

A STUDY OF THE STUDENT PLACEMENT SYSTEM
FOR HIGHER EDUCATION INSTITUTIONS IN TURKEY

116187

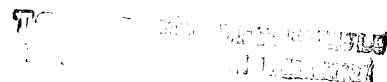
A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY
ÖZGÜR YAZLALI

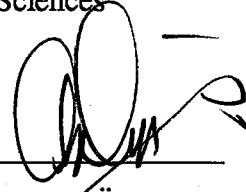
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF INDUSTRIAL ENGINEERING

JULY 2001

116187



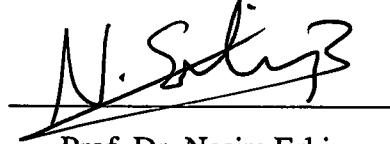
Approval of the Graduate School of Natural and Applied Sciences



Prof. Dr. Tayfur Öztürk

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.



Prof. Dr. Nesim Erkip

Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.



Assoc. Prof. Dr. Gülser Köksal

Co-Supervisor



Prof. Dr. Murat Köksalan

Supervisor

Examining Committee Members

Assoc. Prof. Dr. Gülser Köksal



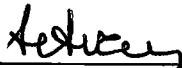
Prof. Dr. Murat Köksalan



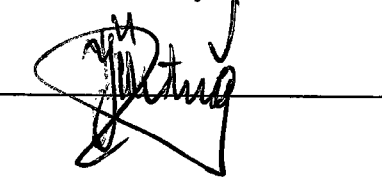
Assoc. Prof. Dr. Tayyar Şen



Dr. Fethi Toker



Prof. Dr. Yıldırım Üçtuğ



ABSTRACT

A STUDY OF THE STUDENT PLACEMENT SYSTEM FOR HIGHER EDUCATION INSTITUTIONS IN TURKEY

Yazlalı, Özgür

M.S., Industrial Engineering

Supervisor: Prof. Dr. Murat Köksalan

Co-Supervisor: Assoc. Prof. Dr. Gülser Köksal

July 2001, 99 pages

In this study, the central placement problem of Student Selection and Placement Center (ÖSYM) of Turkey is analyzed in terms of the appropriateness of its assessment criteria to the intended purposes. The specific concentration is on the assessment of past academic achievement. The two sub-criteria used to weight past achievement: schools' average performance at the central exam (A), and the standardized grade average of the least successful student at each school (C) are analyzed, and elimination of C is proposed and investigated. Another focus of the study is the policy of ÖSYM in dealing with the "outlier" students in their school's grade point average distribution. In this respect, currently used policy is compared with some alternative rules in the literature of outlier detection. The results of the alternative outlier detection methods do not differ enough to have a statistically reliable conclusion in favor of one of them.

Keywords: Student Placement System, Outlier Detection

ÖZ

TÜRKİYE YÜKSEKÖĞRETİM KURUMLARINA ÖĞRENCİ YERLEŞTİRME SİSTEMİ ÜZERİNE BİR ÇALIŞMA

Yazları, Özgür

Yüksek Lisans, Endüstri Mühendisliğı

Tez Yöneticisi: Prof. Dr. Murat Köksalan

Ortak Tez Yöneticisi: Doç. Dr. Gülser Köksal

Temmuz 2001, 99 sayfa

Öğrenci Seçme ve Yerleştirme Merkezi'nin (ÖSYM) genel yerleştirme problemi, değerlendirme kriterlerinin kullanım amaçlarına uygunluğu bakımından incelendi. Özellikle, geçmiş akademik başarının değerlendirilmesinde ağırlıklandırma amacıyla kullanılan iki kriter: okulların ortalama merkezi sınav performansı (A) ve her okuldaki en düşük öğrencinin standart not ortalaması (C) incelendi. Buradaki değerlerin yol açtığı sorunlar ve C'nin kullanılmaması durumu araştırıldı. Çalışmanın diğeri bir odak noktasını, ÖSYM'nin okullardaki not ortalamaları dağılımında, varsa, çok uçlarda kalan öğrencileri belirlemek için kullandığı yöntem oluşturdu. Bu bağlamda halen kullanılan yöntem, literatürdeki bazı alternatif yöntemler ile karşılaştırıldı. Örnek okullar üzerinde yapılan testlerin sonucunda, herhangi bir alternatif yöntem diğeri alternatiflere göre daha avantajlı görülmedi.

Anahtar Kelimeler: Öğrenci Yerleştirme Sistemi, Uç Gözlemlerin Belirlenmesi

ACKNOWLEDGEMENTS

I am grateful to Professor Murat Köksalan for his guidance both here and across the Atlantic Ocean. Professor Gülser Köksal's support and insight has been valuable especially in the second year of this research. I would like to thank both of them for their contribution to this work which has given me the chance to deal with a real life process that I have gone through myself as well together with millions of young people in my country.

I am obliged to mention that this work could not be realized if Professor Yıldırım Üçtuğ, Dr. Fethi Toker and the staff of ÖSYM were not always there, ready to help.

I would like to thank my mother Nuray, and my father Naci for their continuous and unrequited support throughout my life.

And finally, I am indebted to my brother Yener, my friends Mustafa and Canan for their long-lasting patience and tolerance in times I was unbearable.

TABLE OF CONTENTS

ABSTRACT.....	iii
ÖZ.....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS	vi
LIST OF TABLES	viii
LIST OF FIGURES.....	xii
CHAPTERS	
1. INTRODUCTION.....	1
1.1. A Brief History of Student Selection and Placement System in Turkey.....	2
1.2. Current Problem Environment	4
2. PROBLEM DEFINITION AND LITERATURE SURVEY.....	7
2.1. Educational Assessment and Central Placement.....	7
2.2. Previous and Current Systems.....	12
2.2.1. Standardization Procedure.....	16
2.2.2. Previous System (1981 to 1998).....	16
2.2.3. Current System (1999 to Present)	18
2.2.3.1.Student Selection Examination (ÖSS).....	19
2.2.3.2.Secondary Education Academic Achievement Score (OBP).....	21
2.2.3.3.Weighted Secondary Education Academic Achievement Score (AOBP)	24
2.2.3.4.Placement Scores (YÖSS)	31
2.3. Outlier Detection Rules	31
2.3.1. GESD Rule	33
2.3.2. Tukey's Rule.....	34

2.3.3. Median Rule.....	36
2.3.4. Discussion of the Rules	37
2.4. Summary of Potential Problems.....	41
3. ANALYSIS OF REAL DATA	42
3.1. Sampling Methodology.....	42
3.2. Analysis of Main School Categories	44
3.3. Problem I: Outlier Detection.....	49
3.4. Problem II: Placement Criteria and Weights	54
3.4.1. Elimination of ÖYS.....	55
3.4.2. Weighting the Past Academic Achievement	56
3.4.3. Fixing the AOBP Upper Limit to 80 for Every School.....	62
4. COMPARISON OF SOME ALTERNATIVES	65
4.1. Alternative I: Outlier Detection Method of ÖSYM	67
4.2. Alternative II: Outlier Detection Methods with Trimming.....	69
5. CONCLUSIONS AND FURTHER RESEARCH	72
REFERENCES.....	77
APPENDICES	
A. DERIVATION OF AOBP FORMULATION	81
B. FORMULA FOR TUKEY’S RULE	83
C. SUMMARY STATISTICS FOR SAMPLE SCHOOLS.....	84
D. DISTRIBUTION FITTING RESULTS	87
E. COMPARISON OF OUTLIER DETECTION RULES.....	89
F. LOWER-OUTLIERS ACCORDING TO ÖSYM	90
G. COEFFICIENT OF CORRELATION ESTIMATES	91
H. REAL AND NET DIFFERENCES FOR ALTERNATIVES.....	93
I. EFFECTS OF AN INDIVIDUAL STUDENT FOR THE PROPOSED METHOD	99

LIST OF TABLES

TABLE

1.1	Whole Population (all ages), University Entrant Population (ages 15 and 29), and Number of Central Placement Applicants between years 1980 and 2000.....	5
2.1	Comparisons of weights of different placement criteria for an average student performance in an average school in the previous and current systems	13
2.2	Number of students graduated from Science High Schools between 1989 and 1999.....	15
2.3	Coefficients for composite ÖSS scores of the previous system (c_i^k).....	17
2.4	The Coefficients of the ÖYS standard test scores ($c_{i,l}^{\hat{OYS}}$), composite ÖSS scores ($c_{i,k}^{\hat{OSS}}$) and OBP (c^{OBP}) for the placement scores (YÖYS) in the previous system	18
2.5	The coefficients of the standard test scores (c_i^k), for the composite ÖSS scores	20
2.6	Effects of Changing School on OBP.....	22
2.7	The behavior of ÖSYM's outlier detection method for three continuous probability distributions.....	39
3.1	The summary of initial data according to the categories.....	42
3.2	Descriptions of school categories, and comparison of sample schools in these categories (year 2000 data) with the whole Turkey population (year 2001 data)	44
3.3	Summary statistics for 12 school categories and the whole sample	45

3.4	Kolmogorov-Smirnov goodness of fit test results for 12 school categories and the whole sample (for $\alpha = 0.05$ case)	48
3.5	Correlation of coefficient estimates between school parameters for 12 school categories and the whole sample population	49
3.6	Average of p values (\bar{p}_i) and error rates for ÖSYM's method and 2 alternative methods	51
3.7	p_n values and error rates for three alternative outlier detection methods	53
3.8	The changes in the weights of placement criteria and their corresponding aims	55
3.9	CGPA's needed to have equal OBP's in four sample schools in Ankara	57
3.10	Change of OBP's and AOBP's in an example school after the transfer of a student.....	58
3.11	Statistics for the size of schools (n) (average, minimum, and maximum school size values) and histogram summaries for the whole sample in school category <i>Overall</i> ₂ , categorized according to the existence of lower-outliers by ÖSYM's method.....	59
3.12	Coefficient of correlation (r) estimates between students' CGPA and ÖSS scores in Equal Weighted category summarized for 12 categories and the whole sample ($CGPA_i$ vs. $\bar{ÖSS}_i^{EW}$).....	61
3.13	CGPA, OBP, AOBP Comparisons for Two Sample Schools	62
3.14	Average, minimum and maximum values and histogram summaries of arithmetic mean of ÖSS scores of the students in the upper 5% of their school's CGPA distributions ($\mu_{95\%}$), for 12 school categories and the whole sample	64
4.1	Average, minimum and maximum values, and histogram summaries of the net difference (Δ) of students according to the first alternative method, summarized for 12 school categories and the whole sample	68
4.2	The behavior of the difference function (4.1).....	70

4.3	Averages of net difference (Δ) of students for the second alternative method only for the Tukey's case, summarized for 12 school categories and the whole sample.....	71
C.1	Average, minimum and maximum values, and histogram summaries of arithmetic mean CGPA (\bar{X}_{CGPA}) for 12 school categories an the whole sample	84
C.2	Average, minimum and maximum values, and histogram summaries of standard deviation CGPA (s_{CGPA}) for 12 school categories an the whole sample	85
C.3	Average, minimum and maximum values, and histogram summaries of sample skewness (v) for 12 school categories an the whole sample.....	86
C.4	Average, minimum and maximum values, and histogram summaries of ÖSS score average (A^{EW}) for 12 school categories an the whole sample.....	86
D.1	Distribution fitting results (for $\alpha=0.05$ case) in four separate categories for the 12 school categories and the whole sample in numbers of schools and numbers of students	87
D.2	Distribution fitting results (for $\alpha=0.01$ case) in four separate categories for the 12 school categories and the whole sample in numbers of schools and numbers of students	88
E.1	Average of p values (\bar{p}_i) and error rates for ÖSYM's method and two alternative outlier detection rules for the transformed data.....	89
F.1	Statistics for the size of schools (n) (average, minimum, and maximum school size values) and histogram summaries for the 12 school categories and the whole sample, categorized according to the existence of lower-outliers by ÖSYM's method	90
G.1	Coefficient of correlation (r) estimates between students' CGPA and ÖSS scores in Quantitative category summarized for 12 categories and the whole sample ($CGPA_i$ vs. $\dot{O}SS_i^Q$)	91
G.2	Coefficient of correlation (r) estimates between students' CGPA and ÖSS scores in Verbal category summarized for 12 categories and the whole sample ($CGPA_i$ vs. $\dot{O}SS_i^V$)	92

H.1	Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the first alternative method, summarized for 12 school categories and the whole sample	93
H.2	Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the second alternative method (outlier detection according to Tukey's Rule), summarized for 12 school categories and the whole sample.....	94
H.3	Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the second alternative method (outlier detection according to Median Rule), summarized for 12 school categories and the whole sample.....	95
H.4	Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the second alternative method (outlier detection according to GESD Rule), summarized for 12 school categories and the whole sample.....	96
H.5	Average, minimum and maximum values, and histogram summaries of the net difference (Δ) of students according to the second alternative method (outlier detection according to Median Rule), summarized for 12 school categories and the whole sample.....	97
H.6	Average, minimum and maximum values, and histogram summaries of the net difference (Δ) of students according to the second alternative method (outlier detection according to GESD Rule), summarized for 12 school categories and the whole sample.....	98
I.1	The effects of an individual student for the first proposed alternative.....	99

LIST OF FIGURES

FIGURE

1.1	Number of applicants and placements between 1980 and 2000.....	4
1.2	The ratio of number of placements to number of applicants according to graduation status between 1980 and 2000.....	5
2.1	Changes in the relative weights of placement criteria between 1980 and 2000	14
2.1.a	α lower limits for two hypothetical school distributions	26
2.1.b	Determination of AOBP of a student in the $[\alpha, 80]$ range	26
4.1	Histogram summaries of placement scores of students in the sample, and the placement scores average for 12 school categories and the whole sample	67

CHAPTER 1

INTRODUCTION

In Turkey, access to a higher education institution is only through a central placement (YÖK, 1999). Every year almost one and a half million people try to access higher education via the central placement system of Student Selection and Placement Center (ÖSYM) of Higher Education Council (YÖK). Placement of these applicants is performed considering their secondary education academic achievement and their scores on a central examination. Indeed the current placement system can be considered as a bicriteria decision making problem, where these two criteria, success in the national test and past academic achievement, are made up of some other multiple criteria.

The aim of this study is to analyze various factors of the placement process, and their effects on different school categories. Although we analyze the whole system of central placement in general, our specific focus is on the assessment of past academic achievement. One of the areas we concentrate on is the policy of ÖSYM in dealing with the extreme students in their school's grade point average distribution, so that the effect of these extreme students on the rest of the population is minimized. In this respect, performance of ÖSYM's policy is compared with some alternative rules that exist in the literature of outlier detection. We also study the choice of placement criteria, and the way they are weighted. Especially the current system of ÖSYM tries to overcome some of the problems of the previous system on assessing the past

academic achievement, where the focus of the analyses is the appropriateness of the chosen criteria to the intended purposes of ÖSYM in changing the system.

This thesis study consists of 5 chapters. In the first chapter, the central placement system is introduced with a brief history, and the problem environment is described. The second chapter describes the previous and current placement systems in detail, and discusses characteristics of both systems by comparisons with appropriate theoretical background including a presentation on outlier detection rules. The third chapter examines potential problems that have been identified, using some sample school data obtained from ÖSYM. In the fourth chapter, some alternative methods for the assessment of academic achievement and their effects on the central placement system are presented. And the last chapter concludes the report with a brief review of findings, and states some points that need further research.

1.1 A Brief History of Student Selection and Placement System in Turkey

Until 1950's student selection for higher education was not considered to be a problem, since the capacity of the programs exceeded the demand of the applicants. However, with an enormous growth in the student population thereafter, the universities began to deal with increasing numbers of students. To overcome the difficulties aroused, institutions began to apply their own entrance examinations, which consisted of mostly essay type questions. In 1963, considering the difficulty of assessment in those entrance examinations, The Interuniversity Board formed The Interuniversity Entrance Examination Commission to look into the feasibility of enlarging the student selection system of Ankara University to all other universities. By the year 1964, most of the universities in Turkey began to accept students according to the results of the central examination. During this period from 1964 through 1973 the placement system can be regarded as partially centralized in terms of the selection criteria: the students were applying individually to the institutions they desire, and the admissions decisions were made by the university registrars.

By mid 70's, dealing with an ever increasing number of applications had already been a burden for universities. In 1973, considering the huge demand for higher

education, The Interuniversity Board had chosen to apply a fully centralized system in terms of both selection and placement. Between 1974 and 1980, a single-stage examination system was used; applicants were ranked based on the scores they took in Interuniversity Selection Examination (ÜSS). The Interuniversity Student Selection and Placement Center (ÜSYM) were using the scores of ÜSS in the placement of the applicants to the institutions in their preference lists. After 1981, Student Selection and Placement Center (ÖSYM-ÜSYM had been changed to ÖSYM according to the new Higher Education Law introduced in year 1981) had begun to apply a two-stage system. In this system Student Selection Examination (ÖSS) was given to the applicants first, aiming to eliminate some of the applicants who lack the necessary skills needed for higher education. At the second stage, Student Placement Examination (ÖYS), where the success is highly dependent on the candidate's knowledge, was given. The second examination was very similar to the one (ÜSS) used in the one-stage system (ÜSYM, 1980).

In 1999, ÖSYM again turned back to a single stage system, however this time only applying the first-stage examination of the previous system, ÖSS. The reason behind this change, as stated by ÖSYM, was the high correlation between the two exam results, which is almost 100% (YÖK, 1999). In addition to using the previous system's ÖSS as the main factor in the placement process, ÖSYM also made some other changes. One of the radical changes is related to the application process. Until 1999 applicants were submitting their preferences to ÖSYM (prior to announcement of their scores) before final examination, ÖYS. Now, they first learn their scores, and then make preferences accordingly.

Another difference between the latest two systems is the method used for assessment of the applicant's secondary education academic achievement. In the previous system, a score called Secondary Education Academic Achievement Score (OBP) was being used, which was obtained by standardizing the applicant's high school *cumulative grade point average* (CGPA). In 1999, ÖSYM began to use a modified version of the old academic achievement scores. The new score called Weighted Secondary Education Academic Achievement Score (AOBP) is basically an OBP,

however, weighted according to the ÖSS score average of all students in the applicant's school.

1.2 Current Problem Environment

In year 2000, more than 1,400,000 candidates applied to ÖSYM for placement in higher education institutions. Figure 1.1 shows the total number of applicants and number of placements in two categories (recent graduates- those graduated from high school in the year they applied, and others) by the system each year from 1980 to 2000. Although not shown in the figure, the 'others' category consists of three sub-categories, (1) those who could not enter any institution in the previous years, (2) those who have entered and are currently registered students in an institution, and (3) those who have graduated from a higher education institution.

