percent
0	47	3.5	3.5	3.5
1	105	7.9	7.9	11.4
2	151	11.3	11.3	22.7
3	170	12.7	12.7	35.5
4	185	13.9	13.9	49.3
5	311	23.3	23.3	72.6
6	365	27.4	27.4	100.0
Total	1334	100.0	100.0	

Table 1.15 Life Tables Comparing Retention of Groups With Different Remediation Outcomes

Survival Variable RETENT for CMP_REM = 1, non-remedial

Intrvl Start Time	Number Entrng This Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Proba-Bility Densty	Hazard Rate
0	311.0	78.0	272.0	1.0	0.0037	0.9963	0.9963	0.0037	0.0037
1	232.0	33.0	215.5	5.0	0.0232	0.9768	0.9732	0.0231	0.0235
2	194.0	31.0	178.5	7.0	0.0392	0.9608	0.9350	0.0382	0.0400
3	156.0	24.0	144.0	28.0	0.1944	0.8056	0.7532	0.1818	0.2154
4	104.0	11.0	98.5	30.0	0.3046	0.6954	0.5238	0.2294	0.3593
5	63.0	10.0	58.0	11.0	0.1897	0.8103	0.4245	0.0993	0.2095
6	42.0	23.0	30.5	19.0	0.6230	0.3770	0.1600	0.2644	0.9048

The median survival time for these data is 5.24

Survival Variable RETENT for CMP_REM = 2, completed remediation in 1 yr

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Proba-Bility Densty	Hazard Rate
0	576.0	106.0	523.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
1	470.0	128.0	406.0	5.0	0.0123	0.9877	0.9877	0.0123	0.0124
2	337.0	60.0	307.0	12.0	0.0391	0.9609	0.9491	0.0386	0.0399
3	265.0	45.0	242.5	26.0	0.1072	0.8928	0.8473	0.1018	0.1133
4	194.0	27.0	180.5	49.0	0.2715	0.7285	0.6173	0.2300	0.3141
5	118.0	20.0	108.0	32.0	0.2963	0.7037	0.4344	0.1829	0.3478
6	66.0	43.0	44.5	23.0	0.5169	0.4831	0.2099	0.2245	0.6970

The median survival time for these data is 5.64

Survival Variable RETENT for CMP_REM = 3, not completed remediation in 1 yr

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Proba-Bility Densty	Hazard Rate
0	447.0	180.0	357.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
1	267.0	91.0	221.5	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
2	176.0	51.0	150.5	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
3	125.0	25.0	112.5	7.0	0.0622	0.9378	0.9378	0.0622	0.0642
4	93.0	23.0	81.5	11.0	0.1350	0.8650	0.8112	0.1266	0.1447
5	59.0	20.0	49.0	7.0	0.1429	0.8571	0.6953	0.1159	0.1538
6	32.0	28.0	18.0	4.0	0.2222	0.7778	0.5408	0.1545	0.2500

The median survival time for these data is 6.00+

Table 1.16 Life Tables Comparing Numbers of Years Taken To Graduate

Survival Variable LENGTHGR for CMP_REM =1, non-remedial

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-viving	Cumul Propn Surv at End	Proba-Bility Densty	Hazard Rate
0	311.0	0.0	311.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
1	311.0	0.0	311.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
2	311.0	0.0	311.0	1.0	0.0032	0.9968	0.9968	0.0032	0.0032
3	310.0	0.0	310.0	13.0	0.0419	0.9581	0.9550	0.0418	0.0428
4	297.0	0.0	297.0	22.0	0.0741	0.9259	0.8842	0.0707	0.0769
5	275.0	0.0	275.0	50.0	0.1818	0.8182	0.7235	0.1608	0.2000
6	225.0	0.0	225.0	15.0	0.0667	0.9333	0.6752	0.0482	0.0690
7.0+	210.0	210.0	105.0	0.0	0.0000	1.0000	0.6752	**	**

Survival Variable LENGTHGR for CMP_REM =2, completed remediation in 1 yr

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-Viving	Cumul Propn Surv at End	Proba-bility Densty	Hazard Rate
0	576.0	0.0	576.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
1	576.0	0.0	576.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
2	576.0	0.0	576.0	1.0	0.0017	0.9983	0.9983	0.0017	0.0017
3	575.0	0.0	575.0	14.0	0.0243	0.9757	0.9740	0.0243	0.0246
4	561.0	0.0	561.0	34.0	0.0606	0.9394	0.9149	0.0590	0.0625
5	527.0	0.0	527.0	58.0	0.1101	0.8899	0.8142	0.1007	0.1165
6	469.0	0.0	469.0	40.0	0.0853	0.9147	0.7448	0.0694	0.0891
7.0+	429.0	429.0	214.5	0.0	0.0000	1.0000	0.7448	**	**

Survival Variable LENGTHGR for CMP_REM =3, not completed remediation in 1 yr

Intrvl Start Time	Number Entrng this Intrvl	Number Wdrawn During Intrvl	Number Exposd to Risk	Number of Termnl Events	Propn Termi-nating	Propn Sur-Viving	Cumul Propn Surv at End	Proba-bility Densty	Hazard Rate
0	447.0	0.0	447.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
1	447.0	0.0	447.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
2	447.0	0.0	447.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
3	447.0	0.0	447.0	0.0	0.0000	1.0000	1.0000	0.0000	0.0000
4	447.0	0.0	447.0	6.0	0.0134	0.9866	0.9866	0.0134	0.0135
5	441.0	0.0	441.0	11.0	0.0249	0.9751	0.9620	0.0246	0.0253
6	430.0	0.0	430.0	12.0	0.0279	0.9721	0.9351	0.0268	0.0283
7.0+	418.0	418.0	209.0	0.0	0.0000	1.0000	0.9351	**	**

Note: The median survival time for these data is 7.00+; 1334 observations

** These calculations for the last interval are meaningless.

Figure 1.4 Number of Years Taken to Graduate: Survival Function

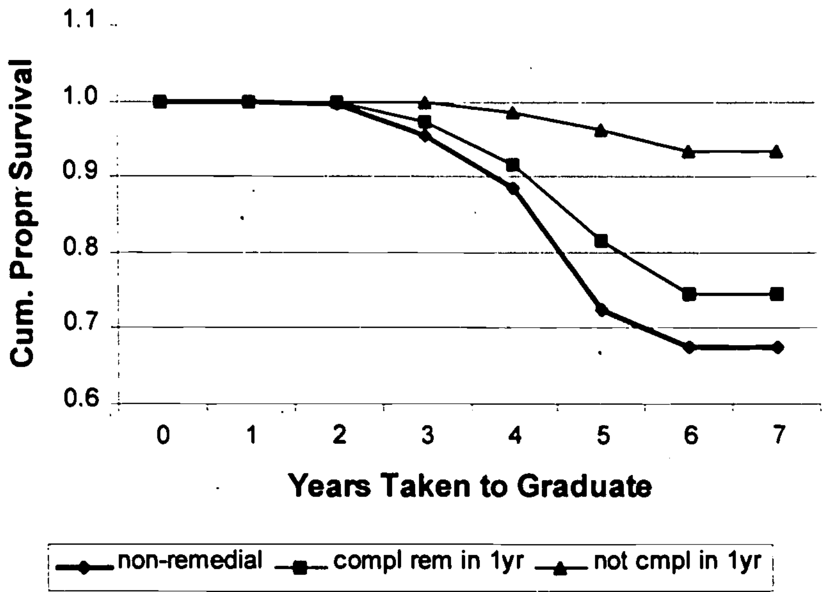


Figure 1.5 Number of Years Taken to Graduate: Hazard Function

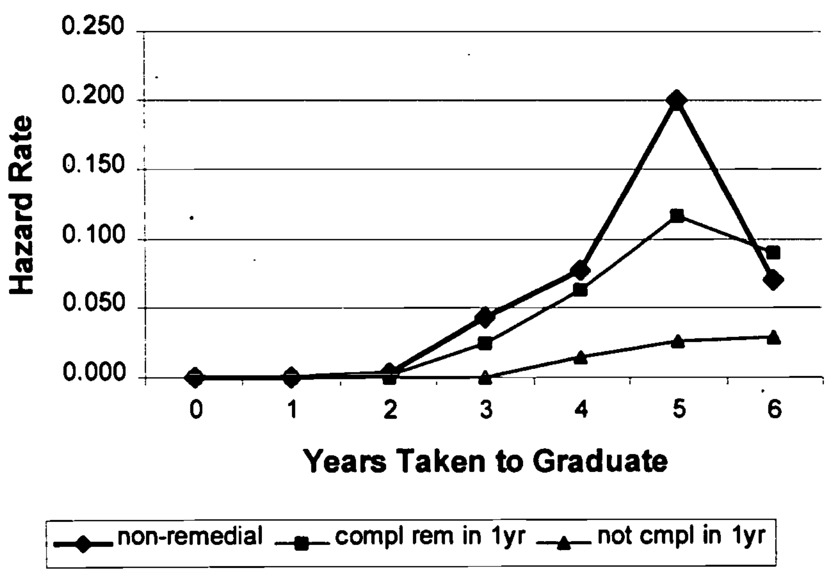


Figure 1.6 Number of Years of Retention (Including Interruption): Survival Function