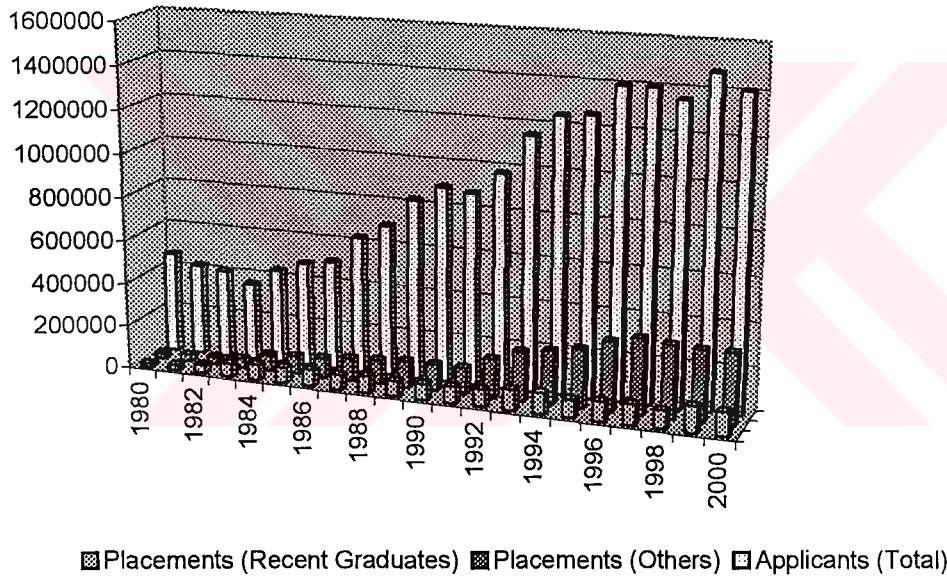


Figure 1.1: Number of applicants and placements between 1980 and 2000 (Source: ÖSYM (2001))

Figure 1.2 shows the ratio of number of placements to number of applicants in 2 categories (recent graduates and other) according to their graduation status for years 1980 to 2000.

Indeed the increasing number of applicants may be an indicator of the importance given to a university diploma in Turkey. According to OECD (1997), people who

enter higher education institutions in Turkey are mostly between ages 16 and 30. Comparing this population with the number of applicants for the periods during which central placement system have been applied, provides information about the demand for higher education. Table 1.1 gives a summary of population statistics and number of applicants between 1980 and 2000 in every five years. Obviously the demand (ratio of number of applicants to university entrant population) has significantly increased between 1980 and 2000 (almost doubled), whereas the population that is likely to have university education (ratio of entrant population to whole population) has almost been stationary during this period.

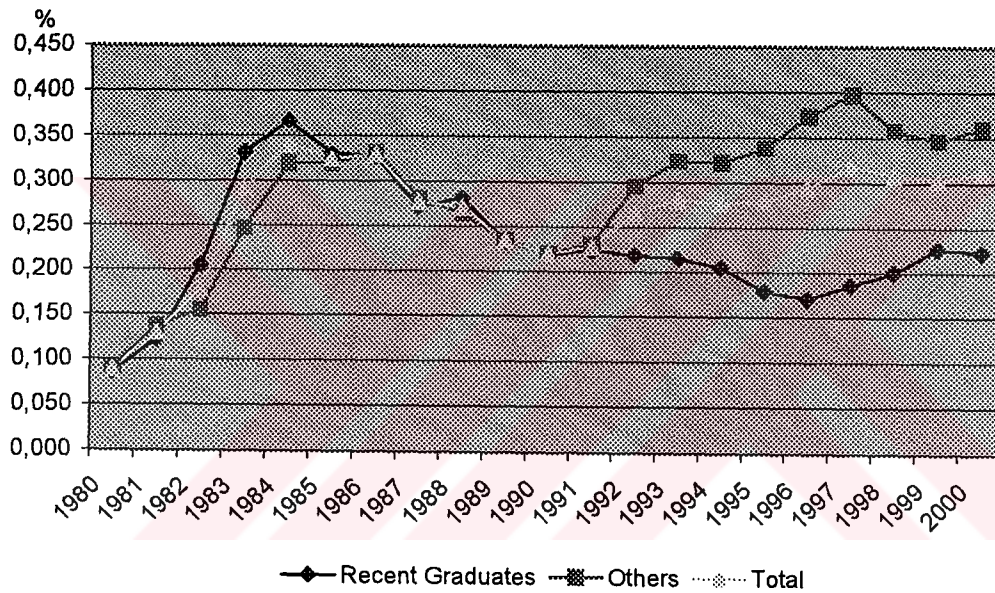


Figure 1.2: The ratio of number of placements to number of applicants according to graduation status between 1980 and 2000

Table 1.1: Whole Population (all ages), University Entrant Population (ages 15 and 29), and Number of Central Placement Applicants between years 1980 and 2000

Year	Whole Population (all ages)	University Entrant Population (ages 15-29)	Number of Applicants	Entrant/ Whole	Applicant/ Entrant
1980	44,736,957	12,392,312	466,963	0.277	0.038
1985	50,664,458	14,232,706	480,633	0.281	0.034
1990	56,473,035	16,125,100	892,975	0.286	0.055
1995	61,644,000	17,744,000	1,263,379	0.288	0.071
2000	66,835,000	19,308,000	1,407,920	0.289	0.073

Source: DİE (1995, 2001), ÖSYM (2001)

There may be different reasons behind the demand; probably the most important one is the desire to have better access to employment opportunities, and hence higher earnings. It would be meaningful to expect increasing demand for higher education, as the awareness of the economic and social benefits of university education increases. The results of employment and wage structure survey of State Institute of Statistics-DİE (1997) indicate that employees with university education earn more than other employees in different educational categories (DİE, 1997, Table 3, p.25), and they have better career opportunities (the increase in average wage with increasing age is higher for the employees in this category, DİE, 1997, Table 7, p.37).



CHAPTER 2

PROBLEM DEFINITION AND LITERATURE SURVEY

2.1 Educational Assessment and Central Placement

In both industrialized and developing countries national testing has always been one of the main strategic tools of educational policy makers as a mean to improve the educational quality (Chapman and Snyder, 2000). The efficiency (resulting from the usage of selection type questions) and the objectiveness of the central examinations are the key issues behind this popularity. Especially the multiple-choice type questions, with their flexibility to measure a variety of learning outcomes and adaptability to a wide range of subject-matter contents, are exclusively used in many standardized tests (Linn and Gronlund, 1995). Since the testing and assessment is totally central, the influence of family and political connections, and wealth is minimized. Although the standardized national testing is regarded as the most objective method of assessment, there are some criticisms in terms of the fairness of the system, especially about the effects of location on the success of applicants.

In Turkey, there are significant differences in the access rates to higher education between applicants from rural and urban areas, and developing and developed regions (ÜSYM, 1979). Since many higher quality secondary schools are located in urban and developed regions of the country, individuals from those areas perform better at the highly competitive university entrance examinations, thus access better higher education institutions. However, these location-specific disparities in the access rates do not necessarily indicate the degree of unfairness of the test (Dundar

and Lewis, 1999). Indeed they may be regarded as the indicators of the significant problems of Turkish educational system.

Since the essence of the central placement system in Turkey, is assessment of students with respect to some criteria, it would be meaningful to analyze the general principles of an educational assessment process first. According to Linn and Gronlund (1995), followings can be listed as the principles that will make an assessment process most effective:

- i. *Clear specification of what is to be assessed has a crucial importance in the process.* The measurement methods, which will be used in the assessment, depend on the clear descriptions of what is to be assessed. Thus, the whole process of central placement system and its individual components should also have their own objectives. According to ÖSYM, ÖSS aims to (1) assure a balance between the demand for higher education and the places available in higher education institutions, (2) select and place students with the highest probability of success in all available higher education programs considering their preferences and performance on the test (ÖSYM, 2000a). Although the stated objectives cannot be easily translated into statistical measures, they provide some insight to the intended use of the exam. Especially the second objective has a very broad meaning, which should be analyzed further. In this respect, the Scholastic Aptitude Test I (SAT I) of Educational Testing Service (ETS) of USA may be useful. SAT I is a test given to high school graduates, and it is commonly used for college admissions in USA. Therefore, ÖSS and SAT I have some common properties; since they are both national tests and used for selection purposes (it should be noted that there is not a central placement system in USA). ETS claims that SAT I can be useful in predicting the first year grades of applicants in the college. For this purpose, they have been constantly examining the correlation between test scores and first year college grades (ETS, 2001).
- ii. *Assessment procedures should be selected considering their relevance to the characteristics or performance to be measured.* An assessment procedure appropriate for some uses may not be so for the others. Considering the

demand for higher education, use of multiple-choice tests can be justified, however the type of questions, and their intended purposes should be open to debate.

In general there are two types of tests available: achievement and aptitude. The former is designed to assess what students have learned, and the latter predicts their ability to learn new tasks. Although there seems to be a clear distinction between the two types in definitions, both of them actually serve both of the uses to some extent. Linn and Gronlund (1995) argue that past achievement can also be a predictor of future achievement, thus both type of tests can be used in predicting future learning. However, they also state some reasons why aptitude tests should also be considered. In addition to the efficiency of administration of aptitude tests, their flexibility to be used with students from a wide range of backgrounds, and ability to predict future achievement in areas where students have no prior knowledge are the main reasons. They conclude that aptitude tests are more convenient measures and predict over a wide range of future experiences (Linn and Gronlund, 1995). ÖSYM introduces ÖSS as a general ability test based on common concepts in Turkish language, mathematics, social and natural sciences, and ÖYS as an achievement test (YÖK, 1999). With the implementation of the new system, ÖSYM has made a radical change, and shifted the dominance in the placement from an achievement test to an aptitude test.

- iii. *Comprehensive assessment requires a variety of procedures.* This item emphasizes the difference between the uses of procedures like multiple-choice questions and essay questions. Multiple-choice questions, measure a variety of learning outcomes and skills from simple to complex, they are adaptable to most types of subject contents, and they are considered as the most versatile type of test item available. However, it is obvious that some skills and learning outcomes cannot be measured via any form of selection item, such as ability to present and organize ideas. For such skills performance-based assessment methods are generally used like essay

questions, On the other hand it is obvious that objectively assessing those types of questions in a nation-wide examination is nearly impossible.

- iv. *Proper use of assessment procedures requires an awareness of their limitations.* According to Linn and Gronlund (1995) any educational or psychological instrument is subject to various types of measurement errors. Inadequate sampling of instructional content, chance factors influencing assessment, and incorrect interpretation of measurement results constitute major types of errors. Thus, one should realize the limited nature of information provided by the tests and use them accordingly.
- v. *Assessment is a means to an end, not an end in itself.* The use of assessment procedures implies that some useful purpose is being served and that the user is clearly aware of this purpose. The central placement system in Turkey is sometimes criticized with respect to this item, since the dominant components of the system, national tests, require students to devote much of their time and energy to get high scores. This amount of concentration on tests may cause students to consider entering into a program as an ultimate ambition.

High-stakes national testing is a term used to define these central nation-wide examinations given to all primary and secondary school graduates for the purpose of qualification, or selection and placement to future instruction. A test is generally considered to be high-stakes, if it has some real consequences for the applicants. High-stakes tests have long been used in the various stages of the Turkish educational system. From the post-primary school examinations to the university access, Turkish students are always in a competition aiming to have better access to the next educational institution they desire. In order to make decisions about students from all over the country with different backgrounds, this type of testing is necessarily needed, however, the effects of testing on students should not be ignored. Linn and Gronlund (1995) have itemized some of the main criticisms about national testing:

- *Tests create anxiety.* Especially considering the importance of university entrance examination for the Turkish students, there may be some serious psychological problems before, during and after the examination.

- *Tests categorize and label students.* Competitive environment as a result of the huge demand for higher education makes the categorization unavoidable. Higher quality institutions like Boğaziçi and Middle East Technical Universities have students from the upper ranks of the assessment system. Consequently the rest of the universities take students from the lower ranks, according to their perceived quality from the viewpoint of students and their families. As a result, some of the students unsatisfied by their current placements again take university examinations to access to a desired institution, and they form a significant portion of re-applicants (ÖSYM, 2001).
- *Tests damage students' self-concepts.* The stereotyping of students as a consequence of the similarity in preparing for the central examinations is counted as a serious problem. Another misinterpretation is the overgeneralization of the test results for the students; especially students receiving low scores may develop a general sense of failure. However, interpretation of the test results should be made according to the measurement capabilities of the items (questions) used in the test. Obviously getting high (or low) scores does not necessarily guarantee academic success (or failure) in the institution students are placed. It should be noted that the test scores and future academic achievement is only a matter of correlation.
- *Tests create self-fulfilling prophecies.* According to Linn and Gronlund (1995) test scores create teacher expectations concerning the achievement of individual students; then the teachers teaches in accordance with those expectations, and the students respond by achieving at the expected level. This criticism is directed primarily for the scholastic aptitude tests. Although until 1999, ÖSS was not dominant in the placement process, it is now (as an aptitude test) the determining factor in the process. Therefore within a few years this criticism may also be valid for Turkey case also.

In addition to the above, Linn and Gronlund (1995) also mention some public concerns related to the national tests. They especially emphasize the undesirable shifts in the curricula of schools as a consequence of the heavy demand of national

tests on students' time and effort. They claim that direct preparation for the test is likely to affect classroom activities and distort the curricula. ÖSYM also has a similar argument, and claims that the common use of private tutoring and tactics to solve more and more questions in order to learn the tricks of the tests is an important problem for the Turkish educational system (YÖK, 1999). Dunder and Lewis (1999) claim that private tutoring in Turkey is so common that almost 20% of the applicants in Turkey took some private coaching in 1991. They also argue that many other countries (e.g., Japan, Taiwan, South Korea, Hong Kong, Greece), where places available are limited and allocated competitively and centrally, use private tutoring commonly.

Another important point Linn and Gronlund (1995) mentioned is the misuse and misinterpretation of test scores. They claim that if the public understands the limits of the standardized tests better, it would be easier to overcome these problems. Considering the limited assessment capacity of the national tests, the authors argue that in any admission or other educational decision, test scores should also be supplemented by past records of achievement and other types of assessment data. With an aim similar to the above argument, ÖSYM has been considering academic achievement in the placement decisions since 1981. Therefore the inclusion of academic achievement can be considered as an act to increase the fairness of the placement system. However, the problem arises with the central implementation of the system. Obviously assessing the past academic achievement is not easy without using a standardized method. Standard achievement tests like the previous system's ÖYS may help the decision makers, however the problems of those tests, as a natural consequence of using some restricted forms of questions, should also be considered.

2.2 Previous and Current Systems

ÖSYM, with the introduction of the new system in 1999, has made substantial changes in the central placement. Not only the number of criteria, but also the weights given to different types of criteria have been drastically changed. In the new practice there are two main criteria (a national test score and past academic

achievement), whereas in the former one three criteria were used (one of the national tests of the previous system is no longer in use).

Table 2.1 gives an idea about how an average student would be evaluated in an average school in the previous and current placement systems (recall that ÖYS is the achievement test in the previous system, ÖSS is the aptitude test administered in both systems, OBP was the score to assess the past achievement in the previous system, and AOBP is the score to assess the past achievement in the current system).

Table 2.1: Comparisons of weights of different placement criteria for an average student performance in an average school in the previous and current systems

Criterion	Previous System (1998)				Current System (1999)			
	Total Sub-Criterion Coefficients	Sub-Criterion Average	Criterion Total	Relative Weight (%)	Total Sub-Criterion Coefficients	Sub-Criterion Average	Criterion Total	Relative Weight (%)
ÖYS	5.5	50.0	275.0	76	-	-	-	-
ÖSS	1.1	50.0	55.0	15	2.2	50.0	110.0	80
OBP	0.6	50.0	30.0	8	-	-	-	-
AOBP	-	-	-	-	0.5	55.0*	27.5*	20
Total			360	100			137.5	100

*: Assuming that an average school has ÖSS average around 110, and its least successful student has an OBP of 30

ÖSYM, eliminating the previous system's ÖYS, has increased the weights of two other components in the current placement system. ÖSS had the highest increase, on the average from 15% to 80%, whereas the weight of past academic achievement has also increased significantly (8% of OBP as compared to 20% of AOBP). In Figure 2.1 the changes in the relative weights of different placement criteria applied since 1980 are shown (the relative weights are approximated for an average student in an average school).

Note that, an achievement test has always been the major criterion of the central placement until 1999; however, an aptitude test has become the dominating component with the application of current system. YÖK (1999) explains the rationale behind this change with the high correlation (nearly 100%) between the two exam results (ÖSS and ÖYS) in the previous years. As a result of eliminating one of the previous system's national tests, the weight given to the minor component, past

academic achievement, has also increased to the current level; YÖK (1999) declares its intention to continuously increase the weight of past academic achievement in the future placement systems. Therefore, the need for a fair assessment of this criterion will be even more important in the future.

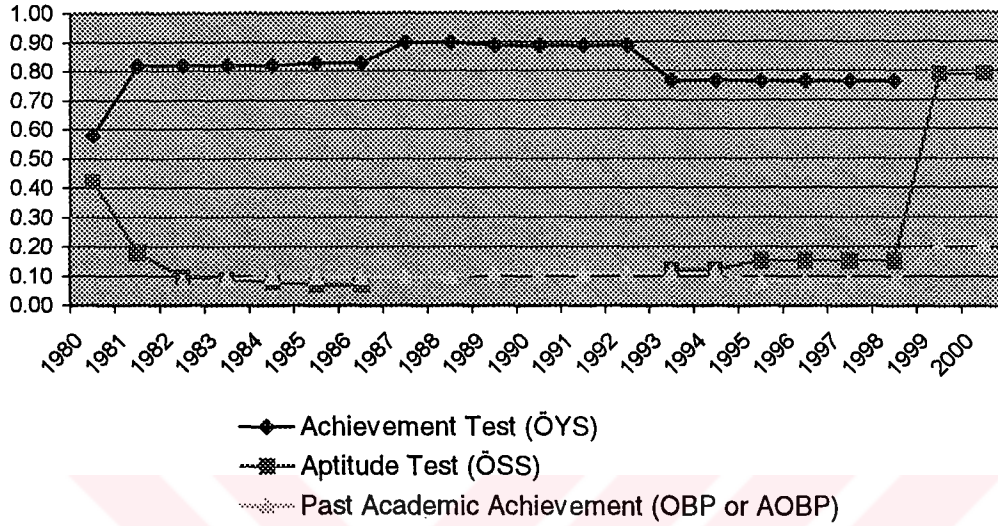


Figure 2.1: Changes in the relative weights of placement criteria between 1980 and 2000 (Source: ÜSYM (1980), ÖSYM (1981 - 1999, 2000b))

Besides the weights, the procedure used to assess past achievement has also been changed in the new system. Until 1998, Secondary Education Academic Achievement Score (OBP) was used for assessment of the students' past academic achievement, which was the standardized form of the CGPA distribution of students in a school. Although OBP could be used to rank students according to their academic achievement levels in a school, it was poor in inter-school comparisons (one of the main criticisms against the previous system). This was especially due to OBP's lack of ability to discriminate schools based on their quality of education, and to the standardization process depending highly on the distribution characteristics of the schools.