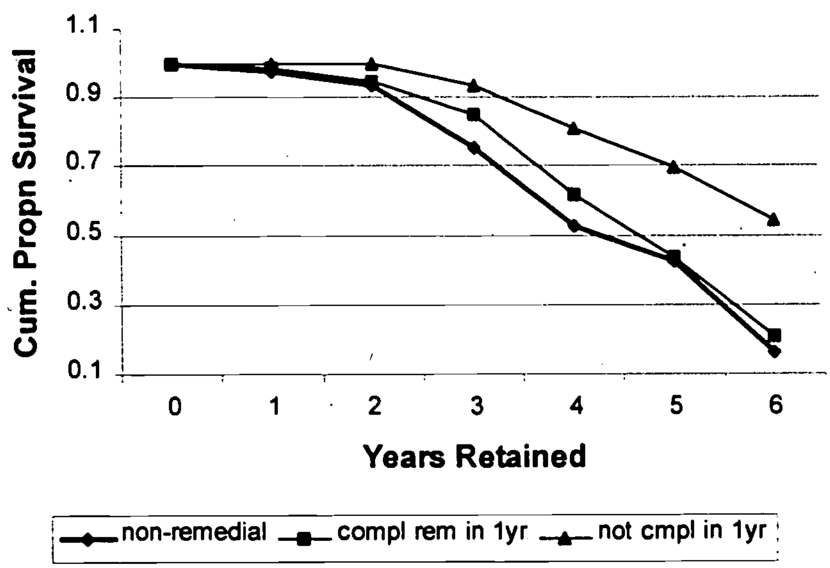
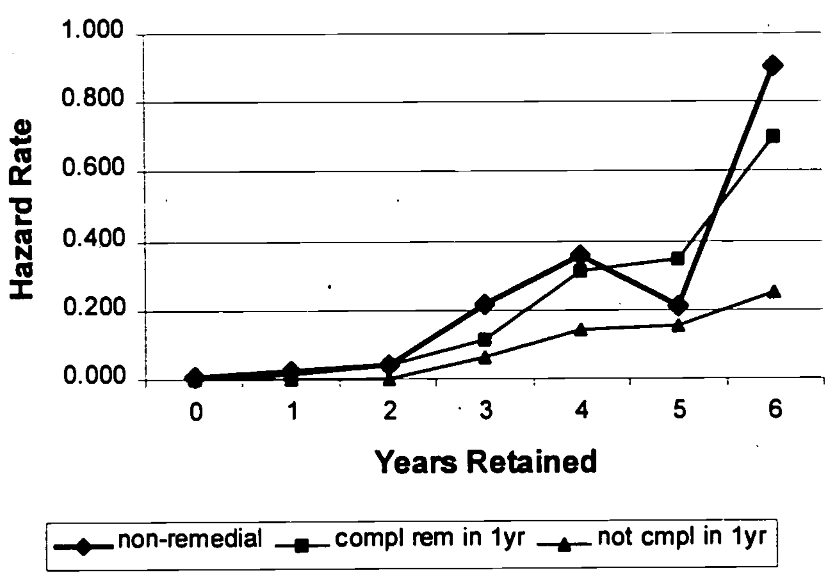


Figure 1.7 Number of Years of Retention (Including Interruption): Hazard Function



Conclusion and Discussion

Using one college as a case study, this panel study compared the students who received remedial education with those who did not in terms of their performance and achievement. Potential factors contributing to the need for remediation and those affecting student retention and graduation were also explored. The purpose was to illuminate the complicated implications of remediation, rather than to conduct a direct and straightforward impact assessment, in a particular context of the alleged crisis of CUNY resulted from its open admissions policy. The main findings of the study were:

(1) Factors effecting students' need for remediation: Students' native language played a major role in determining their need for remediation, that is, ESL students had a greater need for remediation in English than non-ESL students. In addition, those students who were household heads had a greater need for remediation in math. ESL students possessed better math skills than native speakers of English, while working full-time was associated with better reading and writing skills. The role of student age was marginal, with older students having performed slightly better in language, though not in math, tests at entry. It is noticeable that employment and household statuses had larger effect sizes than students' high school performance index. Although the latter generally had better significance test results, in this case-population study such results were only hypothetical while the effect sizes were more meaningful. The result indicates the limitation of high school average as a criterion for assessing students' preparedness.

(2) Difference of performance between remedial and non-remedial students: In terms of their ongoing academic performance, the findings seemed to speak positively for the remedial students. Taken as a time series, the data clearly showed a trend for the remedial students to close their gap with non-remedial students in a normal period of college study (approximately 4 years) as indicated by their average GPA's. In fact, in their fifth year of study the difference was actually reversed. Time series analysis further revealed that the "closing gap" trend was due to constant improvement of the remedial students on one hand and relatively unstable performance of the non-remedial students on the other. The present study was unable to identify the reasons due to lack of means of control, though it was natural to assume that remediation might have had a positive impact. Whatever the exact causes, the finding would lend support to the open admissions policy provided that there was no significant "grade inflation" particularly favoring the remedial students. The point is, if the impact of remedial instruction was hard or impossible to gauge, those students would at least benefit from the chance, time, and educational settings for them to prove themselves. Taking away such conditions altogether would mean eliminating a large category of students who would eventually perform well or even better than the high performers at entry. The finding seemed to suggest that the entrance test results alone could not predict the academic success of a student since the chance for a remedial student to improve and catch up was great if he or she stayed on the path. It was amazing indeed to see from the results how dramatic improvement could be made even for those who flunked all three basic skills tests at entry.

(3) Comparison of students with different remediation needs and outcomes in terms of their retention and graduation patterns: Except for one variable (i.e., employment status) with somewhat mixed results, the findings of T-Tests and One-Way analyses of variance were consistent and interestingly complementary to the above results from a longitudinal perspective. Specifically, data showed that a greater need for and a worse outcome of remediation resulted in a larger number of years staying out of any degree program. This finding suggested that if the

institution wants to maintain its open admissions policy based on the kind of optimistic results shown in the above, it should be prepared to allow the remedial students longer time before they can enter or return to study in degree programs.

Using two survival variables (i.e., retention/dropout and graduation), we examined the survival and hazard functions featuring the different lengths of time between students' entry and the two end events. The latter exhibited a generally positive time dependence, that is, hazard rate increased over time, which was pertinent to our particular research population and the time frame. Consistent with the T-Test and ONEWAY analysis results on the "staying out" variable, the findings clearly distinguished between the student groups with different remediation needs and outcomes. That is, a greater need for or a worse outcome of remediation resulted in more years taken to graduate, which also meant a longer period of retention.

All in all, the results signified the need of the ESL, working, and older (in terms of math) and younger (in terms of language skills) students for remediation and their great promise to improve given the chance. The findings also suggested that we should be prepared to allow the remedial students longer survival (i.e., retention) time before they can graduate with a degree in hand. This might have an impact on graduation rates, though no calculation was feasible and no conclusion could be drawn due to the censoring problem (particularly lack of transfer data). Given the promising performance and retention the remedial students have shown, the question is not whether remediation should be provided but how to help the remedial students move forward as quickly as possible and eventually complete their programs. On the other hand, the information gathered in this study seemed to support that college remediation has a positive effect on student achievements as a hypothesis for future validation study. Although the data did not allow for a reliable direct assessment of the impact of remedial instruction under controlled conditions, the present study has laid a necessary groundwork for future pursuit of a formal design of program evaluation.

This study was limited by both the data and the time constraint on analysis. Certain information was not included in the working data set, such as residence and transportation, number of credits completed in a given time period, and transfer out of the college. Besides, a substantial number of cases had missing values on a few variables, which were excluded from our analysis. Due to limitations in resource, we were unable to further probe into the role (except for some preliminary results) of gender, ethnicity, family income, and a few other variables related to the students' immigration history as well as language capability. With regard to the final outcomes particularly, we hypothesized that unpreparedness at entry is not the sole determinant of the patterns of student retention and graduation. Other factors include language, economic condition, employment, household status, and full-time/part-time student status, which were used at a higher level of control in our modeling under which the groups with different remediation needs and outcomes were compared. This report, however, has not fully covered the results.

All these data and analytical issues can be pursued in ensuing research projects with an ongoing effort to study facts related to higher educational policies, including the research questions raised in the New York City Mayor's Task Force's report regarding remediation. Our future investigation, for instance, may show more detailed retention/graduation patterns (e.g., continuous, interrupted, and terminated registration). Other variables may include the first or the longest period of time of being out of school for those students who ever dropped out but later returned. Further inquiry may involve new variables to assess the implication of the new policy of CUNY to limit remedial coursework to one year. In terms of the factors affecting students' performance in credit courses, multiple regression techniques may be employed to construct a more complete and integrated analytical model. To examine the change over time, we may conduct an internal turnover analysis to show the actual within-group process. We should also study various methodological issues in modeling the different modes of student entry into and departure from college. The Student Attrition Model (Bean, 1978, 1980, 1981; Price, 1977) that emphasizes the importance of the intention to remain enrolled or to depart from college may prove useful in illuminating the underlying mechanism, so may the Student Integration Model (Spady, 1970; Tinto, 1975) that stresses a matching between students' motivations and academic ability and the institution's academic and social characteristics. Future study may also link retention and graduation with issues of academic standards, specifically the subject of grade inflation. If resources are made available, our study can be expanded to include direct surveys of students, faculty, and administrators on issues of fundamental importance, including the validity of such standard tests as the SAT. It is also hoped that on the basis of the descriptive studies more rigorous experimental designs can be employed to further pursue the issues involved. Also, where feasible, the information can be compared to national data sets (e.g., IPEDS, NPSAS, and BPS) to create comparable national and local statistics for student characteristics and educational outcomes.

Although the results presented speak for themselves, they point to a great need for more intensive studies on a number of interrelated issues. Only on a solid basis of research, can we judge any institution not simply by what students need when they enter, but what they have when they leave, as suggested by the president of the case college (Springer, 1997). Here we need some reflection on the philosophy of education as we start addressing the technical issues of instruction as well as specific problems in finance and governance. Probably some bedtime reading of the history of science about those great characters who were one time or another low performers would also help.

Notes

1. It is noted that "Rather than urging that such instruction be shifted to private companies and private colleges, as the Mayor has advocated, the report calls for experiments to stimulate competition" (Arenson, 1999).
2. The improvement of remedial students would be shown even more dramatic if their performance index is calculated by each term rather than based on the cumulative GPA.

References

Adelman, C. (1995), "A's aren't that easy," *New York Times* May 17.

Arenson, Karen W. (1996), "Study details success stories in open admissions at CUNY," *The New York Times*, May 7.

Arenson, Karen W. (1997), "At CUNY, a debate on grade inflation," *New York Times* July 28.

Arenson, Karen W. (1999), "Mayor's Task Force says CUNY is adrift and needs an overhaul," *New York Times* June 8.

Cabrera, Alberto F.(1994), "Logistic regression analysis in higher education: An applied perspective," in Smart, John C. (ed.), *Higher Education: Handbook of Theory and*

Research, Vol.10, Agathon Press, New York. pp. 225-256.

Chen, Sheying (1998). *Mastering Research: A Guide to the Methods of Social and Behavioral Sciences*, Nelson-Hall Publishers, Chicago.

Cheng, David (1997), "Institutional research: In search of powerful tools," *AIR Currents*, 35(1b), Summer.

Cheng, David, & Chen, Sheying (1999), "Factors affecting grading practice," paper presented at the 39th AIR Annual Forum, May 30-June 2, Seattle.

DesJardins, Stephen L., Ahlburg, Dennis A., & McCall, Brian P. (1997), "Using event history methods to model the different modes of student departure from college," paper presented at the 37th Annual Forum of the Association for Institutional Research, May 18-21, Lake Buena Vista, FL.

Editorial (1997), "Educational defeatism," *Staten Island Advance*, August 11.

Gittell, Marilyn, & Holdaway, Jennifer (1996), "Why good students leave CUNY," in Retention Study Committee (n.d.), *Succeeding at CUNY: A Report on Retention and Graduation*, City University of New York.

Lavin, David E., & Crook, David (1990), "Open admissions and its outcome: Ethnic differences in long term educational attainment," *American Journal of Education*, 98(4):389-425.

Lavin, David E., Alba, Richard D., & Silberstein, Richard A. (1981), *Right Versus Privilege: The Open Admissions Experiment at the City University of New York*, The Free Press, New York.

Lavin, David E., & Hyllegard, David (1996), *Changing the Odds: Open Admissions and the Life Chances of the Disadvantaged*, Yale University Press, New Haven.

Lavin, David E., Murtha, James, & Kaufman, Barry (1984), *Long-Term Graduation Rates of Students at the City University of New York*, CUNY Office of Institutional Research and Analysis, New York.

Lavin, David E., et al. (1997), "Graduation rates of CUNY's community college students: The influence of time and transfer," (n.p.).

Lore, Diane C., & Murtha, Kerry (1997), "Most CSI freshmen aren't ready for college," *Staten Island Advance*, August 3.

Noel, L., Levitz, R., & Saluri, D. (eds.), *Increasing Student Retention: Effective Programs and Practices for Reducing the Dropout Rate*, Jossey-Bass, San Francisco.

Retention Study Committee (n.d.), *Succeeding at CUNY: A Report on Retention and* Graduation, City University of New York.

Schmidt, Benno C., et al. (1999), *The City University of New York: An Institution Adrift*. Report of the Mayor's Advisory Task Force on the City University of New York, June 7.

Smodlaka, Ivan (1996), *University Skills Immersion Program: Five Year Report*, City University of New York.

Springer, Marlene (1997), "CSI grads among finest in nation," *Staten Island Advance*, August 18.

Tamada, Mike, & Inman, Claudia (1996), "Survival analysis of faculty retention data: How long do they stay?" paper presented at the California Association for Institutional Research Annual Conference, Costa Mesa, CA.

Tino, Vincent (1993), *Leaving College: Rethinking the Causes and Cures of Student Attrition*, 2nd Ed., The University of Chicago Press, Chicago.

Volpe, Edmond L. (1997), "The CUNY and CSI mission far from fantasy," *Staten Island Advance*, August 20.

Xiao, Beiling (1997), "Using discrete-time survival analysis to examine time to Master of Science degree attainment," paper presented at the 37th Annual Forum of the Association for Institutional Research, May 18-21, Lake Buena Vista, FL.

Part II

Factors Affecting Grading Practices[b]

[b] Two presentations based on this part of study have been made at the 39th Annual Forum of the Association for Institutional Research (AIR), Seattle, May 30-June 2, 1999 and the 25th Anniversary Conference of North East Association for Institutional Research (NEAIR), Philadelphia, November 14-17, 1998. The authors gratefully acknowledge Allen Natowitz for providing assistance in literature search.

Background

Remediation and grading are two related issues concerning the same subject of academic standards in education. The effect of remediation (or any instruction) could be distorted if issues concerning grading are not resolved.

The current research interest in grading was triggered by a mounting concern over grade inflation in the American educational system (Zangenehzadeh, 1988; Summerville et al., 1990; Franklin et al., 1991; Agnew, 1993; Hensley, 1993; Farley, 1995; Arenson, 1997; Yardley, 1997). Although nobody seems to know exactly what "grade inflation" means, a straightforward explanation would point to the increase of average grade over time, specifically the increase in the number of A's and B's and/or the decrease in the amount of D's and F's awarded by an institution (Summerville et al., 1990; Mullen, 1995). In other words, grade inflation happens when "students receive higher grades than their predecessors without a corresponding rise in achievement" (Yardley, 1997). As *USA Today* reports, "since 1987 the portion of students with an A average rose from 28% of test-takers to 37%. But those A students' combined verbal and math scores dropped 14 points at the same time" (Marklein, 1997a).