In general, schools like Science and Anatolian High Schools of Turkey were subject to serious unfairness because of the old system. Since these schools accept students according to nation-wide examinations administered at the end of primary education, their student populations differ significantly from ordinary high schools in terms of

students' aptitude and achievement levels. Treating them like the other schools has always been one of the major problems of the central placement systems in Turkey. Especially the 'elite' students of Science High Schools compete for the upper percentiles in the central placement system, and hence the marginal benefits (increase in their ranks as a result of the increase in their academic achievement scores) as a result of changing their schools were more valuable for them than for students in other categories. The obvious consequence of this problem was the mass-departures of students from Science High Schools as seen in Table 2.2, which gives the number of graduates of Science High Schools between 1989 and 1999 (YÖK, 1999). The increase in the number of Science High School graduates from 1998 to 1999 is the natural result of the new system, which eliminates some of the disadvantages of the previous system concerning schools in this category.

Table 2.2: Number of students graduated from Science High Schools between 1989 and 1999

Year	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Number of Students Graduated	422	526	483	720	748	1088	1079	1138	848	848	2319

Source: YÖK (1999)

The current system is not using OBP as a criterion for placement in the crude sense; instead it determines an applicant's academic achievement score using two criteria; (1) the allowable academic achievement score range of her school, and (2) her place in this allowable range. The first criterion, allowable score range of a school, is mainly determined according to the averages of its students' ÖSS scores, which is independent of individual OBP's of the students. However, the second criterion uses OBP to place the applicant in the allowable range according to her place in the OBP distribution of the school.

Before analyzing the two central placement practices in detail, the standardization procedure, which is used extensively in both of the systems in various stages, should be explained.

2.2.1 Standardization Procedure

The basic idea of the procedure is to transform a score into another, so that it is comparable with the other scores of the candidate. The transformation is linear and does not change the shape of the score distribution; it only affects the location and the scale parameters. As a result of the standardization, a score distribution transforms into a distribution having an arithmetic mean of 50 points and a standard deviation of 10 points. The procedure is as follows:

Let X_1, X_2, \dots, X_n be random variables from a continuous probability distribution function $F(x)$. Then the sample mean and sample standard deviation are defined as follows (these are the two estimators used by ÖSYM (2000) during the calculations):

$$\bar{X}(n) = \frac{\sum_{i=1}^n X_i}{n}, \text{ and } s(n) = \sqrt{\frac{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}{n-1}}$$

According to above definitions, the transformed (standardized) random variables, Y_i are calculated as follows:

$$Y_i = \left[\frac{X_i - \bar{X}(n)}{s(n)} \right] \times 10 + 50, \quad i = 1, 2, \dots, n \quad (2.1)$$

This process is used in determining both *academic achievement scores* and the *standard test scores* in both systems. For the former transformation, the CGPA distribution of a school is transformed into an OBP distribution using the procedure described above; for the latter, raw score distribution of the applicants in various test categories are transformed into standard test scores. One should note that first standardization procedure is applied to every school population separately; however, the latter is applied to whole population of central examination applicants.

2.2.2 Previous System (1981 to 1998)

The previous system, which was applied until 1999, consisted of three components:

- i. Student Selection Examination Score (ÖSS Score)

- ii. Student Placement Examination Score (ÖYS Score)
- iii. Secondary Education Academic Achievement Score (OBP)

Since ÖSS and OBP of the previous system are still in use, detailed analyses of these are left to Section 2.2.3.

ÖSS of the previous system was an examination to assess general abilities of the applicants with the purpose of eliminating the ones who (probably) lack the necessary skills needed for higher education (candidates whose ÖSS scores are less than 105 points were not allowed to take the second-stage examination; ÖYS). For every candidate I , 3 composite ÖSS scores ($\ddot{O}SS_i^k$, $k = 1, 2, 3$) were calculated from 2 standard ÖSS test scores ($S\ddot{O}SS_i^l$, $l = 1, 2$) with different weight combinations according to the coefficients c_i^k in Table 2.3 and the following formula:

$$\ddot{O}SS_i^k = \sum_{l=1}^2 (c_i^k \times S\ddot{O}SS_i^l), \quad k = 1, 2, 3$$

Table 2.3: Coefficients for composite ÖSS scores of the previous system (c_i^k)

Composite ÖSS scores ($\ddot{O}SS_i^k$)	Standard ÖSS test scores ($S\ddot{O}SS_i^l$)	
	Quantitative (1)	Verbal (2)
Quantitative (1)	1.8	0.4
Verbal (2)	0.4	1.8
Equal Weighted (3)	1.1	1.1

Source: ÖSYM (1998)

ÖYS as an achievement test was the final stage of the previous system, where the success strictly depended on the level of knowledge of the applicants on the subject contents of the exam. Different from ÖSS, where students were required to answer all the questions in the test, a category or a set of categories of questions were chosen and answered by the students in ÖYS. According to the scores calculated for those categories, the students were ranked for the placement. The composite ÖYS placement scores in six categories ($Y\ddot{O}YS_i^t$, $t = 1, 2, \dots, 6$) were calculated using the standard ÖYS test scores in five categories ($S\ddot{O}YS_i^l$, $l = 1, 2, \dots, 5$), composite ÖSS

scores and OBP (OBP_i , OBP score of i^{th} applicant) of the candidates according to the following formula (coefficients used in the below formula are listed in Table 2.4):

$$YÖYS_i^t = \sum_{l=1}^5 c_{t,l}^{\ddot{O}YS} \times S\ddot{O}YS_i^l + \sum_{k=1}^3 c_{t,k}^{\ddot{O}SS} \times \ddot{O}SS_i^k + c^{OBP} \times OBP_i, \text{ for every } t$$

Table 2.4: The Coefficients of the ÖYS standard test scores ($c_{t,l}^{\ddot{O}YS}$), composite ÖSS scores ($c_{t,k}^{\ddot{O}SS}$) and OBP (c^{OBP}) for the placement scores (YÖYS) in the previous system

Composite ÖYS Scores ($YÖYS_i^t$)	Standard ÖYS Test Scores ($S\ddot{O}YS_i^l$)					Composite ÖSS Scores ($\ddot{O}SS_i^k$)			Past Academic Achievement
	Natural Sciences (1)	Mathematics (2)	Turkish Language (3)	Social Sciences (4)	Foreign Language (5)	Quantitative (1)	Verbal (2)	Equal Weighted (3)	OBP
Natural Sciences (1)	2.7	1.7	1.1	-	-	0.5	-	-	0.6
Mathematics (2)	1.7	2.7	1.1	-	-	0.5	-	-	0.6
Turkish Language and Mathematics (3)	-	2.2	2.2	1.1	-	-	-	0.5	0.6
Social Sciences (4)	-	1.1	1.7	2.7	-	-	0.5	-	0.6
Turkish Language and Social Sciences (5)	-	0.5	2.5	2.5	-	-	0.5	-	0.6
Foreign Language (6)	-	-	1.1	1.1	3.3	-	0.5	-	0.6

Source: ÖSYM (1998)

2.2.3 Current System (1999 to Present)

In the current system, composite ÖSS placement score (YÖSS) is the main criterion in the placement of applicants, and it consists of two sub-components:

- Student Selection Examination Score (ÖSS Score)
- Weighted Secondary Education Academic Achievement Score (AOBP)

YÖSS is a linear combination of these two scores.

2.2.3.1 Student Selection Examination (ÖSS)

According to ÖSYM (2000a), ÖSS has two objectives:

- i. To select those candidates who will be considered in the placement decisions
- ii. To select and place those candidates qualifying for the placement decisions, in higher education programs of their highest preference, compatible with their relevant weighted composites as the criterion measures

The first objective of ÖSS is for the programs demanding special skills like arts, physical education, etc. where there is an additional qualification examination. Obtaining 105 points from ÖSS is a minimum requirement for those programs. The second objective of ÖSS is for the rest of the higher education programs. According to this objective, placement of the applicants is made considering their ÖSS scores, secondary education academic achievement, programs in their preferences lists, and the capacity restrictions of those programs.

ÖSS, consisting of four tests in different categories, aims to assess the applicants' general ability in two main categories: verbal and quantitative. Each main category consists of two sub-categories: mathematics and natural sciences tests form the quantitative, and Turkish language and social sciences tests form the verbal sections. In any sub-category, there are generally 45 multiple-choice questions, for which the applicants have to choose the correct answer from a set of 5 answers. For each applicant and for each category and sub-category, *Raw Test Scores* are calculated by subtracting $\frac{1}{4}$ of the number of wrong answers from the number of correct answers in that category or sub-category (a correction against guessed answers in multiple-choice questions). ÖSYM, considering all the valid raw test scores of all applicants, calculates the arithmetic mean and standard deviation of those raw scores for each category and sub-category (the applicants may have negative raw scores for some categories and these are not included in the calculations). The arithmetic means and standard deviations are used in the calculation of standard test scores from the raw scores according to equation (2.1).

Letting $RÖSS_i^l$ and $SÖSS_i^l$ be the raw and standard test scores of applicant i in ÖSS for the standard test score category l ($l = 1, 2, \dots, 6$), and $\bar{X}_{ÖSS}^l(m_l)$, $s_{ÖSS}^l(m_l)$ (where m_l

is the number of applicants having valid raw test scores in score category l) be the sample mean and standard deviation of raw test scores in category l . Then standardization procedure is as follows:

$$S\ddot{O}SS_i^l = \left[\frac{R\ddot{O}SS_i^l - \bar{X}_{\ddot{O}SS}^l(m_l)}{s_{\ddot{O}SS}^l(m_l)} \right] \times 10 + 50, \text{ where}$$

$$\bar{X}_{\ddot{O}SS}^l(m_l) = \frac{\sum_{i=1}^{m_l} R\ddot{O}SS_i^l}{m_l}, \text{ and } s_{\ddot{O}SS}^l(m_l) = \sqrt{\frac{\sum_{i=1}^{m_l} (R\ddot{O}SS_i^l)^2 - \frac{\left(\sum_{i=1}^{m_l} R\ddot{O}SS_i^l\right)^2}{m_l}}{m_l - 1}}$$

Having valid raw scores in all of the sub-categories of the test, 3 composite ÖSS scores are calculated for each applicant i ($\ddot{O}SS_i^k$, $k = 1, 2, 3$) and reported to them. In addition to ÖSS, an additional Foreign Language Examination (YDS) is administered for the applicants trying to enter foreign language departments of the institutions. These composite scores and corresponding coefficients are shown in Table 2.5 (Table 2.5 does not consider YDS scores).

Using the coefficients in Table 2.5 composite ÖSS score of applicant i in category k (where $k = 1, 2, 3$) is calculated as follows:

$$\ddot{O}SS_i^k = \sum_{l=1}^6 (c_l^k \times S\ddot{O}SS_i^l), \text{ for every } k$$

Table 2.5: The coefficients of the standard test scores (c_l^k), for the composite ÖSS scores

Composite ÖSS Scores ($\ddot{O}SS_i^k$)	Standard ÖSS Test Scores ($S\ddot{O}SS_i^l$)					
	Verbal (1)	Turkish Language (2)	Social Sciences (3)	Quantitative (4)	Mathematics (5)	Natural Sciences (6)
Quantitative (1)	0.4	-	-	1.8	-	-
Verbal (2)	1.8	-	-	0.4	-	-
Equal Weighted (3)	-	0.8	0.3	-	0.8	0.3

Source: ÖSYM (2000b)

2.2.3.2. Secondary Education Academic Achievement Score (OBP)

In the older system, assessment of past academic achievement was based on the OBP of an applicant, which was the standardized form of CGPA distribution according to the procedure explained in Section 2.2.1.

Letting $CGPA_i$ and OBP_i be the cumulative grade point average and secondary education academic achievement scores of the i^{th} applicant (since any applicant can not be a graduate from any two schools at the same time there is no need for an index denoting the school, i.e. applicant i is strictly the student of school j), and letting $\bar{X}_{CGPA}^j(n_j)$ and $s_{CGPA}^j(n_j)$ be the sample mean and standard deviation of the CGPA distribution in the school j of the applicant i (ÖSYM considers only the recent graduates of the school at the year of calculation) (where j is the index of schools officially recorded at ÖSYM, and n_j is the number of recent graduates of school j)

$$OBP_i = \left[\frac{CGPA_i - \bar{X}_{CGPA}^j(n_j)}{s_{CGPA}^j(n_j)} \right] \times 10 + 50 \quad (2.2)$$

where

$$\bar{X}_{CGPA}^j(n_j) = \frac{\sum_{i=1}^{n_j} CGPA_i}{n_j}, \text{ and } s_{CGPA}^j(n_j) = \sqrt{\frac{\sum_{i=1}^{n_j} (CGPA_i)^2 - \frac{\left(\sum_{i=1}^{n_j} CGPA_i\right)^2}{n_j}}{n_j - 1}}$$

OBP measures the distance between the CGPA of an individual student and average of CGPA's in the school in terms of the sample standard deviation of CGPA's. Indeed it is a measure of deviation from the central tendency in units of sample standard deviation. For example if two students in a school take 30 and 70 as an OBP respectively, this only means the second student has a CGPA 2 standard deviations above the average (since in the standardized form sample standard deviation is set to 10), whereas the other student is 2 standard deviations below the average. If the students in this school were ranked according to their past academic achievement,

obviously the one with the higher OBP score will be at a higher rank. Therefore, it can be concluded that OBP is a meaningful measure for the intra-school comparisons of the students (i.e., comparisons of the students in a school). However in the previous system, the direct use of OBP as a placement criterion necessitates the inter-school comparisons also, where OBP is criticized to be poor.

OBP is only a local measure of academic achievement, and such scores do not carry much information related to the real value of academic knowledge, which may be necessary if the aim is to rank the students from different schools on a universal academic achievement scale. However if the aim of ÖSYM is to reward (or punish) the students, who are equally above (or below) the average in terms of standard deviations of their schools, with equal scores, then the OBP is the correct measure. The two policies described in this paragraph represent two different aims, which should be implemented to some extent for the ideal assessment of past academic achievement. A policy that ignores the quality of education in a school will create a disadvantage for the schools with higher education levels like Science High Schools; on the other hand measuring the academic achievement only considering the quality and ignoring the efforts of students in the schools with lower educational quality will create a negative effect for these schools. The desire to find a reasonable solution considering both dimensions of the problem is probably the rationale of ÖSYM in designing the new academic achievement system, which will be analyzed in detail in the next section.

Table 2.6: Effects of Changing School on OBP

Difference Between Schools	$\bar{X}_1 > \bar{X}_0, s_1 = s_0$	$\bar{X}_1 < \bar{X}_0, s_1 = s_0$	$\bar{X}_1 = \bar{X}_0, s_1 > s_0$	$\bar{X}_1 = \bar{X}_0, s_1 < s_0$
Student is above the mean $CGPA_i > \bar{X}_1$	×	✓	×	✓
Student is below the mean $CGPA_i < \bar{X}_1$	×	✓	✓	×

(\bar{X}_j, s_j) : sample mean and standard deviation of school j (0: initial, 1: last), ✓ : changing school is advantageous, × : changing school is not advantageous

Another crucial part of the standardization is the effects of the two parameters determining the OBP levels: mean and standard deviation of CGPA's, which differ significantly between the schools. Although a student cannot easily affect those school-wide measures individually, changing his or her school may do the same effect. Table 2.6 summarizes the effects of changing school. The first conclusion about the table is for the effect of mean CGPA on OBP: whether or not the student is above the mean in the new school, changing school is advantageous as long as the student goes to a school with a lower mean (column 2). However, being above or below the mean affects the standard deviation case, where a student entering a new school with a lower standard deviation is advantageous only if she is above the mean CGPA in the new school.

Although the standardization process does not change the shape of the distribution, the upper and lower limits imposed by ÖSYM change the original CGPA distribution. The rules are as follows:

- If a student's OBP is more than 80 points, then her OBP is assumed to be equal to 80 (generally the most successful, or some of the most successful students, in terms of CGPA, may be above this limit), otherwise OBP does not change.
- If a student's OBP is less than 30 points, then it is assumed to be equal to 30 (generally the least successful, or some of the least successful students, in terms of CGPA, may be below this limit), otherwise OBP does not change.

Indeed, considering that standard deviation is 10 points in the OBP distribution, the upper limit is 3 standard deviations above the arithmetic mean, and the lower limit is 2 standard deviations below the mean (i.e. the rule of ÖSYM can be represented with the following range $[\bar{X} - 2s, \bar{X} + 3s]$, where the parameters used are sample mean- \bar{X} , and standard deviation- s). Since OBP was directly used as a placement criterion in the previous system, the imposed lower limit was an advantage to unsuccessful students, protecting them from taking too low scores. On the contrary, the upper limit was a disadvantage to successful students, since it prevents them taking scores higher than 80.

Applying this rule, ÖSYM intends to detect the extreme students- outliers (at both ends of their schools), so that they wouldn't get too high or low OBP scores. If there exists any outlier student beyond those limits, his or her OBP is trimmed to the nearest upper or lower limit. ÖSYM's asymmetric $[\bar{X} - 2s, \bar{X} + 3s]$ criterion seems to have been designed considering only the right-skewed CGPA distributions. However, expecting only this kind of distribution characteristics for all of the schools is not a reasonable assumption. A criterion that is more robust against the different types of distribution characteristics would be a better choice. Some alternative methods to the outlier detection rule of ÖSYM are discussed in Section 2.3.

2.2.3.3 Weighted Secondary Education Academic Achievement Score (AOBP)

AOBP's purpose is again to assess the applicants' high school academic achievement, so that the academic knowledge of the secondary education is reflected the final score, hence the placement. The applicant with a higher CGPA has a higher AOBP like it was in the predecessor. Although the AOBP is based on OBP, there are significant differences between the two. The most important one is the new system's ability to discriminate schools based on a few criteria.