The focus of concern has been on college and postgraduate education, where the phenomenon has been "widespread if not outright pandemic" (Yardley, 1997). In common sense, this would be more of a problem with less-than-first-class schools, though recent studies revealed that elite institutions face that serious issue as well or even more (Adelman, 1995; Strauss, 1997). For a public university such as the City University of New York (CUNY) that has been frequently bashed for taking in everybody (i.e., open admissions) and wasting taxpayers' money (e.g., remedial education), the suspicion has been intensified and the debate is lasting (cf. Part I; Arenson, 1997). The administration and faculty, accordingly, have been spending extra energy in looking at the potential issue and trying to find out the facts. In one of its senior colleges, for example, grading practices have been a focus of discussion: at the College Personnel and Budget Committee, at meetings of chairpersons in both divisions, and in the departments themselves.[1] System-wide, the CUNY Board Committee on Academic Policy, Program and Research (CAPPR) has been pushing for information regarding patterns of grading and grade distribution as part of its overall pursuit of rigorous standards in CUNY's academic programs.[2]

Nevertheless, national studies show that the stories of grade inflation are probably false accusations. It is discovered that at most schools, there is no grade inflation; contrary to the widespread lamentations, grades actually declined slightly in the last two decades (Adelman, 1995). A 1992 Department of Education survey found no real change in the distribution of letter grades in four-year colleges between 1985 and 1990. And "Although these studies include no data from the last few years, there is no reason to think this trend has changed" (ibid.). At CUNY, a case could also be made that grade inflation is not a problem (CUNY University Faculty Senate Newsletter, 1998), or "the City University has always graded much tougher than other public universities, and certainly than private universities and colleges in the United States" (Arenson, 1997).

Research Problem

Grade inflation seems to have set the tone for most of the studies on grading: first, many researchers have gone after the trend of grading patterns, trying to decide whether grades have indeed increased over time; second, many researchers have focused their attention on the question of whether students have actually learned more to deserve higher grades than their predecessors (e.g., Zangenehzadeh, 1988; Summerville et al., 1990; Franklin et al., 1991; Agnew, 1993; Hensley, 1993; Farley, 1995; Arenson, 1997; Scocca, 1998). As a result, many have provided evidence to have successfully validated (e.g., Summerville et al., 1990; Farley, 1995) or dismissed (e.g., Adelman, 1995; Olsen, 1997) the public suspicion of grade inflation. These research efforts have laid a solid foundation for further studies on this subject.

However, a more careful review of literature has led us to believe that there are at least two conceptual issues that have not been fully addressed. First, grades are measures of educational achievements, but they only make sense on a comparative basis. Comparisons can be made under unified or standardized conditions. The problem is, except for some nationally or internationally standardized tests (e.g., SAT, GRE, and TOEFL) and various state administered professional license examinations, classroom and non-classroom assessments are not standardized. The grading criteria and the factors affecting them would vary from campus to campus, from department to department, and from course to course. Therefore, what the grades tell us apply only to the students who are taught and tested exactly the same way. In a context that an objective standard is absent, it is apparent that the term grade inflation is problematic, and thus it is nearly impossible to determine whether the higher or the lower grades are in fact the grades that are in fact accurate.[3]

Second, if we can assume that the grades were obtained by using some absolutely objective and good criteria, then more low grades would mean worse preparation and learning on the part of students and/or worse teaching performance on the part of faculty. Since no such absolute criteria exist, the common assumption is that not only all classes of students are equally prepared, but also all faculty members are teaching equally well. Therefore, the more low grades given, the more rigorous academic standards seen. In other words, high marks, originally meant to be indicative of educational success, would only be seen as a lamentable tendency for faculty to "inflate" grades. We know, however, that neither of the above assumptions is true, and more high grades *may* indicate better teaching performance (Agnew, 1993). The question is, then, if the faculty should not fake educational success by giving more high grades, should they take pride in educational failure by giving more low grades (what an irony!)? Under such a "wild guess" condition, we consider the issue of grade inflation or deflation unresearchable in the absolute sense. Yet the public is sometimes led too far with various hypotheses by unjustified methodology and results.

In real terms, there are a number of variables which must be considered relating to students, institutions and institutional policies, and even the changeable political climate (Adelman, 1995). It appears that grade inflation, as an instructional practice, will remain a "myth" before we are able to factor in those seemingly countless associates of grades. Instead of asking whether there is grade inflation, or going after the trend of change of grading pattern over time as many researchers have done, it would be more appropriate to ask what are the potential factors that would affect faculty's grading practices. So far, considerable amount of time and energy has been devoted to examining the correlation between student performance and the grades, while students normally do not even participate in this measurement activity known as grading. The present study will not pretend to find the "ultimate" facts or the "absolute" reality of grade inflation, thus adding another piece of testimony to the existing literature dismissing or validating the accusation of grade inflation. Instead, we will focus on some potentially important factors associated with faculty grading practices. Specifically, we will explore some potential factors affecting the process by asking whether grading practices differ by faculty employment status, by faculty rank, by discipline, and/or by course level. The purpose is to provide some necessary knowledge for public understanding and faculty awareness of the problem, and for policy intervention if this is ever deemed desirable.

Research Hypotheses

A logical reasoning would suggest a number of factors that are potentially important in affecting grade distribution, although there have been few empirical studies with conclusive findings. In addition, considerable effort has been spent on institutional research in various universities, particularly in CUNY at both the university level and the college level. This effort offers many useful tips for further research and analyses.

A number of conjectures can be found in institutional research documents that are offered to explain grade distribution in general and suspected grade inflation in particular. Since autonomy is a highly regarded value in higher education and grading has always been considered to be a faculty prerogative,[4] it is natural to probe into faculty grading practice by directly asking how instructors would evaluate students. One hypothesis says, "It is likely that most faculty members operate in this area within a particular personal philosophy (grading on a curve, allowing only a set percentage of As or Bs, etc.) or filter a broader philosophy through a personal framework."[5] The reports of the CUNY colleges that are based on formal and informal surveys and interviews with faculty members, however, conclude that faculty members generally do not grade on a curve but rather mastery of the subject matter and performance of the students. "Experience over time determines faculty judgment of what constitutes mastery of subject matter and, consequently, the assignment of grades according to levels of performance within college grading policies."[6] It is reasonable, therefore, to assume faculty experience or seniority as a potentially important factor affecting grade distribution. But we are not too sure about the direction of this hypothesis, since experience may help prevent grade inflation while the sense of security associated with tenure may also lead to ignoring college grading policies. Just as an elite institution was in a position to overlook media complaints, senior professors might grade students in whatever ways they deemed as appropriate. In contrast, non-elite schools as well as junior faculty would have to carefully watch for, if not simply follow, the tides in the policy space if they were to survive and achieve their desired status.

There is another question as to whether the increased use of adjuncts may affect grading patterns.[7] Specifically, there is a belief that adjuncts grade higher,[8] and this question has engendered a lively debate on CUNYTALK, the on-line forum for CUNY faculty.[9] We would like, therefore, to examine the academic data as to whether there has been a difference between full-time and adjunct faculty in grading practice. The direct examination of such difference is a more valid way of looking at the issue than comparing the ratios of adjunct to full-time faculty at different colleges, since correlating grades with such ratios may simply lead to a mistake called "ecological fallacy" (Babbie, 1998).

On the part of students, increase in high grades may have to do with the pattern of course-taking.[10] It has been suggested that students understand and are adept at "using the system":[11] grading patterns may be skewed when greater numbers of students opt for courses in which grades tend to be higher, or where the grading tends to be more subjective, such as those in the humanities, as opposed to courses in math and science, where the measures are more objective.[12] In other words, grading patterns differ by discipline or department (Summerville et al., 1988; Cluskey et al., 1997). This is the third hypothesis to be tested in the present study.

Within each discipline, different course levels (e.g., lower, upper, and graduate divisions) might have made a difference in grade distribution. Similarly, students in associate degree programs might have had grade distributions different from those in baccalaureate degree programs. The general direction would be higher grades for upper level courses (as opposed to lower level courses) and baccalaureate programs (as opposed to associate degree programs) since the students are supposed to be more acquainted with or better prepared for the learning tasks. There are certainly other reasons, particularly for such courses as internships (Ciofalo, 1988). We will look at these potential differences by testing related hypotheses using empirical data.

There are many other hypotheses that are also worth formulating and testing. The present study, however, focuses on utilizing institutional data that are most adequate for exploring the potential impact of the above basic factors on grading practice. This article will specifically examine the following hypotheses:

(1) Full-time faculty vs. adjunct faculty: Adjunct faculty give higher grades than full-time faculty.

(2) Faculty experience/seniority: Faculty rank makes a difference in grading.

(3) Disciplinary difference: Grades are generally higher in the humanities and social sciences than in science and technology disciplines.

(4) Course levels: The higher the course levels, the higher the average grades.

Methods

Measurement

Although grading practices can be described in many ways (e.g., Riley et al., 1994), this study focuses on *grade distributions* among various groups of students. Multiple measures are often needed to capture such important characteristics as central tendency, dispersion, and skewness of a frequency distribution in a study based on comparison. Which parameter to use has to do with the analytic procedure to be utilized. Some statistical procedures, such as T-Test and ANOVA, automatically take grade averages if the raw data are numerical equivalents of individual grades (i.e., A = 4, A- = 3.7, B+ = 3.3, B = 3, B- = 2.7, C+ = 2.3, C = 2, D = 1, and F = 0). All such aggregate measures, however, are based on the assignment and interpretation of individual grades, which serve not only as indicators of student performance but also as basic units that constitute faculty grading practice.

In the peculiar research context of low grades as "desirable" outcomes, there is a complicating issue in the treatment of grades data, i.e., the meanings of withdrawals (W's) and "Other" grades such as an "Incomplete." This is also a theoretical issue since it is considered a possible factor that may explain the increase in higher grades.[13] A measure taken by some institutions to offset such a potential inflation effect is to treat the original non-penalty grade of W as low as an F. Probably this is suggested by Adelman's (1995) point that the real problem is not grade inflation but withdrawals, incompletes and repeats. As Adelman argues, "The time students lose by withdrawing is time they must recoup. All they have done is increase the cost of school to themselves, their families and, if at a public institution, to taxpayers" (ibid.). And this increasing volume of withdrawals and repeats does not bode well for students' future behavior in the workplace, where repeating tasks is costly. Therefore, it seems justified to consider such grades as punitive as D's and F's.

We should yet ask what is the purpose of grading research. The concern of grade inflation is about faculty functioning as opposed to student performance. Therefore, we should ask what the grades mean to faculty teaching and grading before agreeing that an institution should treat W's as D's or F's in view of the accusations about grade inflation.

The policy of an institution may distinguish between W's (formal withdrawals, which carry no penalty) and WU's (unofficial withdrawals, which carry penalty). However, from a professor's perspective, both kinds of withdrawal could practically be derived from the same situation: they could mean that students were unsatisfied with the faculty teaching, with some signing off officially and others failing to do so on time. In other words, the number or ratio of W's and WU's is possibly an indication of the failure of teaching.[14] The grade of WU is assigned by the faculty and it is possible for a faculty member to use it to penalize a student for unsatisfactory record of attendance. But why the student is absent in the first place may have to do with the effectiveness of faculty teaching. It should be noted that this is not necessarily the case because of various valid excuses such as a medical withdrawal, financial problem, or the student's escaping of rigorous standards. The grade of "I" (incomplete) is somewhat similar to a W or WU in terms of this kind of uncertainty. It may mean that a student could not fulfil the course requirements for various reasons so that the professor had to ask for more work before he/she feels comfortable assigning a grade, or it may mean that a professor simply wanted to give the student a second chance. In any case, joining these three grading events with the low/failing ones can alter the entire distribution, though by doing so there will be a lot of unanswered questions which may lead to the distortion of the real picture.

All in all, it is hardly a sound logic to use W's, WU's, etc. to achieve the effect of grade "deflation,"[15] or to evaluate faculty grading practices. Especially, under no circumstance can W's be considered a good sign of rigorous grading practice on the part of faculty, if this is ever possible for D's or F's, since W's are actually assigned by the students. If D's and F's speak negatively for students but positively for faculty (high standards), W's *may* speak negatively for both. For all the reasons stated above, a discerning study of grading practice, including those that would use grades to adjust student evaluations (e.g., Zangenehzadeh, 1988), cannot treat W's, WU's, and I's equally with D's and F's as "desirable" low grades.

For the reasons above, our approach to grading study is to code W's, WU's, and I's as separate categories in our categorical data analyses. This way we could examine how these special grades are awarded and what are the potential factors affecting them. In other types of analyses that require higher levels of measurement, our tradeoff is to treat these grades as missing values so that they would not obscure the findings with arbitrary determination of their meanings.

Units of Analysis

Since the study deals with students, faculty, and the institution, what is our primary unit of analysis constitutes a good question. Many studies use individual students as units of analysis, since grades are their achievements. The instructor could also be the unit of analysis, since grading has always been considered to be a faculty prerogative.[16] In addition, a course or course section, a program, a discipline, a department, a division, a college/university, or even a country could all serve as a unit of analysis, since grading is supposed to reflect some kind of organizational policy.

However, the problem of using student as primary unit of analysis is that students normally do not participate in, or make decision about, grading themselves (with such exceptions as W's, i.e., withdrawals). Although a grade distribution is among the students, grading practice is about the behavior of faculty. Therefore, the faculty member should serve as our primary unit of analysis.

A frequently seen mistake in research practice is to correlate grade averages of different institutions with their ratios of adjunct to full-time faculty in order to prove or disprove the assertion that adjunct faculty tend to "inflate" grades. But how do we know that high ratios of adjuncts would not drive full-timers to give even higher grades to compete with adjuncts or for some unknown reason to do so? Clearly, the correlation between such ratios and grade averages, which is a college or some other aggregate thing, does not account for the difference between an adjunct and a full-timer, which is a matter at the individual level. To avoid making assertions about individuals based on the examination of an aggregate, we consider the grade averages of various aggregates (e.g., a program or a department) as characterizing the group memberships of individual faculty members, just like their demographics, full-/part-time or tenured/untenured status, etc.

Moreover, we need to distinguish between the unit of analysis and the unit of data collection. The moment a grade is assigned can be considered as a "grading event," which may involve all things that are relevant, such as faculty and institutional characteristics. This event is usually the unit of original data collection and recording. For the purpose of grading practices research, however, a grading event as a unit may appear to be too detailed and may not make sense in a more aggregated form of analysis. The data, therefore, need to be transformed (or manipulated) to facilitate different kinds of analyses involving different groupings of the grading events (e.g., by faculty member, by course section, by division, etc.).

Data Sets

Our research project focused on one college as a case study. Empirical data were obtained from the campus-wide student information system. A working data set was constructed by extracting and combining data from different academic and administrative databases. The two main sources of data were the Course Masters File and the Course Card File. Our research questions and unit of analysis allowed for aggregated forms of data with counts of identical cases. Since our focus was on the faculty and institutional side, we dropped student ID as a variable, which helped create a lot more identical cases. The benefit was the reduction in the size of the database, with counts used as weights in subsequent data analysis.

Designed as a preliminary study of the complex issue, the project was conducted as a cross-sectional study of the various potentially important factors associated with grade distributions within the college. Longitudinal as well as between-college comparative studies were planned as the tasks for later phases of inquiry. The data analyzed in this study covered the Fall semester of 1997.

Analytical Procedures

Analytical procedures used in the study included Cross-Tabulation and T-Test. The techniques of "quasi-multivariate analysis" or elaboration (Chen, 1998) were performed by applying statistical control where a relationship was suspected to be spurious in order to clarify the net effect of a potential causal influence. Multivariate analysis was also conducted in terms of the analysis of variance (ANOVA) and multiple and logistic regression techniques.

It should be noted that these procedures were used for multiple purposes, not simply statistical inference. As a matter of fact, since we included all the students and the faculty who were active on the roll of the Fall semester of 1997, we actually did not need to make any such inference. The inferential results would make sense when the data were supposed to constitute a random sample. Yet in research practice, tests of significance are widely used to analyze nonrandom data, and some argue that significance at least points to the presence of a relatively considerable effect (ibid.). The inferential results included in the following should only be interpreted in such a manner (i.e., for a hypothetical random sample of a larger population).

Results

Collegewide Grade Distribution

Altogether, there were 31,916 grades/grading events recorded for the Fall of 1997 at the college. Table 2.1 breaks these grading events into three distinctive groups: (1) regular grades ranging from A to F, with further breakdowns of high, medium, and low/failing grades; (2) the grades of official and unofficial withdrawals and "incomplete;" and (3) non-judgmental grades, the grades awarded to auditors ("L"), to someone without immunization record as required by the New York State ("WA"), to whom no grade is submitted by the instructor ("Z"), in a course that has passing/failing grades only, and in a course that a grade is assigned at the end of a sequence ("PEN"). Overall, excluding non-judgmental grades, close to 50 percent of the grades awarded in Fall 1997 were on the higher end of the grading spectrum (B and up), nearly one-quarter of the grades, medium (C to B-), and 8 percent, low/failing (D and F).