The AOBP formula according to ÖSYM (2000b) is as follows:

$AOBP_i^k$: Weighted secondary education academic achievement score of applicant i weighted according to composite ÖSS score category k

A_j^k : Arithmetic mean of the valid ÖSS Scores of students in composite score category k ($k = 1, 2, 3$) in school j

B_j : Highest OBP in the school j

C_j : Lowest OBP in the school j

Let

$$D_j^k = \frac{80 - \left[\left(\frac{A_j^k}{80} C_j \right) - \left(\frac{A_j^k - 80}{10} \right) \right]}{\left(\frac{A_j^k}{80} B_j \right) - \left(\frac{A_j^k}{80} C_j \right)}, \text{ then}$$

$$AOBP_i^k = \left[\left(\frac{A_j^k}{80} C_j \right) - \left(\frac{A_j^k - 80}{10} \right) \right] + \left[\left(OBP_i \frac{A_j^k}{80} \right) - \left(\frac{A_j^k}{80} C_j \right) \right] D_j^k \quad (2.3)$$

where

$$A_j^k = \frac{\sum_{i=1}^{n_j} \ddot{O}SS_i^k}{n_j}$$

The above formulation (2.3) can be transformed to the below one (2.4) (see Appendix A for details):

Let

$$\alpha_j^k = \left[\frac{A_j^k}{80} \times (C_j - 8) \right] + 8$$

then

$$AOBP_i^k = \alpha_j^k + \left[(80 - \alpha_j^k) \frac{OBP_i - C_j}{B_j - C_j} \right] \quad (2.4)$$

Observing the new formula (2.4), it can be noted that the α_j^k term behaves like a lower limit for the AOBP scores of a school. This term is independent of students' OBP— except the least successful student in the school, C , and only depends on the $\ddot{O}SS$ average of the students in the school. Therefore, AOBP of a student in a school is somewhere between α_j^k and 80 range according to her place in her school's CGPA distribution (according the ratio $(OBP_i - C_j)/(B_j - C_j)$, which is between 0.0 and 1.0). Independent of her OBP, the most successful student in any school has always an AOBP of 80 points (since the ratio for her is 1) (To simplify the notation the subscripts and superscripts of α_j^k , A_j^k , B_j , and C_j will be dropped in most of the discussions below).

ÖSYM tries to balance the two different policies of assessing the academic knowledge mentioned in Section 2.2.3.2. It weights the academic knowledge using A with an aim to include the quality education of schools into the AOBP's (assuming that A is related to the educational level of a school). However it fixes the upper limit of AOBP to 80, which means equal reward for every successful student independent of their schools' education quality. Thus, at any school, a student has the chance to have the same maximum academic achievement score independent of her school's educational level, which also means the same chance to enter any higher education institution, based on the performance in the other component of the placement system, ÖSS.

Figure 2.2 presents a visual comparison between the previous and the current system.

According to Figure 2.2 there are two main components of AOBP:

- i. Determination of α as a lower limit for a school
- ii. Determination of AOBP of students in the $[\alpha, 80]$ range with $(OBP_i - C_j) / (B_j - C_j)$ term

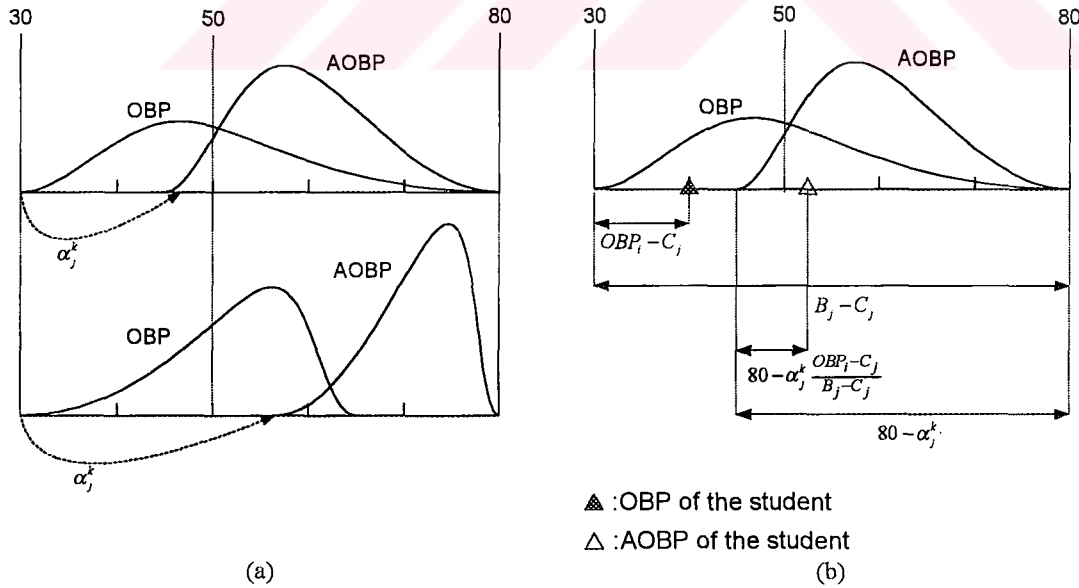


Figure 2.2: (a) α lower limits for two hypothetical school distributions, (b) determination of AOBP of a student in the $[\alpha, 80]$ range

α term in formula (2.4) depends on the parameters A and C for any school, and is independent of the students' OBP in the school (except C). Since α is proportional to A, the current system values the schools' success in terms of ÖSS score averages. Indeed the drawback of the previous system's lack of ability to discriminate the schools based on their educational quality, is tried to be eliminated with the inclusion of A. Since A is the arithmetic mean of ÖSS scores of the students in a school, its usage in the formula to determine the lower limit can be meaningful. Although A is not independent of the performance of any individual student, a single student's contribution is marginal compared to the other components of AOBP, especially when the number of students in the school is large.

From the viewpoint of assessing academic achievement, ÖSS scores can probably be highly correlated with the students' quantitative and verbal knowledge developed in the school, thus it may be used as a parameter while comparing the schools. However weighting the applicants' secondary education knowledge with the performance on a central examination consisting of only multiple-choice questions is a question that should be answered (which is out of the focus of this study). Using this method is a natural consequence of a system, which is seeking a more objective way of assessment of more people.

Unlike A, usage of C in determining the α term should be questioned in detail, especially its dependence on an individual or a group of individuals, and its effects on the whole school population. Since OBP affects directly the placement scores in the previous system, there the lower limit 30 protects individual students from falling far below the arithmetic mean of the school. However, with the new use of C in determining α , the lower limit 30 now protects the whole school population. As discussed in the following section in detail, ÖSYM's rule tends to label higher numbers of outliers, which means the probability of observing a C value of 30 increases as the school gets more crowded. It can be concluded that a hypothetical school will benefit less from the advantage associated with the use of C in the α term, as its population increases.

Second part of the formula (2.4) determines a student's place in her school's AOBP range (i.e. $[\alpha, 80]$ range), which corresponds to the aforementioned transformation of OBP's in $[C, B]$ to AOBP's in $[\alpha, 80]$ range (not the standardization of equation (2.2)). The ratio $(OBP_i - C_j) / (B_j - C_j)$ used here, is between 0.0 and 1.0 depending on the OBP's of students, hence CGPA values. The more successful the student is, the higher the ratio will be.

If the ÖSYM's lower and upper limits $[\bar{X} - 2s, \bar{X} + 3s]$ are applied to the CGPA distributions also, then the transformation with the ratio $(OBP_i - C_j) / (B_j - C_j)$ removes all the effects of the standardization procedure. If the $CGPA_C$ and $CGPA_B$ are the CGPA's of a school determined in a similar way as the C and B are determined, then the following equality holds:

$$\frac{OBP_i - C_j}{B_j - C_j} = \frac{\left[\frac{CGPA_i - X_{CGPA}^j}{s_{CGPA}^j} 10 + 50 \right] - \left[\frac{CGPA_C - X_{CGPA}^j}{s_{CGPA}^j} 10 + 50 \right]}{\left[\frac{CGPA_B - X_{CGPA}^j}{s_{CGPA}^j} 10 + 50 \right] - \left[\frac{CGPA_C - X_{CGPA}^j}{s_{CGPA}^j} 10 + 50 \right]} = \frac{CGPA_i - CGPA_C}{CGPA_B - CGPA_C}$$

Analyzing the components of AOBP will provide an insight to the formula and its reaction to the changes in its components. For this purpose, a change in AOBP of a student is tried to be defined in terms of changes in its components (OBP, A, B, C).

Let $AOBP_{i,0}^k$ and $AOBP_{i,1}^k$ be the AOBP of student i in composite score category k , before and after the change respectively, and $\Delta AOBP_i^k$ be the difference $\Delta AOBP_i^k = AOBP_{i,1}^k - AOBP_{i,0}^k$. The formulations below are derived from the derived equation of AOBP (2.4).

AOBP vs. OBP

$$\Delta AOBP_i^k = \frac{80 - \alpha_j^k}{B_j - C_j} \Delta OBP_i,$$

where

$$\Delta OBP_i = OBP_{i,1} - OBP_{i,0}$$

It is obvious that any change in the OBP of an applicant directly affects the AOBP in the same direction. Since only the amount of change is important, it is independent of the applicant's OBP. However the amount of change is affected by the $(80 - \alpha_j^k) / (B_j - C_j)$ ratio, which depends on the school characteristics.

AOBP vs. A

$$\Delta AOBP_i^k = \frac{B_j - OBP_i}{B_j - C_j} (C_j - 80) \frac{\Delta A_j^k}{80},$$

where

$$\Delta A_j^k = A_{j,1}^k - A_{j,0}^k$$

Since the lower limit α is directly proportional to A, any positive change in the ÖSS scores average A of a school will positively affect all the AOBP's in that school. Note that the change of AOBP is directly proportional with $(B_j - OBP_i) / (B_j - C_j)$ term, which is greater for the students with lower OBP's. Consequently when a change in the ÖSS scores average of a school occurs, the AOBP of students with lower OBP's are affected more than the students with higher OBP's. The change in A is also affected by C_j , where the schools with higher C values are affected more.

AOBP vs. B

$$\Delta AOBP_i^k = \frac{-\Delta B_j}{(B_{j,1} - C_j)(B_{j,0} - C_j)} (80 - \alpha_j^k) (OBP_i - C_j)$$

A positive change in B will negatively affect AOBP. An increase in the OBP of the most successful student will decrease AOBP of all the students in school j . In other words an increase in the OBP of an individual, B, means a decrease in the AOBP of the others (note that with a positive increase in B, ranks in the school are not changed, thus AOBP of the most successful students is again 80, the upper limit).

The effect of change in B, unlike the above components, is directly proportional to OBP of a student (because of $OBP_i - C_j$ term). The higher the OBP of a student is, the more she will be affected by the change in B. Thus the most disadvantageous person after this change is the second most successful student.

AOBP vs. C

Since C is used in two different terms in the AOBP formulation (2.4), its effects are divided into two: The first C analyzed is the one in the calculation of lower limit α (note that the other C used in the ratio $(OBP_i - C_j)/(B_j - C_j)$ in the AOBP formulation (2.4) is not changed). Then, the difference in AOBP can be defined as follows:

$$\Delta AOBP_i^k = \frac{A_j^k}{80} \left(\frac{B_j - OBP_i}{B_j - C_j} \right) \Delta C_j,$$

where

$$\Delta C_j = C_{j,1} - C_{j,0}$$

The positive change in the value of C affects the AOBP scores in the same direction.

In order to analyze the effects of change in C in the ratio $(OBP_i - C_j)/(B_j - C_j)$, the following equation is derived (note that, this time the C in α term is not changed):

$$\Delta AOBP_i^k = \frac{-\Delta C_j}{(B_j - C_{j,1})(B_j - C_{j,0})} (80 - \alpha_j^k) (B_j - OBP_i),$$

where

$$\Delta C_j = C_{j,1} - C_{j,0}$$

This effect of C is similar to the effect of B on AOBP; an increase in C will decrease the AOBP of others, and the highest change will be observed on the one closest to C term, the second least successful student (because of $B_j - OBP_i$ term).

2.2.3.4 Placement Scores (YÖSS)

The placement scores are the final scores, which are used to place the students into the institutions. YÖSS for an applicant i in composite score category k ($YÖSS_i^k$) is a linear combination of ÖSS score and AOBP score of the applicant according to the following formula:

$$YÖSS_i^k = ÖSS_i^k + c_{i,j}^{AOBP} \times AOBP_i^k$$

The coefficient $c_{i,j}^{AOBP}$ of AOBP is determined according to applicant's secondary education major. The aim in determining the coefficients is to reward an applicant to choose a major that is considered to be a direct continuation of her high school major. The coefficient $c_{i,j}^{AOBP}$ takes one of the following values 0.2, 0.5, 0.65. If ÖSYM considers the higher education program as direct continuation of the major, it is 0.5; otherwise it will be 0.2. If an applicant is a graduate of vocational or technical school and the intended program is considered to be as continuation, then it will be 0.65.

2.3 Outlier Detection Rules

Identifying outliers in data has always been one of the primary concerns of data analysis. Consequently there is a vast amount of literature on this topic, discussing the concepts of outlyingness, and rules to deal with them. Barnett and Lewis (1984), as one of the reference books in this area, defines the outlier in a set of data as *an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data*. Obviously 'appears to be inconsistent' is the critical point of the definition, which emphasizes the subjective judgment of the observer on outliers. Hence, those limits determined by ÖSYM describe a policy against the outlier students in schools, whether it is robust or not.

Declaring an observation as an outlier depends on its appearance in relation to the postulated model about the data, F . If some of the observations in the data set do not come from the distribution F , but from a different distribution G , then they are called

as *contaminants*. Note that contaminants do not have to be the extreme observations in our data set, instead they may be hidden in somewhere between the observations coming from the original model F ; similarly the outliers identified according to some criteria do not have to be only contaminants, which means they may be the results of the inherent variability of the data. Thus identification based on a single variable, which is the CGPA of the student in ÖSYM case, does not guarantee the detection of real contaminants. Changing school and entering into a new one, is analogous to contamination of the original model F with an observation coming from another distribution G .

Barnett and Lewis (1984) groups the aims in examining the outliers into two:

- i. *Accommodation*: Inferring characteristics of the basic model, F , thus robust methods of analysis, which minimizes the impact of outliers, will be relevant.
- ii. *Testing of discordancy*: Detecting outliers for further actions (such as rejecting them, modifying the original model F to incorporate those outliers, and so on)

OBP and AOBP differ significantly in terms of these two aims. OBP as a transformed form of CGPA distribution affects directly the placement score in the previous system, thus parameters used in standardization, sample mean and sample standard deviation, are very important. Consequently the impact of outliers on these parameters should be minimized in the ideal use of OBP in the previous system. In this case, accommodation techniques using robust location and dispersion (spread) estimators (such as Winsorized mean and variance) can be considered, since they are designed as robust estimators against the presence of outliers. However, AOBP, using OBP only for transformation purposes eliminates the direct effect on placement. The major concern in the new system is to detect the outlying students, which affects the transformation process (reader should note that the transformation mentioned is not the standardization procedure of (2.2); Section 2.2.3.3 provides the details). Therefore, discordancy testing is much more relevant for the outlier students problem in the current system.

Although there are various tests proposed for outlier detection in the literature, only a few of them have still considerable attention, since they are resistant to problems associated with the presence of multiple outliers in the sample data. These problems are *masking*: presence of other outliers makes the detection difficult, and *swamping*: tending to declare too many observations as outlier when there is at least one real outlier in the sample. The first test is the Rosner's (1983) Generalized Extreme Studentized Deviates procedure (which will be referred as GESD hereafter), which is an extended version of outlier t test. The second test is an extension of well-known boxplot rule of Tukey (1977) (which will be referred as Tukey's Rule hereafter). And the last test considered is an alternative version of Tukey's Rule, which uses Median instead of the quartiles in the original version, based on Carling (2000) (which will be referred as Median Rule hereafter).

2.3.1 GESD Rule (Rosner, 1983)

Let X_1, X_2, \dots, X_n be a random sample of size n , the corresponding k extreme studentized deviates ($R_i, i = 1, 2, \dots, k, k \leq n$) are defined iteratively as follows:

Step 1. Define $R_1 = (\max |X_i - \bar{X}|) / s$ considering the whole sample.

Step 2. Remove X_i corresponding to R_1 from the sample, update \bar{X} and s .

Repeat steps 1 to 2, k times; at each iteration removing one of the observations and updating the statistics. Note that k is the maximum number of outliers in the sample, which is determined by the user at the beginning.

If there are actually l outliers present in the data (i.e., the null hypothesis, H_l , claiming there are l outliers in the data, is correct), then the critical values of the procedure are found at a type I error rate of α as follows:

Find $\lambda_i, i = 1, \dots, k$ such that

$$\Pr \left\{ \bigcup_{i=l+1}^k (R_i > \lambda_i | H_l) \right\} = \alpha, \quad l = 0, 1, \dots, k-1$$

Rosner (1983), using the approximation of Quesenberry and David (1961), determines the critical values, λ_i , for the two-sided outlier problem with the following formula:

$$\lambda_{i+1} = \frac{(n-l-1)t_{n-l-2,p}}{\left\{ \left[n-l-2+t_{n-l-2,p}^2 \right] (n-l) \right\}^{1/2}}, \quad l=0,1,\dots,k-1,$$

where $p=1-(\alpha/2)/(n-l)$ and $t_{d,p}$ represents the p^{th} percentile of a t distribution with d degrees of freedom.

According to the simulation runs of Rosner (1983), the nominal α level (either $\alpha=0.01$ or $\alpha=0.05$) of the test is realized for $n \geq 25$. The most crucial drawback of GESD is the assumption that data comes from a Gaussian distribution.

2.3.2 Tukey's Rule

Probably, the boxplot rule of Tukey (1977) is one of the most common rules, described in various textbooks and implemented in current statistical packages such as SPSS, Minitab, SAS, Statgraphics and Systat (Frigge et al., 1989). Tukey (1977) introduces it as a rule of thumb to identify extreme observations, hence potential outliers. The rule is based on summary statistics *fourths*, which are approximately sample quartiles. The initial rule of Tukey (1977) is as follows:

Let X_1, X_2, \dots, X_n be a random sample of size n , and their corresponding order statistics be $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. Accordingly, lower and upper fourths are defined as $F_L = X_{(f)}$ and $F_U = X_{(n+1-f)}$ with $f = [(n+3)/2]/2$, where $[\cdot]$ denotes the greatest-integer function. f , defined as *depth*, locates the fourths from both ends of ordered sample. The difference between the fourths is defined as $F\text{-spread} = F_U - F_L$, which is approximately the interquartile range, a resistant measure of the spread of the data.

According to the above definitions, the Tukey's rule uses *lower and upper inner fences* (IF_L and IF_U) to identify potential outliers:

$$IF_L = F_L - k_T(F - \text{spread})$$

and

$$IF_U = F_U + k_T(F - \text{spread})$$

Any observations that fall below IF_L or above IF_U are assumed to be an outlier, where the initial choice of Tukey (1977) for k_T was 1.5. One of the advantages of the initial rule of Tukey (1977) was the resistance of its estimators. Another advantage of the rule was its applicability to a wide range of distributions characteristics, since it was designed for exploratory purposes not like a hypothesis test. Hoaglin et al. (1986) propose some distribution-independent measures to provide a basis for standardization of the rule. These are the *outside rate per observation*- p , which is the probability that a particular observation in the sample will be classified as outlier, and the *some-outside rate per sample*- p_n , which is the probability that any sample of size n contains at least one outlier observation (whether at only one end, or at both ends). Note that p_n is analogous to Type I error rate, α , used in hypothesis testing. Although holds only in asymptotic cases, the following relationship exists between these two measures:

$$p_n = 1 - (1 - p)^n$$

Hoaglin et al. (1986) approximate the some-outside rate per sample, p_n , for the initial rule of Tukey on Gaussian data via simulation. Hoaglin and Iglewicz (1987) propose alternative choices for k_T instead of the standard rule (i.e., 1.5), which fix the some-outside rate per sample, p_n (analogous to fixed significance levels in hypothesis tests). The recent work of Carey et al. (1997) proposes a calibration formula for k_T , so that outlier-free Gaussian samples will have a some-outside rate fixed at 0.01 and 0.05 (see Appendix B). The proposed estimators of k_T are valid for sample sizes between 10 and 3000.

Defining the quartiles, so-called fourths, is the other crucial point of the rule. Hoaglin et al. (1986) observe that some-outside rate per sample measure tend to separate according to sample size, n , having the form $4j$, $4j+1$, $4j+2$, or $4j+3$, which is probably the result of the definition of quartiles. Especially the measures differ significantly for small choices of n , since the effect of modularity decreases with increasing n . Frigge et al. (1989) discuss alternative definitions of fourths and recommend the use of *ideal* or *machine fourths* based on the work of Hoaglin et al. (1983), which is more resistant to the modularity than its alternatives. The depth, f , in the definition of the ideal fourth is $f = n/4 + 5/12$, where n is the sample size upper and lower fourths are defined as in the above case.

2.3.3 Median Rule (Carling, 2000)

This rule is similar to the previous one, however uses sample median rather than the first and third quartiles used in the previous rule. Carling (2000) defines the lower and upper inner fences as follows:

$$IF_L = M - k_M (F - \text{spread})$$

and

$$IF_U = M + k_M (F - \text{spread}),$$

where M is the sample median, and k_M is the constant defined accordingly. The two rules are very similar in the sense that both of them have a breakdown point roughly around 25% (breakdown point is defined as the maximum fraction of outliers in the sample that the rule can cope with), and for the asymptotic symmetric case they are equivalent, if $k_M = k_T + 0.5$ holds. Carling (2000) also uses the ideal fourth definition given in the previous rule to find the F-spread.

According to simulation results on samples generated from Generalized Lambda Distribution of Ramberg et al. (1979), Carling (2000) proposes the following formula for k_M in terms of upper outside rate per observation in a sample of size n (p^U), skewness (α_3), and kurtosis (α_4) as a reasonable fit to calculate constant k_M :

$$100p^U = 8.07 + \frac{3.71}{n} + \frac{17.63}{k_M} - \frac{23.64}{nk_M} + 0.83\alpha_3 + 0.48\alpha_3^2 + 0.48(\alpha_4 - 3) - 0.04(\alpha_4 - 3)^2$$

(2.5)

Carling (2000) adds a subscript of n to p^U in his original definition of upper outside rate per observation, however effects of the sample size, n , in formula (2.5) decrease as n increases. This means the terms with n are included as a correction factor for small sample sizes, and this subscript is removed so that the notation of same measure agrees with the one in Tukey's rule (p). Ramberg et al. (1979) propose the estimators for α_3 and α_4 as follows:

$$\alpha_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3 / n}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 / n \right]^{3/2}}, \text{ and } \alpha_4 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4 / n}{\left[\sum_{i=1}^n (X_i - \bar{X})^2 / n \right]^2}$$

Carling (2000) proposes the use of p^U for symmetric and right-skewed cases, since $2p^U \geq p^U + p^L$ holds for these two cases (where p^L denotes the lower outside rate per observation, and $p = p^U + p^L$ is the outside rate per observation measure defined in the previous rule). However, he misses the left-skewed distributions, which may also be observed in ÖSYM case.

Carling (2000) compares only the Median and Tukey rules, and concludes that Median rule has a better performance in terms of resistance and efficiency in the non-Gaussian cases. Especially his proposed estimator for k_M using the skewness and kurtosis of the data incorporates the distribution characteristics better than the alternative (Tukey); however, his work does not consider the well-known GESD procedure.

2.3.4 Discussion of the Rules

The first two rules introduced above (GESD and Tukey), are probably the most common ones used in the area of outlier detection. GESD and Tukey differ in their conceptual frameworks. GESD is basically a hypothesis test, and apparently it has Type I and II error rates for the Gaussian null hypothesis. Carey et al (1997).

describe GESD as a test-based detection rule with iterative peeling, since it consists of serial tests of studentized deviates against the corresponding critical values. The application of the test in the reverse order, which is also called *outward testing*, increases the test's performance against the masking errors. The only drawback of GESD is its strict dependence on the Gaussian assumption for the population.

With its robust location estimators (lower and upper fourths) against different distribution families, Tukey's rule is a member of resistant rules in the literature. Although it seemed to have an advantage over GESD because of its resistant estimators, Brant (1990) observes comparable performances of both types of rules in the non-Gaussian cases. Especially it is noted that applying GESD Rules with a large k (e.g., $k \approx n/5$) incurs little penalty in the pure Gaussian cases, whereas provides performance competitive with the resistant rules in the Non-Gaussian cases. Carey et al. (1997) again compare the performances of Tukey's rule and GESD rule under the Gaussian assumption, and observe a potential advantage of GESD against the calibrated resistant rules in some cases. They also recommend the use of calibration for resistant rules, which provides them advantage over the versions that are not calibrated.

Although the Tukey's rule was defined initially as a resistant rule, the works of Hoaglin and Iglewicz (1987) and Carey et al. (1997) aim to provide a basis for standardization, which uses the calibration of the fences to fix the rate at which outlier-free samples are erroneously declared to possess one or more outliers (analogous to Type I error rates in the hypothesis testing). Obviously the calibration efforts need a null hypothesis about the data, which also restrict the resistant nature of the Tukey's rule. The formula of Carey et al. (1997) uses an error-based rubric under the Gaussian assumption, which fixes the p_n value defined earlier in a similar way the GESD rule fixes the Type I error rate.

Median rule being a variation of Tukey's rule is also a member of resistant rules family, however its standardization rubric differs significantly from Tukey's. Median rule aims to fix the probability at which an observation in data set is erroneously

labeled as an outlier beyond the upper limit ($p^U = p - p^L$). ÖSYM is also using a similar standardization rubric, and fixes the same kind of a rate as the Median rule. If it is assumed that all of the schools follow a theoretical continuous distribution, then ÖSYM, limiting the distributions in the $[\mu - 2\sigma, \mu + 3\sigma]$ range, can be considered as a rule fixing the probability of an observation being labeled as an outlier to some p value ($p = p^L + p^U$), which is the outside rate per observation as defined above.

Obviously p of ÖSYM depends on the distribution characteristics, such as the skewness, and will differ between schools. Table 2.7 shows the behavior of ÖSYM rule in three different theoretical distributions. The rule tends to declare more outliers on the skewed tails as expected, however it declares more lower-outliers in the left-skewed distributions when compared to the upper-outliers declared in the right-skewed ones (lower-outlier is used to denote the observations that are labeled as outliers because they are below the lower limit, and upper-outlier is for the opposite end), which is because of the asymmetric limits ÖSYM.

Table 2.7: The behavior of ÖSYM's outlier detection method for three continuous probability distributions

Distribution	Skewness	p^L	p^U	p
Lognormal(0, 0.5)	>0	0.0000	0.0154	0.0154
Normal(0,1)	0	0.0228	0.0014	0.0242
Beta(3, 1.5)	<0	0.0363	0.0000	0.0363

These four outlier detection rules can be divided into two separate groups based on their standardization rubrics: (1) those fixing p_n (GESD and Tukey), and (2) those fixing p (Median and ÖSYM). Although the equality $p_n = 1 - (1 - p)^n$ between the two rates holds in the asymptotic case, it is obvious that fixing one of the rates means changing the other with changing sample sizes. Fixing some-outside rate per sample (p_n) means continuously decreasing the outside rate per observation (p) for increasing sample sizes; however, fixing p means increasing p_n with increasing sample sizes. Having different standardization rubrics means the two groups have

different aims in detecting the outliers, and hence a direct comparison of their performances is not meaningful.

Probably there is not a universal optimal solution to the problem of outlier detection, and any solution available will depend highly on the subjective judgment of the user, here ÖSYM. Although fixing some-outside rate per sample or Type I error rate is more common in the area of outlier detection, the choice of the standardization rubric is up to the user. Instead of trying to find the best method, it would be meaningful to analyze the impacts of the available policies and their fitness to the intended purposes.

Fixing p increases the expected number of outliers (np , n : sample size) with the increasing sample sizes. The increasing number of outliers means a decrease in the discriminating power of the OBP; as the number of outliers increases, the number of OBP's that are trimmed to the C and B values. Trimming increasing number of students at both ends of their distributions is only meaningful if the aim is to reward (and to punish) the highest (and the lowest) percentiles equally within their groups. In other words, the current policy of ÖSYM does not try to discriminate these two groups of people from each other. If it is believed that education and assessment are designed mainly for an average student, trimming the extreme percentiles can be considered as a meaningful policy, since students at these extremes may not be discriminated as well as those in the middle percentiles.

However, if the policy is to detect the real outlier students first, and then treat the rest of the data independent of these extreme students, then a policy fixing the measure p_n will be more suitable than fixing p . Fixing p_n will be a statistical hypothesis test conducted at a fixed error rate. Even the concept of trimming the extreme percentiles for educational assessment purposes may also be applied with a policy fixing p_n (i.e., ÖSYM may be indifferent between the students of extreme percentiles in term of educational assessment). For this purpose, some fixed levels of the upper and lower percentiles may be trimmed after detecting the real outliers.

2.4 Summary of Potential Problems

As a result of the analyses in the previous sections, two major types of problems that ÖSYM is confronted are identified as follows:

- i. The choice of lower and upper limit for CGPA distribution in schools.
- ii. The way the selection and placement criteria are chosen and weighted.

The first problem item includes the comparisons of theoretical outlier detection methods and the method of ÖSYM. The concept of outliers is one of the ill-structured areas of data analysis, where the subjective judgment of user is extremely important, and hence the rule of ÖSYM represents a policy on labeling extreme students as outliers. It is shown that ÖSYM's rule tries to fix the probability of labeling an observation as outlier, p , to some value. However, fixing the probability p generally cannot be achieved, because of the rule's lack of ability to deal with different distribution characteristics. In this respect some alternative outlier detection methods in the literature are considered, in order to provide a basis for comparison of the performance of ÖSYM's rule with respect to alternative approaches to outlier detection.

The second problem item consists of the analyses of various policies of ÖSYM on the choice of placement criteria and weights, such as moving from an achievement test dominant system to an aptitude test dominant system, and weighting past achievement to include the educational quality. The main concern of this item is the appropriateness of the chosen placement criteria to the intended purposes of ÖSYM. Especially the criteria chosen to weight the past academic achievement of students are subject to analysis. It is shown that the new AOBP system of ÖSYM is trying to balance two different dimensions of assessing the past achievement: quality of education of schools and local success in the schools independent of the educational quality.

CHAPTER 3

ANALYSIS OF REAL DATA

Since the analyses of problems stated explicitly in Section 2.4 are based on the CGPA distribution characteristics of schools obtained from ÖSYM, it would be meaningful to describe the methodology used for sampling the schools, and briefly identify the characteristics of schools in various categories before the detailed analysis of the problems.

3.1 Sampling Methodology

A small sample of the students from 46 schools in 5 main school categories is obtained in May of 2000 from ÖSYM for preliminary analysis from the year 1999 data (the first year of the operation of new system). A statistical sampling methodology is not used; instead schools are chosen from the main urban centers so that the location-specific disparities between students are minimized. The summary of initial data according to the school categories is given in Table 3.1.

Table 3.1: The summary of initial data according to the categories

School Category	Number of Schools	Number of Students
Science High Schools	9	697
Anatolian High Schools	9	1880
General High Schools	9	4359
Private High Schools	9	1374
Vocational High Schools	10	3850
Total	46	12160

Although the data is not chosen in a systematic way, it is very useful in the sense that some indicators of the potential problems are identified, especially for some school categories.

To avoid possible bias resulting from the inadequacy of the sample size and to have the chance to draw conclusions for the whole population, a new set of data is obtained. For this purpose, an updated list of populations of current seniors in all registered schools in Turkey is obtained from ÖSYM at the first stage. Although it is intended to obtain the new sample from the year 2000 graduates of the schools, the population summaries of schools for the current year 2001 are obtained. Assuming that there might be only minor differences between the data of two consecutive years, sample schools are determined according to year 2001 data.

A 10% sample is considered to be sufficient for all of the categories, except Science High Schools. Since it is expected that the Science High Schools be affected more by the past academic achievement assessment system of ÖSYM, all of the schools in this category are included in the sample

Although the sample obtained from ÖSYM consisted of more than 800 schools, some of them are eliminated since they do not satisfy the minimum student population requirement of ÖSYM (according to ÖSYM (2000b) a school should have at least 5 students for statistical calculations).

The second sample contains schools from 73 different categories (according to category definitions of ÖSYM, 2000), however only a few of them contain statistically significant numbers of schools. Therefore all those 73 categories are further summarized into 12 main categories for the purpose of analysis. The summary of the all schools in Turkey (according to year 2001 data), and the schools in the sample (year 2000 data), and descriptions of the school categories are given in Table 3.2. In general, the desired representation level of 10% is obtained in any category, except for Science High Schools category (5), where the representation level is almost 100%; for this reason, the estimators for the whole sample are

calculated in two different ways: all the 39 schools (*Overall₁* will be used to represent this category hereafter), and only a randomly selected sample of 4 schools of all 39 schools (*Overall₂* will be used to represent this category hereafter) in the Science High Schools category (5) are included to the whole sample statistics.

Table 3.2: Descriptions of school categories, and comparison of sample schools in these categories (year 2000 data) with the whole Turkey population (year 2001 data)

Category	Descriptions	Whole Turkey Population (2001)		Sample Schools (2000)		Representation Level (%) by Schools	Representation Level (%) by Students
		Number of Schools	Number of Students	Number of Schools	Number of Students		
1	General High Schools	2308	249613	203	24062	8.80	9.64
2	General High Schools with Foreign Language	802	40661	73	4551	9.10	11.19
3	Anatolian High Schools	341	23165	34	2992	9.97	12.92
4	Private High Schools	415	20186	40	1428	9.64	7.07
5	Science High Schools	40	2968	39	2912	97.50	98.11
6	Fine Arts and Teachers High Schools	185	7366	10	603	5.41	8.19
7	High Schools for Religious Education	550	45283	56	3558	10.18	7.86
8	Trade High Schools	503	63694	72	4526	14.31	7.11
9	Technical High Schools	421	8380	40	1106	9.50	13.20
10	Industrial Vocational Schools	604	81468	94	8134	15.56	9.98
11	Vocational High Schools for Girls	582	24315	70	2320	12.03	9.54
12	Other Vocational Schools	723	22911	67	3178	9.27	13.87
<i>Overall₁*</i>	Whole Sample population, additional 35 schools in category 5 are <u>included</u>	7474	590010	798	59370	10.68	10.06
<i>Overall₂*</i>	Whole Sample population, additional 35 schools in category 5 are <u>excluded</u>	7439	587450	763	56810	10.21	9.63

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

3.2 Analysis of Main School Categories

Summary statistics of those 12 main school categories can be seen in Table 3.3. These measures are chosen so that the CGPA distribution characteristics and central examination performances of schools in different categories can be understood better. Table 3.3 gives the averages of various school statistics summarized

according to different categories; the interested reader is referred to Appendix C, where further summaries for those statistics are also included. The statistics presented in Table 3.3 are: (1) sample mean CGPA (\bar{X}_{CGPA}^j , according to notation of Section 2.2.3.2), (2) sample standard deviation CGPA (s_{CGPA}^j , according to notation of Section 2.2.3.2), (3) sample skewness v^j , and (4) sample mean ÖSS-Equal Weighted scores (A_j^{EW} will be used to represent the average of ÖSS Equal Weighted scores of school j hereafter). Although the analyses are conducted for all three ÖSS composite score categories, only the results on Equal Weighted scores will be presented in the following sections because of the similarities of results with the other score categories.

Table 3.3: Summary statistics for 12 school categories and the whole sample

Category	Number of Schools	Number of Students	Average \bar{X}_{CGPA}^j	Average s_{CGPA}^j	Average v^j	Average A_j^{EW}
1	203	24062	3.143	0.653	0.526	104.854
2	73	4551	4.044	0.491	-0.210	128.748
3	34	2992	3.970	0.566	-0.239	138.323
4	40	1428	3.711	0.578	0.057	126.568
5	39	2912	4.511	0.297	-0.812	168.694
6	10	603	3.999	0.527	-0.234	121.996
7	56	3558	3.508	0.697	0.207	104.713
8	72	4526	3.283	0.594	0.397	100.014
9	40	1106	3.577	0.470	0.215	111.867
10	94	8134	3.138	0.522	0.408	100.231
11	70	2320	3.728	0.539	-0.007	99.416
12	67	3178	3.551	0.528	0.228	103.491
<i>Overall₁*</i>	798	59370	3.512	0.564	0.192	111.658
<i>Overall₂*</i>	763	56810	3.466	0.576	0.239	109.064

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

Skewness as a measure of symmetry is the most striking measure in the table. Especially for the successful (in terms of ÖSS scores average) schools, the skewness is increasingly negative, which means that their CGPA distributions are skewed to left. For example in the Science High Schools category (5) the average skewness is -0.812, and 36 of the 39 schools in this category have a negative skewness value (see Table C.3 of Appendix C). In general other categories have an average

skewness in $[-0.500, 0.500]$ range (10 of 12 categories), which means the corresponding distributions are more or less symmetric. On the other hand, the first category, General High School category, having the largest total number of students (almost 50% of students in the sample belong to category 1) have an average skewness higher than 0.500, meaning that the CGPA distributions of schools are skewed to right. OBP of ÖSYM having non-symmetric limits is probably designed with an expectation of only right-skewed student populations like the ones in category 1. This result shows that different school categories have different distribution characteristics, yet they are treated the same way according to the system of ÖSYM.