While this distribution of grades seems to reinforce the notion of "grade inflation" given such a high percentage of high grades, it also points to the fact that how arbitrary and debatable the statistics can be if the grades of withdrawals and incompletes are not sufficiently studied and fully understood (as discussed in our methods section). Table 2.2 presents an alternate way of grouping the grades by which one can effectively argue that there is no grade inflation since over one-fifth of the grades are now in the low range. But there is no assurance that the students who received W's and WU's would necessarily get low grades if they did complete the courses. It is uncertain either with regard to faculty performance since the meaning of D's and F's could be very different from WU's and W's as discussed earlier.

Table 2.1 Grade Distribution and Grouping (1)

Grades and Grouping			Quality Points per Credit	Frequency	Percent
Regular Grades					
High	A		4.0	5,115	16.5%
	A-		3.7	2,916	9.4%
	B+		3.3	3,362	10.8%
	B		3.0	3,970	12.8%
		Subtotal		15,363	49.6%
Medium	B-		2.7	2,267	7.3%
	C+		2.3	1,992	6.4%
	C		2.0	2,918	9.4%
		Subtotal		7,177	23.2%
Low	D		1.0	1,602	5.2%
	F		0.0	878	2.8%
		Subtotal		2,480	8.0%
Grades in Question					
	W - Withdrawal		N/A	2,827	9.1%
	WU - Unofficial Withdrawal		0.0	1,743	5.6%
	I - Incomplete		N/A	1,406	4.5%
		Subtotal		5,976	19.3%
Non-Judgmental Grades					
	P - Passing		N/A	150	N/A
	L - Auditor		N/A	18	N/A
	PEN - Pending		N/A	41	N/A
	AW - Administrative Withdrawal		N/A	106	N/A
	Z - No Grade Submitted		N/A	605	N/A
		Subtotal		920	2.9%
		Total		31,916	

Table 2.2 Grade Distribution and Grouping (2)

Grades and Grouping		Frequency	Percent	Cumulative Percent
Regular Grades:				
High	A, A-, B+, & B	15,363	48.1%	48.1%
Medium	B-, C+, & C	7,177	22.5%	70.6%
Low	D, F, W, & WU	7,050	22.1%	92.7%
Grades in Question:				
Incomplete	I	1,406	4.4%	97.1%
Non-judg.	P, L, PEN, AW, & Z	920	2.9%	100.0%
	Total	31,916		

Bivariate Analysis

Full-Time vs. Part-Time (Adjunct) Faculty. A total of 594 faculty members were involved in grading and included in the study. Of the 594 faculty members, 218 (36.7%) were full-timers, and 376 (63.3%) were adjuncts (part-timers). Full-time faculty were responsible for 15,440 grades/grading events, which account for 46.8% of the total. Adjunct faculty were responsible for 17,544, or 53.2% of the total grades/grading events.

Table 2.3 clearly indicates that, measured by mean quality points per credit, adjunct faculty gave grades 0.107 point higher than full-time faculty. Table 2.4 shows that adjunct faculty gave more high grades than full-time faculty by a margin of 5.4% (a difference of 10.4% in direct comparison of percentages), and they gave fewer low grades than full-time faculty by a margin of 1.2% (a difference of 13.9% in direct comparison of percentages). Row percentages are used in Table 2.4 to facilitate such comparisons. The results seem to render support to our first hypothesis, that is, adjunct faculty give higher grades than full-time faculty.

It is noticeable that while students withdrew officially from full-time faculty's classes at a higher rate than that from adjuncts' (10.3% vs. 8.1% of W's), a higher proportion of students received a grade of unofficial withdrawal (WU) from adjuncts. In addition, full-time faculty seemed to be more willing to give an incomplete grade (5.4%) than adjuncts (3.8%).

Faculty Rank/Seniority. Data were available on the ranks of full-time faculty members while adjuncts carried no formal titles in the database. Of the 218 full-time faculty members, 69 were full professors, 71 associate professors, 67 assistant professors, and 11 under other titles such as lecturers. Senior faculty (full and associate professors) were responsible for 9,116 grades/grading events, which account for 60.2% of the subtotal. Junior faculty (assistant professors and faulty with other titles) were responsible for 6,016, or 39.8% of the subtotal of grades/grading events.

Our second hypothesis says that faculty rank makes a difference in grading. Table 2.3 shows that, measured by mean quality points per credit, there was no significant difference between junior and senior faculty in assigning grades. In Table 2.4, although junior faculty assigned more high grades than senior faculty (48.1% vs. 45.8%), they also assigned more low grades than the latter (9.1% vs. 8.3%). This actually means a greater degree of dispersion, or a greater discriminating power, of junior faculty's grades. In addition, junior faculty assigned or received fewer W's, WU's, and I's than senior faculty (9.6% vs. 10.7%, 4.9% vs. 5.3%, and 5.2% vs. 5.5%, respectively).

Disciplinary Difference. Academic disciplines or departments at the case college are organized in two divisions: the Division of Humanities and Social Sciences (H&SS) and the Division of Science and Technology (S&T). In Fall 1997 19,069 grading events took place in the Division of H&SS and 11,649 in S&T. The T-Test in Table 2.3 points to the fact that, measured by mean quality points per credit, students' grades were 0.113 point higher from the courses taken in the Division of H&SS than those from S&T. Table 2.4 shows that the H&SS Division was responsible for 51.0% of the high grades awarded, whereas S&T, 46.8%. On the other hand, H&SS's low grades accounted for 7.4% of the total, while S&T's accounted for 9.6%. Here our third hypothesis seems to be supported: Grades are higher in the humanities and social sciences than in science and technology disciplines.

It is interesting that, while the faculty in the H&SS Division gave more unofficial withdrawals (WU's) and incomplete grades (I's) than S&T faculty did (6.0% vs.5.1% and 5.2% vs. 3.1%, respectively), the latter received far more W's from the students (7.4% vs. 12.2%). It seems that, though S&T faculty is less likely to "inflate" grades, it might be of greater concern in terms of a need for pedagogical elaborations to help students overcome the difficulties.

Table 2.3 T-Test of Numbered Grades of "A" to "F"

	N	Mean	S.D.	Mean Difference
Part-Time	13,434	2.943	0.979	
Full-Time	11,586	2.836	1.018	0.107*
Senior	6,882	2.836	1.005	
Junior	4,704	2.838	1.037	-0.002
H&SS	15,016	2.930	0.976	
S&T	9,102	2.817	1.055	0.113*

* p<.01.

Table 2.4 Chi-Square Tests of Grade Groups

		Grading Group							
		High	Medium	Low	W	WU	I	Total	Chi-Square*
Part-Time	N	8,528	3,690	1,216	1,327	988	617	16,366	
	%	52.1%	22.5%	7.4%	8.1%	6.0%	3.8%	100%	
Full-Time	N	6,835	3,487	1,264	1,500	755	789	14,630	
	%	46.7%	23.8%	8.6%	10.3%	5.2%	5.4%	100%	159.29**
Senior	N	4,015	2,135	732	940	469	482	8,773	
	%	45.8%	24.3%	8.3%	10.7%	5.3%	5.5%	100%	
Junior	N	2,820	1,352	532	560	286	307	5,857	
	%	48.1%	23.1%	9.1%	9.6%	4.9%	5.2%	100%	15.23
H&S	N	9,409	4,248	1,359	1,365	1,113	968	18,462	
	%	51.0%	23.0%	7.4%	7.4%	6.0%	5.2%	100%	
S&T	N	5,349	2,657	1,096	1,391	581	352	11,426	
	%	46.8%	23.3%	9.6%	12.2%	5.1%	3.1%	100%	328.308**
Lower	N	12,019	6,349	2,309	2,585	1,625	1,047	25,934	
	%	46.3%	24.5%	8.9%	10.0%	6.3%	4.0%	100%	
Upper	N	2,672	794	171	221	110	280	4,248	
	%	62.9%	18.7%	4.0%	5.2%	2.6%	6.6%	100%	592.083**

* DF=5 ** p< .0001.

Course Levels. Given the fact that the college academic offerings range from associate degree programs all the way to the Masters, the frequencies of grades/grading events by course level are pyramidal: the higher the course level, the fewer the students/grades. Does course level affect faculty grading practice?