Another important observation related to the Table 3.3 is about the averages of CGPA standard deviations of schools (column 5). Almost all of the categories have average standard deviations between $[0.400, 0.600]$, however same measures for category 1, 5, and 7 are beyond the mentioned limits. Category 1 and 7 have higher CGPA standard deviations on the average (0.653 and 0.697), whereas category 5 has a lower average of 0.297. Since the standard deviation is measure of spread of the data, it can be concluded that allowable OBP range (i.e. range $5s$ between ÖSYM's limits of $[\bar{X} - 2s, \bar{X} + 3s]$) based on this measure will be wider for the schools having higher CGPA standard deviations, which may affect the probability of an individual student's CGPA falling beyond the allowable limits.

To learn the characteristics of categories better, CGPA distributions are fitted into some theoretical probability distributions using Kolmogorov-Smirnov¹ goodness of

¹ Law and Kelton (1991) argue that the original form of K-S Test (all-parameters known) can be applied directly for any continuous distribution with estimated parameters, and for discrete distributions, however, it produces conservative results with the probability of Type I error smaller than specified. They also give a formulation to adjust the test statistic D_n , based on the work of Stephens (1974), which obviates the need for large tables. Briefly the null hypothesis, (H_0 : the data is distributed by Normal) is rejected if:

$$\left(\sqrt{n} - 0.01 + \frac{0.85}{\sqrt{n}} \right) D_n > c_{1-\alpha}$$

where $c_{1-\alpha}$ are the critical values having a Type I error of α .

fit test (which will be referred as K-S Test hereafter). Since the K-S Tests do not require grouping the data into intervals (which is very problematic in small sample sizes), and they are valid for any sample size (in the all-parameters known case), they are chosen as the tool for goodness of fit testing. Although there are only a few theoretical distributions that K-S Tests can be applied in the all-parameters unknown case, these distributions are sufficient to represent almost all of the schools in the sample. Normal and Lognormal are the two distributions, for which K-S Tests can be applied, and into which the school distributions are probably fit. In the previous paragraphs it is mentioned that for a significant portion of the schools (10 out of 12 school categories), symmetry assumption could not be rejected according to the skewness estimates, thus Normal distribution can be assumed for these symmetrical cases. Again a significant portion of the data was found to be skewed right, which can be described by the Lognormal distribution. For some categories like Science High Schools, both of the above distributions would probably be inappropriate, however not fitting into these two distributions can be an indicator of a left-skewed distribution characteristic for those categories.

As a result of the test the schools are classified into one of the following four fitted distribution classes: *Only Normal*, *Only Lognormal*, both *Normal and Lognormal*, neither Normal nor Lognormal (*Neither*). The results of the tests summarized in Table 3.4 are only for $\alpha = 0.05$ case, since the results are similar in the other error rate of $\alpha = 0.01$ (see Appendix D for tables of distribution fitting results for $\alpha = 0.01$ and $\alpha = 0.05$ cases, where school sizes are also included)

In general 70% to 90% of the schools in all school categories are fitted into *Only Lognormal* and *Normal and Lognormal* categories. However, lognormal distribution assumption is not valid for Science High Schools category (5), where in 54% of total schools in this category, the assumption is rejected at $\alpha = 0.05$ level, and this result again supports the previous claim about the distribution characteristics of schools in this category.

Since a random variable X is distributed by Lognormal only if the natural logarithmic transformation ($\ln(X)$) of X is distributed Normally, the critical values $c_{1-\alpha}$ can also be used to fit the data into a Lognormal distribution.

Table 3.4: Kolmogorov-Smirnov goodness of fit test results for 12 school categories and the whole sample (for $\alpha = 0.05$ case) (where numbers in parentheses denote the proportion of schools in that category satisfying the distribution)

Category	Number of Schools	Number of Schools			
		Only Normal	Only Lognormal	Normal and Lognormal	Neither
1	203	4 (0,02)	49 (0,24)	97 (0,48)	53 (0,26)
2	73	9 (0,12)	2 (0,03)	52 (0,71)	10 (0,14)
3	34	3 (0,09)	1 (0,03)	18 (0,53)	12 (0,35)
4	40	2 (0,05)	2 (0,05)	26 (0,65)	10 (0,25)
5	39	4 (0,10)	0 (0,00)	14 (0,36)	21 (0,54)
6	10	1 (0,10)	0 (0,00)	7 (0,70)	2 (0,20)
7	56	2 (0,04)	5 (0,09)	40 (0,71)	9 (0,16)
8	72	2 (0,03)	12 (0,17)	47 (0,65)	11 (0,15)
9	40	0 (0,00)	1 (0,03)	37 (0,93)	2 (0,05)
10	94	5 (0,05)	19 (0,20)	56 (0,60)	14 (0,15)
11	70	2 (0,03)	2 (0,03)	60 (0,86)	6 (0,09)
12	67	4 (0,06)	6 (0,09)	54 (0,81)	3 (0,04)
<i>Overall₁*</i>	798	38 (0,05)	99 (0,12)	508 (0,64)	153 (0,19)
<i>Overall₂*</i>	763	35 (0,05)	99 (0,13)	495 (0,65)	134 (0,18)

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

To analyze the relationships between the parameters further, a correlation study based on school summary statistics is made for each school category, and the following classes of estimators are derived from the data:

- Correlation coefficients between \bar{X}_{CGPA}^j and s_{CGPA}^j
- Correlation coefficients between \bar{X}_{CGPA}^j and A_j^k , where $k = Q, V, EW$
- Correlation coefficients between s_{CGPA}^j and A_j^k , where $k = Q, V, EW$

Table 3.5 contains correlation estimates for the 12 major school categories and the whole sample for the Equal Weighted composite score category. The first estimate (estimates between \bar{X}_{CGPA}^j and s_{CGPA}^j at column 2), although not so strong, indicates that there is a negative correlation between sample mean and sample standard deviation of CGPA distributions, which means the schools with higher CGPA average, have generally lower CGPA standard deviation. This indicates that CGPA range is generally narrower in the schools where the CGPA average is high.

Considering the correlation estimates between \bar{X}_{CGPA}^j and A_j^{EW} , it can also be concluded that there exists a strong and undeniable positive correlation between the CGPA average and ÖSS score average, which means that schools with higher ÖSS score average have generally higher grades (column 3). Especially the relationship is very strong for Science High Schools category (5) with a correlation coefficient estimate around 0.83. The last class of correlation estimates is between sample standard deviation and ÖSS scores average, which are not strong enough to have a conclusion.

Table 3.5: Correlation of coefficient estimates between school parameters for 12 school categories and the whole sample population

Category	\bar{X}_{CGPA}^j vs. S_{CGPA}^j	\bar{X}_{CGPA}^j vs. A_j^{EW}	S_{CGPA}^j vs. A_j^{EW}
1	0.15230	0.22919	-0.18568
2	-0.67029	0.30247	-0.38259
3	-0.18051	0.55548	0.06368
4	-0.04329	0.67150	0.04533
5	-0.82428	0.82892	-0.71799
6	-0.67121	0.68877	-0.36674
7	-0.57453	0.60609	-0.45452
8	-0.32132	0.38399	0.11534
9	-0.20284	0.00917	0.43408
10	0.10916	0.68496	0.10912
11	-0.33607	0.54986	-0.05485
12	-0.44673	0.35004	-0.09162
Overall ₁ *	-0.43250	0.72531	-0.32979
Overall ₂ *	-0.31480	0.64343	-0.14065

*: Overall₁: Whole sample population, additional 35 schools in category 5 are included, Overall₂: Whole sample population, additional 35 schools in category 5 are excluded.

3.3 Problem I: Outlier Detection

As discussed earlier in the report, the outlier detection methods differ in their standardization rubrics. Since directly comparing any two methods from different standardization rubrics is not meaningful, the rules from different rubrics are tried to be defined in the other way, and the comparisons are made accordingly.

First comparison is based on the standardization rubric of ÖSYM and Median rules, which fixes p , the probability that an observation in a sample is labeled as an outlier. In application of the rules, it is assumed that the sample schools are distributed

Normally, even though there are a lot of cases in which this assumption is violated, as shown in Section 3.2. A hypothesis about the sample school distributions is a necessity; otherwise, applying the rules (except ÖSYM) is not possible since the p values cannot be defined. Thus, using the Normality distribution assumption, p value is set to 0.0241 (which is the p value of ÖSYM's rule under the Normal assumption, see Table 2.7) for Median and GESD rules (p^U of Median rule is set to 0.01205 because of the assumption of Normality, and hence symmetry, and p_n of GESD is approximated using the equation $p_n = 1 - (1 - p)^n$ where the p value is set to 0.0241). Since the calibration formula for the Tukey's rule was proposed only for two values of p_n (0.05 and 0.01), this rule cannot be used with a standardization rubric fixing p . According to the above applications of the three rules, p values of the sample schools are summarized for the 12 school categories and the whole sample in Table 3.6. In order to measure the deviation of the rules in different schools and categories the following two statistics are defined and presented in the table.

Mean Squared Deviation from Sample Mean is used to measure the deviation of the rules from the central tendency. Let \bar{p}_i be the sample mean of p values of schools of category i , and p_j be the p value of school j in category I , n_i be the number of schools in category i , then the statistic is as follows:

$$MSD_{\bar{p}_i} = \frac{\sum_{j=1}^{n_i} (p_j - \bar{p}_i)^2}{n_i}$$

To measure the deviation from the nominal p value of 0.0241 similar to the above one, the following equation is used for each category:

$$MSD_{0.0241} = \frac{\sum_{j=1}^{n_i} (p_j - 0.0241)^2}{n_i}$$

If the \bar{p}_i 's are examined from Table 3.6, it is obvious that both measures for ÖSYM and GESD deviate significantly from the nominal value of 0.0241 in almost all of the categories, where ÖSYM is below, and GESD is above the nominal level. However the same statistics for Median rule is closer to the desired level than its alternatives. It should also be mentioned that, in all of the three rules \bar{p}_i 's differ substantially between different school categories (e.g., Science High Schools category has the highest outside rate in all three rules). When the error levels are considered, ÖSYM is slightly superior to the alternatives in both types of the error measures. To understand the reason behind these different error behaviors, some of the sample schools are examined and it is observed that GESD rule (when trying to fix p) is very conservative against the deviations of sample data from unimodality (i.e., GESD rule rejects any observation, if it is a member of another mode in the tails).

Table 3.6: Average of p values (\bar{p}_i) and error rates for ÖSYM's method and 2 alternative methods

Category	ÖSYM			Median			GESD		
	\bar{p}_i	$MSD_{\bar{p}_i}$	$MSD_{0.02}$	\bar{p}_i	$MSD_{\bar{p}_i}$	$MSD_{0.02}$	\bar{p}_i	$MSD_{\bar{p}_i}$	$MSD_{0.02}$
1	0.0048	0.0001	0.0005	0.0138	0.0007	0.0008	0.0370	0.0033	0.0035
2	0.0240	0.0005	0.0005	0.0125	0.0008	0.0009	0.0203	0.0015	0.0015
3	0.0198	0.0002	0.0003	0.0045	0.0001	0.0005	0.0078	0.0010	0.0013
4	0.0097	0.0004	0.0006	0.0044	0.0003	0.0007	0.0324	0.0032	0.0032
5	0.0348	0.0003	0.0004	0.0237	0.0006	0.0006	0.0504	0.0030	0.0037
6	0.0163	0.0002	0.0003	0.0069	0.0001	0.0004	0.0114	0.0004	0.0005
7	0.0037	0.0001	0.0005	0.0060	0.0003	0.0006	0.0386	0.0047	0.0049
8	0.0048	0.0001	0.0005	0.0159	0.0011	0.0012	0.0475	0.0042	0.0048
9	0.0124	0.0005	0.0006	0.0212	0.0012	0.0012	0.0519	0.0040	0.0048
10	0.0091	0.0002	0.0005	0.0213	0.0015	0.0015	0.0425	0.0032	0.0036
11	0.0156	0.0005	0.0006	0.0243	0.0023	0.0023	0.0387	0.0047	0.0049
12	0.0075	0.0002	0.0005	0.0134	0.0007	0.0009	0.0319	0.0028	0.0029
<i>Overall₁*</i>	0.0110	0.0003	0.0005	0.0150	0.0010	0.0010	0.0365	0.0034	0.0035
<i>Overall₂*</i>	0.0100	0.0003	0.0005	0.0147	0.0010	0.0011	0.0359	0.0034	0.0035

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

In order to decrease the effects of Normality assumption on the sample schools, the CGPA's of schools are normalized via both natural logarithmic and square transformations, and runs similar to the above ones are made on the transformed data. Right-skewed distributions (i.e., schools with skewness estimator of 0.5 or greater) are normalized via logarithmic transformation ($X \rightarrow \ln(X)$), and left-skewed distributions (i.e., schools with skewness estimator of -0.5 or smaller) are normalized via square transformation ($X \rightarrow X^2$). Although normalizing the data exactly is not possible, with these transformations errors of all 3 rules decrease, as expected. GESD rule's performance for the whole sample is closest to the nominal level of 0.0241 at its application on transformed data, which is superior to two other rules, yet it has again the highest error rates (see Table E.1 of Appendix E).

The distribution fitting results presented in the previous section could also be used to normalize the data, however, they are not chosen because of their bias for the sample size (it is observed that the Kolmogorov-Smirnov goodness of fit test has generally rejected the Normal distribution assumption in the crowded schools, and has failed to reject the same assumption for small sizes). The deviation from unimodality assumption is another reason in using the skewness measure, since the K-S test is not designed to handle those kinds of disturbance in the data. [see Table D.1 and D.2 in Appendix D for the distribution fitting results with school sizes]

Since ÖSYM does not use any calibration for the outlier limits, applying its rule while fixing some-outside rate per sample (p_n) is not possible, and hence only the other three rules can be compared with the standardization rubric of fixing p_n . GESD and Tukey's rule are applied according to their original definitions, however, the p value of the Median rule is approximated similar to the approximation of p_n of the GESD rule in the previous standardization rubric using the equation $p_n = 1 - (1 - p)^n$. Like the previous comparison, the rules are initially applied to the data with a Normal distribution assumption, and the same transformation procedure is used to improve the results. MSD is again chosen as the measure of deviation from the central tendency, and nominal p_n levels (although the tests are conducted

for two different nominal levels of p_n , only $p_n = 0.05$ case will be presented). Since some-outside rate per sample is a sample based measure, MSD's of this measure are calculated only for the whole sample considering the deviations of p_n values for 12 school categories (i.e., since the nominal level of p_n is expected for any school category, the deviations of p_n values from the nominal level of 0.05 in each category are used to calculate the error terms). The p_n , and the MSD values for the initial data and data after Normalization procedure are summarized in Table 3.7.

Table 3.7: p_n values and error rates for three alternative outlier detection methods

Category	p_n for the initial data			p_n for the transformed data		
	Tukey	Median	GESD	Tukey	Median	GESD
1	0.1182	0.2118	0.1182	0.0345	0.1133	0,0345
2	0.0548	0.2877	0.1096	0.0000	0.1781	0,0411
3	0.0294	0.0588	0.0294	0.0000	0.0294	0,0000
4	0.0500	0.0500	0.0750	0.0000	0.0500	0,0500
5	0.3077	0.6410	0.3846	0.1795	0.5385	0,2308
6	0.0000	0.2000	0.0000	0.0000	0.0000	0,0000
7	0.0000	0.0536	0.0000	0.0000	0.0179	0,0000
8	0.0694	0.1806	0.0833	0.0139	0.0972	0,0000
9	0.0250	0.1000	0.0500	0.0000	0.1000	0,0000
10	0.1809	0.2766	0.1596	0.0957	0.2234	0,0745
11	0.0286	0.2143	0.0857	0.0143	0.1571	0,0429
12	0.0448	0.1194	0.0896	0.0448	0.0896	0,0448
<i>Overall₁*</i>	0.0890	0.2055	0.1078	0.0351	0.1378	0,0426
<i>Overall₂*</i>	0.0799	0.1874	0.0944	0.0288	0.1206	0,0328
<i>MSD_{p_n}</i>	0.0073	0.0242	0.0095	0.0027	0.0192	0.0039
<i>MSD_{0.05}</i>	0.0079	0.0464	0.0118	0.0031	0.0259	0.0038

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

Tukey's rule shows the best performance on the initial data; which is closest to nominal level of 0.05 and has the lowest error levels. The GESD rule is slightly worse than Tukey's on the initial data, but its performance has been improved after the transformation. Based on the results summarized in the Table 3.7, it is not easy to choose between the two rules (Tukey and GESD), since their results are not different enough in either of the case to claim statistical significance. However,

Median rule (when trying to fix p_n) seems to be inferior to the two alternatives in both of the data sets.

Because of similarities of performances of the four rules, choosing any of them is not meaningful. Especially for the standardization rubric fixing p , none of the three rules (ÖSYM, Median, GESD) performed well on the initial data (i.e. data before transformation); however, GESD rule has performed superior to other two rules on the transformed data, yet it has the highest error rates among the alternatives tested. All of the three rules tested in this rubric are poor in handling the different distribution characteristics. For the second standardization rubric fixing p_n , GESD and Tukey's rule have performed better than Median rule in both of the data sets (initial and transformed); however, again like the first rubric, the three rules (Tukey, Median, GESD) are not resistant against different distribution characteristics. When the implementation efforts of the rules are considered, GESD is a bit inferior to three other alternatives, which is because of its iterative application. However the difference is not, again, significant since all the runs of the tests are completed within a few minutes on a personal computer (the rules are coded using Microsoft Visual Basic 6.0).

3.4 Problem II: Placement Criteria and Weights

Changing the weights of assessment criteria radically, ÖSYM has several objectives. Although some of the objectives are not explicitly stated, the changes and the corresponding aims can be summarized as follows:

The first item listed in Table 3.8 is, probably, the most important difference between the two placement systems. In the following section the aim of ÖSYM in moving to an aptitude test dominant system will be analyzed. The second and third items are two different dimensions of ÖSYM's policy in assessing the past academic achievement.

Table 3.8: The changes in the weights of placement criteria and their corresponding aims

Change	Aim
<i>Elimination of ÖYS:</i> Moving from an achievement test dominant system to an aptitude test dominant system (i.e. ceasing the implementation of ÖYS)	Decreasing the private tutoring and study tactics based on solving more kinds and numbers of test questions, which is claimed to be negatively affecting the national education system, and hence the society and national economy in the long term (YÖK, 1999)
<i>Weighting the Past Academic Achievement:</i> Weighting past academic achievement with ÖSS scores average and OBP of the least successful student	Discriminating the schools based on their quality of education
<i>Fixing the AOBP Upper Limit to 80 for Every School:</i> Assigning the highest available past academic achievement score, 80, to the most successful students of every school	Rewarding the success in any school independent of the quality of education

3.4.1 Elimination of ÖYS

YÖK (1999) argues that the dominance of the achievement tests in the previous systems led students to concentrate more on studying multiple-choice test type questions instead of lectures in their high school curricula. The first type of study has short-term purposes directly (and maybe only) for placement, however the latter has a long-term purpose of educating the individual members of society. ÖSYM claims that the re-applicants, which constitute almost 2/3 of the total number of applicants (910,686 out of a total of 1,407,920 in 2000, ÖSYM (2001)), prepare mostly with the first type of study tactics, and lose the formation achieved during the high school education.

Making two crucial assumptions: (1) test-type study is mostly for preparation to achievement type exams, and (2) two types of exams (achievement-ÖYS and aptitude-ÖSS) are assessing almost the same abilities and learning outcomes, ÖSYM is no longer administering ÖYS in the current system.

High correlation between the two exam results in the previous years (stated as almost 100% by YÖK, 1999) can be an indicator of a close relationship between the two exams; however, it does not necessarily guarantee that these two tests are

assessing the same abilities, and it does not mean they can be used interchangeably. Indeed high correlation is an expected outcome, considering the fact that these two exams were administered within two months of time in the previous systems. Further analysis of this item necessitates a study considering the educational aspects of the problem, which is beyond the scope of this study, and hence it is left as a further research area.

Note that eliminating one of the components of the previous system, and moving its weight to other two components has also resulted in an increase in the weight of past academic achievement in the current placement process (as approximated in Section 2.2, the weight of academic achievement has increased for an average student in an average school from 8% to 20%). This increase is also in line with ÖSYM's aim to decrease test-type preparation for placement. YÖK (1999) states that they are planning to increase the weight of past academic achievement more in the future placement systems.

3.4.2 Weighting the Past Academic Achievement

ÖSYM, parallel to its efforts to increase the past academic achievement in the placement process, has also tried to improve the academic assessment procedure via including two weighting criteria, ÖSS score average of the schools and smallest OBP in the schools (A and C). As mentioned earlier, weighting the past achievement aims to incorporate the educational quality of the schools into AOBP, and hence the placement scores, which cannot be achieved with OBP type scores.

Table 3.9 summarizing the CGPA's needed to have equal OBP's in four different example schools in Ankara, will be useful to understand and highlight the pooriness of OBP in measuring the educational quality (the CGPA's are out of 5.00). The first two schools (1: a general high school, 2: a vocational school) do not have any special entrance examination however the last two (3: an Anatolian high school, 4: a science high school) admit their students according to a special entrance examination, and hence the levels of students and the education in these two schools are higher than the average (the averages of ÖSS scores in Equal Weighted category are given to

compare the education and student levels of the schools). To have the same OBP (such as 30, 50, or 70), students from different schools have to get CGPA's in a wide range (e.g., in category 2, a student has to get 1.96/5.00 to have an OBP of 30, whereas, a student in category 4 has to get 4.53/5.00 to get the same OBP). Having the same OBP does not necessarily mean that these four students are equivalent in terms of their academic knowledge; probably the students in "higher quality" schools have higher levels of knowledge. The difference is even more significant for higher OBP's: in the last two schools (3 and 4), where the average of CGPA's are high, having an OBP of 70 is impossible since the corresponding CGPA is higher than maximum available grade 5.00. This also means that the successful students of lower educational quality schools are rewarded more than the successful students of higher quality schools. For example, the most successful student in school 2 has CGPA of 4.73, and the corresponding OBP is $(4.73 - 3.06)/0.55 \times 10 + 50 = 80.214$, however the most successful student of the science high school (4) can only take $(5.00 - 4.83)/0.14 \times 10 + 50 = 62.206$, even though her CGPA is 5.00/5.00 higher than the other student. The disadvantage of the students in higher quality schools is obvious. Another conclusion from Table 3.9 is that OBP is not a suitable measure for inter-school comparisons.

Table 3.9: CGPA's needed to have equal OBP's in four sample schools in Ankara

School (<i>j</i>)	A_i^{EW}	\bar{X}_{CGPA}^j	S_{CGPA}^j	CGPA of least successful student	CGPA of most successful student	OBP=30	OBP=50	OBP=70
1	112.495	3.42	0.65	2.00	4.94	2.11	3.42	4.72
2	96.882	3.06	0.55	1.78	4.73	1.96	3.06	4.17
3	152.572	4.23	0.64	2.35	5.00	2.95	4.23	5.50
4	166.127	4.83	0.14	4.36	5.00	4.54	4.83	5.11

ÖSYM uses ÖSS score average A_i as a weighting criterion of the past achievement, which may be reasonable in the sense that it is an average of student performances in terms of a standard measure. However C , OBP of the least successful student, is only a measure of an individual, and it affects the whole school population to a certain extent. In order to understand the effects of this parameter the numerical example presented in Table 3.10 can be useful. If C value in a hypothetical school is

35 initially, and it is 30 after a transfer of a not so successful student to this school (a contaminant not representing the original population of the school), and if it is assumed that the new student in the school does not change the school parameters of A (assumed to be 110 points in both cases), B (assumed to be 80 points in both cases), mean and standard deviation of CGPA's, (which means the OBP's of the students do not change after the transfer), then α terms will be $(110/80)(35-8)+8=45.125$ and $(110/80)(30-8)+8=38.250$ respectively for the two cases. The AOBP of the least successful student of the first case (since school parameters do not change, student has an OBP of 35 in both cases) will be 45.125 $(45.125+(80-45.125)(35-35)/(80-35))$, and 42.425 $(38.250+(80-38.250)(35-30)/(50-30))$, in the first and second cases respectively. Then the relative disadvantage of the students of the school in the first case (i.e., all the students except the one transferred to the school) will be distributed in $[0, 45.125-42.425=2.700]$, where 0 corresponds to the highest ranking student, and 2.700 correspond to the lowest ranking student. With this amount of change (i.e., $2.700 \times 0.5 = 1.350$ points in terms of placement scores), a student may jump thousands of ranks, which means the chance to enter a more desired institution. Note that the disadvantage of all the students mentioned in this paragraph is because of transfer of a single student to school.

Table 3.10: Change of OBP's and AOBP's in an example school after the transfer of a student

Students	OBP		AOBP	
	School Before Transfer	School After Transfer	School Before Transfer	School After Transfer
1	80.000	80.000	80.000	80.000
2	77.000	77.000	77.675	77.495
.
.
n-1	37.000	37.000	46.675	44.095
n	35.000	35.000	45.125	42.425
n+1		30.000		38.250

There are two main objections against the use of C as a weighting criterion. First of them is the dependence of C value on an individual or a group of individuals. In the previous chapter, the dependence of C value on the sample size was especially emphasized. It was concluded that as the student population in a school increases,

the probability of observing a lower C value increases, which directly means the advantage of the school because of having a C value greater than 30, decreases (until the lower limit 30). In order to analyze this relationship, the schools in the sample are divided into two groups: (1) those having a C value of 30, and (2) those having more than 30 according to ÖSYM's method, and their histogram summaries for student populations are obtained. Table 3.11, categorizes the schools in two groups and gives the average size of a school in the sample, minimum and maximum sizes of schools, and histogram summaries for the whole sample (n in Table 3.11 denotes the size of a school, and additional statistics for the 12 school categories can be found in Appendix F). The relation with the sample size is very strong for the small sample sizes, for example 80% of total number of schools with a size of n in the $5 \leq n < 50$ range, there are no lower-outliers observed (lower-outlier is used to denote any outlier that falls beyond the lower limit 30), however this ratio decreases until 42% with the increasing sample sizes as expected (there is an unexpected sharp decrease to 29% for the range $200 \leq n < 250$). The relationship is not so strong for crowded schools, which is probably the result of the effects of distribution characteristics on the method, however the increasing trend in the first category and the decreasing trend in the second category cannot be ignored.

Table 3.11: Statistics for the size of schools (n) (average, minimum, and maximum school size values) and histogram summaries (numbers in parenthesis denote the proportion of schools satisfying the category in the corresponding range) for the whole sample in school category *Overall₂*, categorized according to the existence of lower-outliers by ÖSYM's method

Category*	Number of Schools	Average n	Minimum n	Maximum n	Histogram Summaries of Schools						
					$5 \leq n < 50$	$50 \leq n < 100$	$100 \leq n < 150$	$150 \leq n < 200$	$200 \leq n < 250$	$250 \leq n < 300$	$300 \leq n$
Lower-outlier(s) is detected	249	113.269	11	744	93 (0.20)	65 (0.47)	32 (0.52)	19 (0.58)	12 (0.71)	10 (0.56)	18 (0.58)
Lower-outlier(s) is not detected	514	55.654	5	616	370 (0.80)	74 (0.53)	30 (0.48)	14 (0.42)	5 (0.29)	8 (0.44)	13 (0.42)
Total	763	74,456	5	744	463 (1.00)	139 (1.00)	62 (1.00)	33 (1.00)	17 (1.00)	18 (1.00)	31 (1.00)

*: Only the schools in category *Overall₂* is presented.

The fact that C is an OBP score is the other objection against the use of it in the α term. Indeed one of the aims of the new system of ÖSYM is to eliminate the problems caused by the inadequacy of OBP type scores as a means of inter-school comparisons in the previous systems. However, the use of C in the α term means direct use of an OBP type score for the purpose of inter-school comparisons again (note that C and B in the ratio $(OBP_i - C_j)/(B_j - C_j)$ are not directly affecting the AOBP scores as shown in Section 2.2.3.3). Thus, incomparable OBP type scores are used to compare the schools in the current system, like they were used in the previous one.

In Table 3.11, it is shown that around 67% of the schools in the whole sample (514 out of 763 in category *Overall*₂), there are no lower-outlier students observed, which means all these schools have a C value greater than 30. The comparison here is between these two categories of schools, (1) those with C=30, and (2) those with C>30. For example, in the two most successful school categories in terms of ÖSS averages, Anatolian and Science High Schools (categories 3 and 5), only 24% and 5% of the total number of schools (8 out of 34, and 2 out of 39 respectively according to Table F.1 of Appendix F), respectively, benefit from the advantage of having a C value greater than 30. However, around 59% of schools in category 10 (General High Schools), benefit from this additional score advantage, even though they are the least successful school category in the sample in terms of ÖSS scores (e.g. ÖSS-Equal Weighted score average of schools is 96.993 for this category, whereas it is 154.407 for Science High Schools category, and 109.054 for the whole sample)

ÖSYM (2001) claims that they observe close relationships between the CGPA and ÖSS scores of individual students in the schools examined. This relationship is obviously one of the reasons of using ÖSS score average as a weighting criterion for past academic achievement. A similar correlation study has been conducted for the sample schools at hand, and the correlation coefficient estimates between ÖSS Equal-Weighted scores ($\bar{ÖSS}_i^{EW}$ is used to denote the ÖSS Equal Weighted score of

applicant i) and CGPA's of the students are summarized in Table 3.12. Note that this study is different from the one presented in Section 3.2, in the former one the correlation estimates were calculated from school parameters (A, CGPA average, etc.) for every school category, whereas now they are calculated from student parameters (ÖSS scores, CGPA, etc.) for every school.

Table 3.12: Coefficient of correlation (r) estimates between students' CGPA and ÖSS scores in Equal Weighted category summarized for 12 categories and the whole sample ($CGPA_i$ vs. $\bar{ÖSS}_i^{EW}$)

Category	Number of Schools	Average r	Minimum r	Maximum r	Histogram Summaries of Schools						
					$-1.0 \leq r < -0.2$	$-0.2 \leq r < 0.0$	$0.0 \leq r < 0.2$	$0.2 \leq r < 0.4$	$0.4 \leq r < 0.6$	$0.6 \leq r < 0.8$	$0.8 \leq r < 1.0$
1	203	0.550	-0.080	0.872		1	4	25	95	67	11
2	73	0.607	0.299	0.853				4	28	37	4
3	34	0.680	0.407	0.847					4	28	2
4	40	0.584	-0.341	0.881	1	2	3		5	23	6
5	39	0.634	0.373	0.834				1	12	23	3
6	10	0.540	0.204	0.763				2	4	4	
7	56	0.582	0.137	0.888			1	9	20	22	4
8	72	0.406	0.025	0.805			11	27	20	13	1
9	40	0.311	-0.607	0.861	2		8	18	5	6	1
10	94	0.263	-0.077	0.756		5	34	35	14	6	
11	70	0.439	-0.150	0.756		3	9	16	24	18	
12	67	0.449	-0.089	0.826		1	4	22	28	11	1
Overall ₁ *	798	0.492	-0.607	0.888	3	12	74	159	259	258	33
Overall ₂ *	763	0.485	-0.607	0.888	3	12	74	158	248	238	30

*: *Overall₁*: Whole Sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole Sample population, additional 35 schools in category 5 are excluded.

Table 3.12 gives the estimate averages, minimum and maximum observed estimates and histogram summaries of estimates for each school category. Almost 90% of the schools in the *Overall₂* category (674 out of 763) have an estimate between 0.2 and 1.0, which indicates a positive correlation. If the individual categories are examined, it is obvious that the relationship is very strong for the first seven categories, and weak for the last five categories, which are vocational and technical high schools. Again the highest values are observed for the Anatolian and Science High Schools categories (3 and 5). According to the Table 3.12 it can be concluded that, in general,

students with higher CGPA's have higher ÖSS scores, which means ÖSS as an aptitude test can be used to weight academic achievement especially for the first seven school categories. However, weighting the academic achievement with ÖSS may not be meaningful in the last five categories (vocational and technical schools), because of inappropriateness of ÖSS as a means to assess the abilities and learning outcomes obtained in these schools. Therefore, the main component of the current system ÖSS (both as a weighting criterion of past achievement and main component of the placement system) is poor in placing the applicants from vocational and technical schools into higher education institutions (the correlation estimates for ÖSS Quantitative and Verbal score categories are similarly poor for the last five school categories, see Tables G.1 and G.2 in Appendix G).

3.4.3 Fixing the AOBP Upper Limit to 80 for Every School

By fixing the AOBP upper limit to 80 points for every school, ÖSYM is trying to include the other aspect of the problem of centrally assessing the past achievement: local success independent of the educational quality.

In order to observe the implications of this policy, a summary of CGPA's, OBP's, and AOBP's of the least and most successful students in two schools from İzmir is given Table 3.13. The first school is the İzmir Science High School, which is one of the most successful schools in Turkey in terms of ÖSS score average in Equal Weighted score category, the other school is an average general high school in İzmir. The increase in AOBP's of the science high school because of the new system is obvious; both the least and the most successful students have increased their academic achievement scores drastically. Although there is an increase in the scores of the other school, it is not at the same amount observed in the science high school.

Table 3.13: CGPA, OBP, AOBP Comparisons for two example schools

	İzmir Science High School		Other School	
	Least Successful Student	Most Successful Student	Least Successful Student	Most Successful Student
CGPA	4.06	5.00	1.16	5.00
OBP	30.000	62.352	30.000	80.000
A^{EW}	167.354		113.919	
$AOBP^{EW}$	53.111	80.000	39.328	80.000

It can be concluded that ÖSYM is indifferent between the most successful students coming from different schools in terms of their AOBP, and hence, their past academic achievement. The two most successful students in Table 3.13 have the same AOBP, however this does not necessarily mean that their knowledge about the past academic work is equivalent. Although measuring the true levels of academic knowledge may not be possible, ÖSS scores can be used to measure the difference between the most successful students of different schools.

If it is assumed that the true level of academic knowledge of the most successful students in a school can be measured by the average of their ÖSS scores, then this standard measure can be used to analyze the implications of the policy of ÖSYM.

For this purpose the averages of ÖSS Equal Weighted scores of the students in the highest 5% of their school's CGPA distributions are calculated from the sample data and presented in Table 3.14 for each category. In all of the schools, the students in the upper 5% of their school will have an AOBP of 80 (or very close to 80). In other words, ÖSYM rewards the students in these groups with equal past achievement scores, however as shown in Table 3.14 there are significant differences between the schools.

Especially the last five school categories, which consist of mainly vocational and technical high schools, differ significantly from the first seven. The most successful students these vocational and technical schools have generally lower ÖSS scores than the most successful students of the schools in the first seven categories (e.g., in 64% of the schools, 129 out of 203, in the General High Schools category- 1, most successful students have an ÖSS average of 120 points or more; however, the same ratio falls down to 3% (3 out of 94) in Industrial Vocational Schools category- 10). The answer to why this difference occurs between the school categories is probably the inappropriateness of the ÖSS as a means to measure the learning outcomes of the vocational and technical schools.

Table 3.14: Average, minimum and maximum values and histogram summaries of arithmetic mean of ÖSS scores of the students in the upper 5% of their school's CGPA distributions ($\mu_{95\%}$), for 12 school categories and the whole sample

Category	Number of Schools	Average $\mu_{95\%}$	Minimum $\mu_{95\%}$	Maximum $\mu_{95\%}$	Histogram Summaries of Schools									
					$80 \leq \mu_{95\%} < 90$	$90 \leq \mu_{95\%} < 100$	$100 \leq \mu_{95\%} < 110$	$110 \leq \mu_{95\%} < 120$	$120 \leq \mu_{95\%} < 130$	$130 \leq \mu_{95\%} < 140$	$140 \leq \mu_{95\%} < 150$	$150 \leq \mu_{95\%} < 160$	$160 \leq \mu_{95\%} < 170$	$170 \leq \mu_{95\%} < 180$
1	203	123.642	92.577	166.380		6	22	46	64	51	11	2	1	
2	73	144.469	123.726	159.607					7	13	30	23		
3	34	153.703	132.546	167.120						3	8	13	10	
4	40	143.622	92.693	169.350		2	2		5	3	11	9	8	
5	39	163.937	141.874	170.298							2	3	33	1
6	10	137.871	119.270	165.803				1	2	3	3		1	
7	56	125.965	107.114	160.890			2	15	20	14	4		1	
8	72	111.223	91.242	147.716		9	30	20	8	3	2			
9	40	116.180	91.250	160.594		3	13	11	6	4	2		1	
10	94	104.217	87.893	150.485	2	38	33	12	6	1	1	1		
11	70	111.283	92.953	134.713		7	27	23	10	3				
12	67	117.709	91.960	149.406		3	14	24	19	5	2			
<i>Overall₁*</i>	798	124.775	87.893	170.298	2	68	143	152	147	103	76	51	55	1
<i>Overall₂*</i>	763	122.997	87.893	169.350	2	68	143	152	147	103	74	48	26	

*: *Overall₁*: Whole Sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole Sample population, additional 35 schools in category 5 are excluded.

CHAPTER 4

COMPARISON OF SOME ALTERNATIVES

As a result of the analyses in the previous chapters, basically two alternative methods aiming to overcome the drawbacks of the current method of ÖSYM are developed. The proposed alternatives have some common properties, however their approaches to some aspects of the problem differ. The common point between the alternatives is the elimination of C's use as a weighting criterion in the lower limit of AOBP. Note that, the elimination of C is independent of the outlier detection method that determines its value. The objections stated in the previous chapters are against the use of a parameter that is not in a standardized form as the other criterion weighting the α term- A. For this purpose, all the alternatives use $\{A, 30\}$ pair instead of the $\{A, C\}$ in the calculation of α .