Table 2.5 displays an unambiguous pattern: the higher the course level, the higher the average grade, which is exactly our fourth hypothesis. This finding is consistent across both undergraduate (100-level to 500-level) and graduate (600-level and above) courses. The least-significant-difference (LSD) multiple range test was conducted through the One-Way ANOVA procedure to see how different the mean grades were from each other. With significance level set at 0.05, the results showed that grades in all the 600-level and above courses were higher than all the 400-level and lower courses. In other words, average grades in graduate courses were higher than those in undergraduate courses except independent study, internship and special topics courses for undergraduates at the 500 level. Focusing on undergraduate-level courses, a Chi-Square test was performed and the result (see Table 2.4) confirmed a significant grading difference between lower division courses (100- and 200-levels) and upper division courses (300- and 400-levels; 500-level courses are excluded for a more rigorous test). Upper division instructors gave out 62.9% high grades, as opposed to the lower division, 46.3%. Meanwhile, upper division instructors gave less than one-half low grades as compared with lower division instructors (4.0% vs. 8.9%). What is especially intriguing was that while upper division instructors seemed to be more prepared to give out incomplete grades (6.6% vs. 4.0%), they assigned or received by far the fewer WU's and W's (5.2% vs. 10.0% and 2.6% vs. 6.3%, respectively). This suggests an important difference between incompletes and withdrawals.

Elaboration

Table 2.4 suggests that faculty graded differently by full-time and part-time status, discipline, and course level, while there was no significant difference between senior and junior faculty. We should, however, put the tests under more controlled conditions to make sure that the differences found were not spurious. Each hypothesis can be tested by taking into consideration additional variables in each analysis which might be responsible for the potentially spurious differences in grading practice. The logic is that if the said differences disappear or weaken after controlling for the other variables, then the differences may be to some degree spurious. If the differences stay unchanged after controlling for the other variables, then they are probably true or non-spurious (Chen, 1998).

Table 2.5 Difference of Course Levels in Grading Practice

Course Level	N	Mean	S.D.
100-Level	13,652	2.736	1.076
200-Level	7,025	3.006	0.879
300-Level	2,658	3.125	0.856
400-Level	979	3.188	0.798
500-Level	125	3.689	0.498
600-Level	480	3.469	0.584
700-Level	84	3.607	0.560
800-Level	17	3.706	0.588
Total	25,020	2.894	0.999

For the categorical data presented in Table 2.4, statistical control can be carried out via a partial- or sub-table approach. Tables 2.6 and 2.7 present some results of the elaboration. For the full-/part-timer difference, a consistent pattern of grading practices under all the conditions controlling for disciplinary difference and course levels suggests that the results of the bivariate analysis presented earlier are probably true (i.e., nonspurious). However, the conclusion regarding the difference between senior and junior faculty in grading practice can be partly attributed to the disciplinary difference because the finding is reversed for grades awarded in the S&T (Science and Technology) Division. That is, junior faculty in the S&T Division, but not in the H&SS (Humanities and Social Sciences) Division, tended to hand out more high grades than senior faculty, which was especially true in the upper level courses. In contrast, senior faculty in the H&SS Division tended to give more high or medium grades than junior faculty particularly in the lower level courses. These findings render our analysis a specification model, which can also be called a conditional analysis (Chen, 1998).

Table 2.6 Grade by Faculty Full-/Part-Time Status Controlling for Course Level and Discipline

Course Level	Discipline	Grade	Faculty Status Part-Time	Faculty Status Full-Time	Chi-Square
Lower	S&T	High	60%	51%	
		Medium	30%	32%	
		Low	10%	17%	82.967*
	H&SS	High	62%	55%	
		Medium	28%	34%	
		Low	10%	11%	59.220*
Upper	S&T	High	77%	66%	
		Medium	19%	25%	
		Low	4%	8%	17.306*
	H&SS	High	84%	70%	
		Medium	15%	26%	
		Low	1%	5%	61.388*

* p< .01

Table 2.7 Grade by Faculty Junior/Senior Status Controlling for Course Level and Discipline

Course Level	Discipline	Grade	Faculty Status Junior	Faculty Status Senior	Chi-Square
Lower	S&T	High	53%	50%	
		Medium	1%	33%	
		Low	16%	17%	2.221
	H&SS	High	53%	57%	
		Medium	34%	34%	
		Low	13%	10%	21.850*
Upper	S&T	High	75%	62%	
		Medium	15%	30%	
		Low	9%	8%	26.841*
	H&SS	High	70%	69%	
		Medium	24%	27%	
		Low	6%	4%	2.503

* p< .01

Multivariate Analysis

The ANOVA procedure was used to provide a more comprehensive understanding through multivariate analysis. Since ANOVA requires listwise deletion, we combined the two variables of faculty seniority and full-time/part-time status to avoid the potential problem of too many missing cases. As a matter of fact, the original variable "title" in the college administrative database equated the category of "no title" with the category of part-time (adjunct) faculty. This categorical variable can be used in the ANOVA procedure to show the influence of different faculty employment status. To provide a more detailed comparison, multiple and logistic regression techniques are also used.

Tables 2.8, 2.9, and 2.10 contain the results of ANOVA and regression analyses. The results are consistent in that all types of multivariate analysis reconfirmed the influences of course level, faculty status, and disciplinary differences on faculty grading practice, and course level had the highest impact among the variables examined. The multiple regression results further confirmed the impact of course level and that adjunct faculty graded higher on average than full-time faculty, whereas faculty rank had the least influence with junior faculty tended to grade slightly lower than other faculty. The linear model, however, does not seem to fit the data well as indicated by the adjusted R square. The logistic regression focused on both low (i.e., F and D) and high (i.e., B through A) grades while omitted the middle grades (i.e., C through B-). This treatment greatly amplified the difference in grading practice. With the Forward Stepwise (LR) technique independent variables entered the equation in the following order: course level, full-/part-time status, discipline, and senior/junior status. Table 2.10 demonstrates that higher course levels were most closely related to higher grades, and the next was adjunct status. It also shows that if our focus is on high and low grades (excluding middle grades), then senior faculty would give more high grades and/or fewer low grades than junior faculty. This difference even surpassed the influence of different disciplines in terms of both the odds ratio and the B values ($p < .001$). The difference demonstrated under this approach, however, would vanish when middle grades were counted in because senior faculty might assign more lower middle grades while junior faculty more upper middle grades.

Table 2.8 Factors Affecting Grading Practice: ANOVA Results

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig. of F
Main Effects	876.273	7	125.182	127.766	0.000
DIVISION	59.538	1	59.538	60.767	0.000
LEVEL	730.222	4	182.556	186.325	0.000
SEN/JUN/ADJ.STATUS	201.702	2	100.851	102.933	0.000
Explained	876.273	7	125.182	127.766	0.000
Residual	23622.302	24110	0.980		
Total	24498.576	24117	1.016		

Table 2.9 Factors Affecting Grading Practice: Multiple Regression Results

Variable	B	SE B	Beta	T	Sig T
Division	0.102557	0.013378	0.048913	7.666	0.0000
FTPT	-0.196975	0.017771	-0.096487	-11.084	0.0000
LEVEL	0.211536	0.007621	0.180830	27.755	0.0000
SR/JR/ADJ	-0.042906	0.019371	0.018998	2.215	0.0268
(Constant)	2.548007	0.015465		164.756	0.0000

Multiple R	0.19157	R Square	0.03670
Standard Error	0.99988	Adjusted R Square	0.03654

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	4	923.56812	230.89203
Residual	24249	24243.40817	.99977

F = 230.94529 Signif F = .0000

Coding: Division - 1=HSS, 0=S&T; FTPT - 1=ft, 0=pt; SR/JR/ADJ - 1=junior, 0=other.

Table 2.10 Factors Affecting Grading Practice: Logistic Regression Results

GRADE (Dependent Variable Encoding: High —> 1, Low —>0):

Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
Division							
S&T	-0.1976	0.0460	18.4770	1	0.0000	-0.0344	0.8207
(HSS)							
SR/JR/ADJ							
Senior	0.2181	0.0655	11.0997	1	0.0009	0.0255	1.2438
(Junior)							
FTPT							
FT	-0.5489	0.0531	106.7046	1	0.0000	-0.0867	0.5776
(PT/ADJ)							
LEVEL	0.8283	0.0369	504.1845	1	0.0000	0.1898	2.2894
Constant	0.6757	0.0884	58.3773	1	0.0000		

Discussion

College administrators often find themselves caught in a dilemma when their institution is being accused of grade inflation, especially when "hard" data over time seem to support the accusation. On the one hand, since grading is always a faculty prerogative, the administration is supposed not just to refrain from interfering faculty practices but to defend this basic academic freedom. On the other hand, institutions, especially the public ones, are increasingly held accountable for their performance and outcomes, and nothing serves as a more negative indication of an institution's lack of academic standards than grade inflation. Therefore, to college administrators, this is not a matter of whether to intervene with faculty grading or not; it's a matter of how.

To simply compile data or to go after the trend of change in grading patterns over time, as most researchers have done so far, does not help confirm or dismiss the accusation of grade inflation. It is our belief that the judgment of whether there is grade inflation is more of a normative or political issue than an academic or scientific one. In other words, it is the lack of standardized criteria in classroom grading that makes it impossible to speak about grade inflation in any absolute terms. In the last analysis, to understand the potential factors contributing to the variation in grade distribution becomes a prerequisite for any effective policy intervention, currently represented by a desire to keep grades in check or to achieve grade deflation (Agnew, 1993).

Our approach is to identify areas of attention without confirming or dismissing the accusation. The results would provide administrators with specific and in-depth knowledge about faculty grading practices. The findings here suggest that greater attention should be paid to upper level courses, courses offered in the humanities and the social sciences, and part-time faculty grading practice. Faculty rank is not a general concern, though it does make some difference in the details of a grading pattern.

The present study had certain limitations. Chief among them was that the data did not include student information as well as more detailed characteristics of the faculty. Our future endeavor will try to explore the hypothesis that the possible increase in high grades has to do with admissions criteria, or improved preparation of entering students (Mullen, 1995). It is further suggested that high school percentile rank and ACT Composite Scores may account for individual differences among freshmen (ibid.). There is evidence that students who are now entering CUNY with more CPI units are better prepared.[17] The increase in the proportion of transfer students who are historically stronger performers may also count. Another possible factor is that the advent of technology in the classroom and at home, such as word processing, is helping students do better work and thus obtain higher grades.[18] A more detailed look at the grading process may take into account the numbers of grade changes as a result of appeals by students, faculty's allowing students to redo their work or taking a particular course more than once, students' maneuver for better grades (Wiesenfeld, 1996, Zangenehzadeh, 1988; Franklin et al., 1991), government eligibility requirements for certain benefits, and such technical questions as telecourses vs. traditional courses (Searcy et al., 1993) and class size (Franklin et al., 1991). These variables may be included in future survey and other research designs aimed at collecting more detailed data. In those designs we can consider even more factors such as gender, race/ethnicity (Cross, 1993), English proficiency, and credits completed that may potentially influence grade distribution.

Notes

1. Springer, M. (1998), "Grade distribution at the College of Staten Island," memo to CUNY Interim Chancellor, February 5.
2. Kimmich, C. M. (1998), "Report on grade distribution," memo from CUNY Interim Chancellor to College Presidents, January 8.
3. There are institutions and researchers who have tried to address this issue by using measures other than traditional letter grades or GPA, standardized test scores, faculty consensus, or student input (e.g., Marklein, 1997b; Johnson, 1997; Prince, 1997; Cluskey et al., 1997; Duckwall, & Wilson, 1996; Farley, 1995; Dreyfuss, 1993). These approaches, which might be challenged for potential biases, subjectivity or other problems (e.g., Shepard, 1989), will need a more solid knowledge base, particularly an understanding of important facts associated with grading practice.
4. See Note 2.
5. See Note 2.
6. Mirrer, L. (1998), "Grade distribution," memo from CUNY Vice Chancellor for Academic Affairs to Members of CAPPR, February 19.
7. See Note 6.
8. Cheng, D., Hartman, J., Podell, D., & Zeldin, M. (1998), "Grading report," memo to CSI Vice President for Academic Affairs, January 16.
9. Hartman, J. (1998), "Poll of chairpersons," memo to CSI Dean of Science and Technology, January 21.
10. See Note 2.
11. See Note 2.
12. See Note 6.
13. See Note 6.
14. Balfe, J. (1998), "Relation of student evaluations and grades," memo to PSAS faculty at the College of Staten Island/CUNY, February 9.
15. This has been encouraged in many colleges and universities (Agnew, 1993) and linked seriously with decision making in tenure and promotion (Zangenehzadeh, 1988).
16. See Note 2.
17. See Note 6.
18. See Note 6.

References

Adelman, C. (1995). A's aren't that easy. *New York Times* May 17.
Agnew, E. (1993). Departmental grade quotas: the silent saboteur. Paper presented at the Annual Meeting of the Conference on College Composition and Communication (44th, San Diego, CA, March 31-April 3).
Arenson, K. W. (1997). At CUNY, a debate on grade inflation. *New York Times* July 28.
Babbie, Earl (1998), *The Practice of Social Research*, 8th Ed., Wadsworth Publishing Company, Belmont, CA.
Chen, S. (1998). *Mastering Research: A Guide to the Methods of Social and Behavioral*

Sciences. Chicago: Nelson-Hall Publishers.
Cheng, David, & Chen, Sheying (1999). A study of the outcome of remedial education. Paper presented at the 39th AIR Annual Forum, May 30-June 2, Seattle.
Ciofalo, A. (1988). A perspective on internship grading problems and a solution. Writing/Media Department, Loyola College, MD, ERIC_NO-ED298557.
Cluskey, G. R., Jr., et al. (1997). Accounting grade inflation. *Journal of Education for Business* 72(5): 273-277.
Cross, T. L. (1993). On scapegoating blacks for grade inflation. *Journal of Blacks in Higher Education* Fall(1):47-56.
CUNY University Faculty Senate (1998), Faculty earn A+ for objective grading. *University Faculty Senate Newsletter* Spring XXVIII(2):5.
Dreyfuss, S. (1993). My fight against grade inflation: a response to William Cole. *College Teaching* 41(4) 149-152.
Duckwall, J. M., & Wilson, S. J. (1996). Can core components in mathematics courses replace comprehensive common course finals? Paper presented at the Annual Forum of the Association for Institutional Research (36th, Albuquerque, NM, May 5-8).
Ekstrom, Ruth, & Villegas, Ana Maria (1994). *Study of Policies and Practices* (Report No. 94-1), The College Board (ETS), New York, 1994.
Farley, B. L. (1995). "A" is for average: the grading crisis in today's colleges. In: *Issues of Education at Community Colleges: Essays by Fellows in the Mid-Career Fellowship Program at Princeton University*. Mid-Career Fellowship Program, Princeton University, NJ.
Franklin, J., et al. (1991). Grade inflation and student ratings: a closer look. Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7).
Hensley, W. E. (1993). Systems theory in the ivory tower. Paper presented at the Annual Convention of the Speech Communication Association (Miami, FL, November 18-21).
Johnson, V. E. (1997). An alternative to traditional GPA for evaluating student performance. *Statistical Science*, 12(4): 251-269.
Marklein, M. B. (1997a). SAT scores up, but so is grade inflation. *USA Today* August 27.
Marklein, M. B. (1997b). Duke plan could deflate high grades. *USA Today* March 13.
Mullen, R. (1995). Indicators of grade inflation. Paper presented at the Annual Forum of the Association for Institutional Research (35th, Boston, MA, May 28-31).
Prince, G. S., Jr. (1997). Personal attention counts. *USA Today* March 25.
Riley, Herbert, et al. (1994). Current trends in grades and grading practices in undergraduate higher education (results of the 1992 AACRAO survey). ERIC_NO-ED378893.
Scocca, T. (1998). The great grade-inflation lie. *The Boston Phoenix*, April 23-30.
Searcy, R. D., et al. (1993). Grade distribution study: telecourses vs. traditional courses. ERIC_NO-ED362251.
Shepard, L. A. (1989). Inflated test score gains: is it old norms or teaching the test? Effects of testing project. Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, March 27-31).
Strauss, V. (1997). Colleges seek to slow grade inflation. *Washington Post* June 12.
Summerville, R. M., et al. (1988). Grade inflation in the eighties: the case of urban colleges and universities. Paper presented at the Annual Meeting of the American Educational Research Association (28th, Phoenix, AZ, May 15-18).

Summerville, R. M., et al. (1990). Grade inflation: the case of urban colleges and universities. *College Teaching* 38(1) 33-38.
Wiesenfeld, K. (1996). Making the grade. *Newsweek*, June 17.
Yardley, J. (1997). The high cost of grade inflation. *Washington Post* June 16.
Zangenehzadeh, H. (1988). Grade inflation: a way out. *Journal of Economic Education* Summer 19(3): 217-230.

033 031

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

0 31024

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Remedial Education and Grading: A Case Study Approach to Two Critical Issues in American Higher Education

Author(s): Sheying Chen and David X. Cheng

Corporate Source: City University of New York

Publication Date: June 1999

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1	2A	2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature: [signature]	Printed Name/Position/Title: David X. Cheng, Assistant Dean for Research and Planning
Organization/Address: Columbia University, 407 Lerner Hall, MC2611 2920 Broadway, New York, NY 10027	Telephone: (212)854-1225 FAX: (212)854-5353
	E-Mail Address: dxc1@Columbia.edu Date: 4/21/2000

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com