The proposed alternatives differ from each other with their policies of outlier detection. The first alternative is the case where the outlier detection method is not changed (i.e., ÖSYM's rule of $[\bar{X} - 2s, \bar{X} + 3s]$ is used). The second alternative includes the three other outlier detection methods, and a trimming policy because of some educational assessment purposes.

As previously stated, one of the reasons of ÖSYM fixing p , and hence increasing the expected number of outliers with the increasing sample sizes (i.e., np), can be to trim some fixed fractions of the whole school population because assessing and

discriminating them is not meaningful (i.e., it is assumed that ÖSYM is indifferent between the students in the upper and lower 2 percentiles of the schools). Therefore in the second alternative, first the outliers are detected with a policy fixing some-outside rate per sample measure (p_n), and then a $2+2=4\%$ trimming policy is applied for the rest of the school population. Although choosing any trimming percentage is possible, 4% is used in the study.

Since comparison of the proposed alternatives with the ÖSYM's current method is the main concern of this chapter, the difference between the methods are measured in terms of the individual change in the AOBP's of the students. In the following sections, *net difference* between the AOBP's of students are summarized for the 12 school categories and for the whole population only for Equal Weighted ÖSS composite score category. The net difference is used to represent the net change in the AOBP of an individual with respect to the net change of the whole population (e.g. if the AOBP of a student decreases less than the average decrease in the AOBP's of whole population, then the net difference for that student will be a positive term since the net difference for the whole sample will be set to 0). The interested reader is referred to Appendix H, where the real differences are presented. The real difference between the initial AOBP (calculated according to ÖSYM's original method- $AOBP_{i,0}$) and final AOBP (calculated according to alternative methods- $AOBP_{i,1}^k$, $k = \text{ÖSYM, Tukey, Median, GESD}$) of a student i in school j is defined as follows:

$$AOBP_{i,1}^k - AOBP_{i,0} = \left[\alpha_{j,1}^k + (80 - \alpha_{j,1}^k) \frac{OBP_i - C_j^k}{B_j^k - C_j^k} \right] - \left[\alpha_{j,0} + (80 - \alpha_{j,0}) \frac{OBP_i - C_j^{\text{ÖSYM}}}{B_j^{\text{ÖSYM}} - C_j^{\text{ÖSYM}}} \right] \quad (4.1)$$

The α terms used in the formula are $\alpha_{j,1}^k = (A_j/80)(30-8)+8$, and $\alpha_{j,0} = (A_j/80)(C_j^{\text{ÖSYM}}-8)+8$ respectively, whereas C_j^k and B_j^k denote the C and B values for school j determined according to the outlier detection rule of k .

Before analyzing the alternative methods, it would be useful to analyze the placement score (YÖSS) distribution of the sample, since Y-ÖSS is the main factor on the effects of the net difference on the rankings of the students. Figure 4.1 shows the distribution of YÖSS scores of the whole sample (note that coefficient $c_{i,j}^{AOBP}$ - according to notation of Section 2.2.3.4, is taken as 0.5 although it does not necessarily be that value for all the schools), YÖSS score averages of 12 school categories and the whole sample for the Equal Weighted category. Obviously, the same amount change of AOBP will not affect the ranks of students in different categories the same way. If a student is closer to the mode of the YÖSS distribution (around 130 points), her jump in ranks as a result of the same amount of change in AOBP will be higher (e.g., the jump of general high school student- category 1, will be greater than a science high school student- category 5).

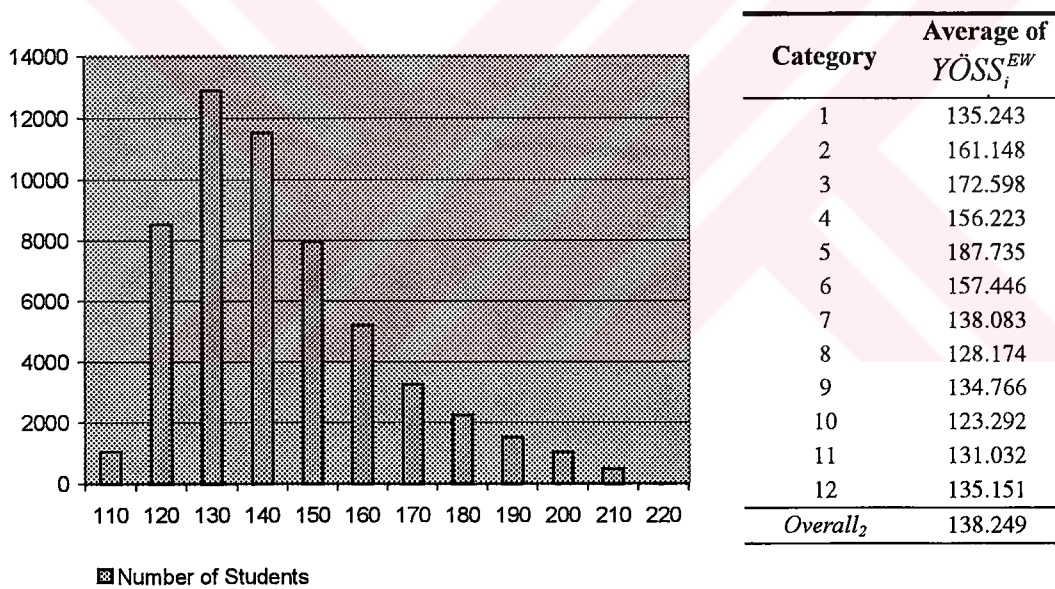


Figure 4.1: Histogram summaries of placement scores of students in the sample, and the placement scores average for 12 school categories and the whole sample

4.1 Alternative I: Outlier Detection Method of ÖSYM

Since this alternative is for showing the difference caused only by the change in the calculation of lower limit α , the outlier detection for the alternative is made according to ÖSYM's method (i.e., $k = \text{ÖSYM}$). Then, equation (4.1) is reduced to the following equation:

$$AOBP_{i,1}^{\hat{O}SYM} - AOBP_{i,0} = (\alpha_{j,1}^{\hat{O}SYM} - \alpha_{j,0}) \left[1 - \frac{OBP_i - C_j^{\hat{O}SYM}}{B_j^{\hat{O}SYM} - C_j^{\hat{O}SYM}} \right] \quad (4.2)$$

Table 4.1 summarizes the average net difference, minimum and maximum net differences, and the histogram summaries for the 12 school categories and the whole sample. Since the C value is the only factor changed between the alternative method and $\hat{O}SYM$'s original method, AOBP's in schools that have an initial C value of 30 (i.e., $C_j^{\hat{O}SYM} = 30$) are not decreased, and hence rewarded in terms of the net difference (note that 249 of the total 763 schools constitute this group, column 4). According to the histogram summary for the individual net differences, it can be concluded that the distribution of net differences are highly skewed-to left, where the greatest portion of the total population (40,115 of 56,810 students) have a non-negative net difference.

Table 4.1: Average, minimum and maximum values, and histogram summaries of the net difference (Δ) of students according to the first alternative method, summarized for 12 school categories and the whole sample

Category*	Number of Schools			Average Δ	Minimum Δ	Maximum Δ	Number of Students	Histogram Summaries of Students						
								$\Delta < -12.5$	$-12.5 \leq \Delta < -10.0$	$-10.0 \leq \Delta < -7.5$	$-7.5 \leq \Delta < -5.0$	$-5.0 \leq \Delta < -2.5$	$-2.5 \leq \Delta < 0.0$	$0.0 \leq \Delta < 2.5$
1	203	163	40	-0,220	-12,363	1,175	24062	39	213	594	2568	4934	15714	
2	73	22	51	0,901	-11,212	1,175	4551	5	9	25	66	215	4231	
3	34	8	26	1,020	-4,855	1,175	2992				18	103	2871	
4	40	30	10	-0,786	-11,763	1,175	1428	6	21	67	235	323	776	
5	39	2	37	1,094	-4,692	1,175	2912				19	17	2876	
6	10	4	6	0,708	-3,959	1,175	603				23	61	519	
7	56	47	9	-0,458	-11,055	1,175	3558	40	68	198	303	586	2363	
8	72	60	12	-0,471	-8,944	1,175	4526			18	175	463	1533	2337
9	40	27	13	-0,962	-13,130	1,175	1106	5	15	58	68	112	188	660
10	94	63	31	0,381	-13,254	1,175	8134	5	19	54	156	315	1200	6385
11	70	41	29	0,266	-12,112	1,175	2320		10	22	40	87	393	1768
12	67	49	18	-0,150	-12,774	1,175	3178	2	18	40	99	256	624	2139
Overall ₁	798	516	282	0,047	-13,254	1,175	59370	12	152	503	1422	4465	10177	42639
Overall ₂	758	514	249	0,000	-13,254	1,175	56810	12	152	503	1422	4446	10160	40115

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

4.2 Alternative II: Outlier Detection Methods with Trimming

In the previous section, the origin of the difference between the alternative and ÖSYM's method is the calculation of α terms. However, in this alternative a change in outlier detection method affects the net differences. Note that the schools with a previous C value of 30 have a potential advantage again (i.e., $C_j^{\text{ÖSYM}} = 30$). An important point that should be noted is about the real differences. In the previous section, the real changes are always negative or zero, which means the AOBP scores of the students have either decreased or remained the same. However, in this approach, positive changes in AOBP's are observed. This is because of the positive effects of outlier detection methods on the AOBP's of students. Especially in some schools, where the previous C value were 30, the AOBP's increase because of this effect.

As previously stated, trimming the upper and lower percentiles can be a meaningful policy, especially when the design and conduct of education systems are considered. A decision maker can be indifferent between the students that belong to a fixed upper (or lower) percentile in terms of educational assessment. In the current approach, it is assumed that one of the reasons of ÖSYM using a fixed p policy is the indifference of the assessment mentioned. Thus, in addition to the outliers detected according to the rules fixing p_n , a trimming policy may also be applied to find the C and B values. The approach followed for outlier detection rule of k in school j (k = Tukey, Median, GESD) is:

Step 1. Detect the outlier CGPA's according to rule k

Step 2. Assign CGPA's corresponding to the 2nd and 98th percentiles from the outlier-free sample to C_j^k and B_j^k (since $2 + 2 = 4\%$ is chosen as the trimming percentiles), interpolate between consecutive CGPA's when necessary.

Using both a trimming policy and outlier detection methods fixing p_n increases the resistance of C and B values against different distribution characteristics, which cannot be achieved by the rules trying to fix p (see Section 2.3.4). Although the rules

fixing p_n , are not that resistant against different distributions, they tend to declare very low numbers of outliers, which can be counted as real outliers. Thus, their application with a trimming policy is more robust than the current method of ÖSYM, and less prone to the errors originated from the different characteristics of distributions.

The difference function (4.1) can be defined as a linear function of OBP of an individual student. Thus examining this function will provide an insight about the behavior of difference function in the schools. The following is the reduced difference function:

$$f(OBP_i) = \left(\frac{80 - \alpha_{j,1}^k}{\Delta_j^k} - \frac{80 - \alpha_{j,0}^k}{\Delta_j^{\hat{O}SYM}} \right) OBP_i + (\alpha_{j,1}^k - \alpha_{j,0}^k) - (80 - \alpha_{j,1}^k) \frac{C_j^k}{\Delta_j^k} + (80 - \alpha_{j,0}^k) \frac{C_j^{\hat{O}SYM}}{\Delta_j^{\hat{O}SYM}}$$

where $\Delta_j^k = B_j^k - C_j^k$ is the difference between the B and C values calculated according to the outlier detection rule k ($k = \hat{O}SYM$, Tukey, Median, GESD) for school j . If OBP^* is defined as follows: $f(OBP^*) = 0$, then the behavior of difference of AOBP's in school j can be modeled as in Table 4.2 ($\text{sgn}(\cdot)$ is the sign of the function, and $k = \text{Tukey, Median, GESD}$):

Table 4.2: The behavior of the difference function (4.1)

	$OBP^* \leq C_j^k$	$C_j^k < OBP^* < B_j^k$	$OBP^* \geq B_j^k$
$\text{sgn} \left(\frac{80 - \alpha_{j,1}^k}{\Delta_j^k} - \frac{80 - \alpha_{j,0}^k}{\Delta_j^{\hat{O}SYM}} \right) = +$	Case3: All the students negative	Case 4: Lower portions are negative; upper portions are positive	Case 1: All the students positive
$\text{sgn} \left(\frac{80 - \alpha_{j,1}^k}{\Delta_j^k} - \frac{80 - \alpha_{j,0}^k}{\Delta_j^{\hat{O}SYM}} \right) = -$	Case 1: All the students positive	Case 2: Lower portions are positive; upper portions are negative	Case3: All the students negative

Table 4.3 categorizes the sample schools into two groups according to their C values (those with $C_j^{\hat{O}SYM} > 30$, and those $C_j^{\hat{O}SYM} = 30$) and into four groups according to the categorization in Table 4.2, and gives the average net difference, minimum and maximum net differences, and histogram summaries of the net differences only for

the 12 school categories (since the results of the other rules are similar only the Tukey's rule is presented in Table 4.3, for the other methods please refer to the Tables H.5 and H.6).

When the individual categories are concerned, the findings in Table 4.3 are similar to the ones presented in the previous section: the highest increase in the average net difference is again observed for the Science High Schools category (5). However, the general pattern of the net differences is different from the one in the previous section. Now the distribution of net differences is almost symmetric, and the net differences are distributed in a wider range.

Table 4.3: Averages of net difference (Δ) of students for the second alternative method only for the Tukey's case, summarized for 12 school categories and the whole sample

Category	Number of Schools	C>30				C=30				Average Δ	Minimum Δ	Maximum Δ	Histogram Summaries of Students							
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4				$\Delta < -15$	$-15 \leq \Delta < -10$	$-10 \leq \Delta < -5$	$-5 \leq \Delta < 0$	$0 \leq \Delta < 5$	$5 \leq \Delta < 10$	$10 \leq \Delta < 15$	$15 \leq \Delta < 20$
1	203	0	0	88	75	8	0	0	32	-0,569	-12,751	10,650	48	1414	12372	9699	528	1		
2	73	0	0	19	3	47	0	3	1	1,657	-11,600	13,505	3	32	370	4072	72	2		
3	34	0	0	4	4	23	0	2	1	1,160	-5,243	5,420		1	212	2772	7			
4	40	0	0	26	4	8	0	0	2	-0,757	-12,151	18,448	6	134	562	693	28	3	2	
5	39	0	0	1	1	31	0	5	1	1,901	-5,080	7,048		1	210	2629	72			
6	10	0	0	3	1	5	0	0	1	1,221	-4,347	6,495			106	490	7			
7	56	0	0	36	11	6	0	0	3	-0,877	-11,443	5,396	44	307	1509	1690	8			
8	72	0	0	41	19	8	0	0	4	-0,618	-22,909	8,426	6	3	336	2244	1819	118		
9	40	0	0	25	2	13	0	0	0	-0,456	-13,519	9,766	25	131	295	636	19			
10	94	0	0	44	19	11	0	1	19	0,719	-15,147	17,829	1	21	299	2895	4191	708	12	7
11	70	0	0	38	3	27	0	0	2	0,571	-12,500	8,812	11	79	590	1625	15			
12	67	0	0	41	8	13	0	0	5	0,434	-13,162	20,546	19	169	1061	1665	263			1
Overall ₁	798	0	0	366	150	200	0	11	71	0,078	-22,909	20,546	7	180	2903	22426	31981	1845	18	10
Overall ₂	763	0	0	365	149	173	0	6	70	0,000	-22,909	20,546	7	180	2902	22216	29704	1773	18	10

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

CHAPTER 5

CONCLUSIONS AND FURTHER RESEARCH

Focus of the study presented in this report is on the central placement problem of ÖSYM, and specifically on the problem of assessing the applicants' past academic achievement. According to the current system of ÖSYM, the placement decisions are based on two main criteria: past academic achievement and success in a national test. National tests are standard measures of students' ability and knowledge, however the information provided by these tests is limited. Since central administration of these tests necessitates the use of selection-type questions (specifically multiple-choice questions in the ÖSYM case), it is not possible to assess all the learning outcomes and skills (e.g., ability to present and organize ideas). Therefore consideration of past academic achievement in placement decisions could lessen the deficiency associated with the use of national tests as the sole criterion, and hence increases the fairness of the system.

The problem is in the central assessment of the past academic achievement using the CGPA's of the individual students. Obviously, CGPA is only a local measure of academic achievement, and hence, to calculate a universal measure using CGPA need to be incorporated some other factors such as the educational levels of the schools. In the previous system, past achievement was assessed directly using OBP, which was not a suitable measure for inter-school comparisons of success. However, AOBP of the current system tries to overcome this problem.

AOBP is a score that considers both dimensions of the problem of centrally assessing past achievement: educational quality and local success independent of the quality. Ignoring the quality of education, which may differ among the schools, and consequently treating all the schools in the same manner, which is the problematic side of OBP type scores, creates disadvantageous situations against the schools with higher educational quality. The natural consequence of this policy was the mass-departures of students from these schools. AOBP includes this dimension via using a varying lower limit α , which is determined according to two sub-criteria: ÖSS scores average of the students, A , and the OBP of the least successful student, C .

The other dimension is related to the problem of rewarding success in any school without considering the educational quality. Unlike the varying lower limit, ÖSYM uses a fixed upper limit of 80 for AOBP's, and assigns the most successful students to this limit in every school. Since the success in any school is rewarded with the same prize, this fixed upper limit policy handles the second dimension of the problem: assessing local success independent of quality.

It is shown that, ÖSS scores and CGPA's of the individual students have generally positive correlation in most of the school categories. Although the relation is not strong in some of the school categories (especially for the technical and vocational schools), the existence of a positive correlation means that it is meaningful to weight past academic achievement with A . However, the use of C as a weighting criterion has some questionable implications, because of the following reasons:

- i. C is an individual parameter, however its use in the first part of the AOBP formula, in weighting the α term, affects all the students in a school.
- ii. C is originally an OBP, which cannot be used as basis for inter-school comparisons; however the way it is used in weighting the α term directly serves for inter-school comparisons.

The two points against the use of C are independent of the outlier detection methods used. If the least successful student is not a lower-outlier (i.e., C is greater than the lower limit 30), then the hypothetical C value tends to decrease (down to the lower limit) as the size of the school increases, since the probability of observing a more

extreme student increases (a disadvantage for the students in crowded schools). Elimination of C 's use in calculating the α term is proposed as an improvement. Thus $\{A,C\}$ pairs in the α term are replaced with $\{A,30\}$ pairs for each school in the alternative methods proposed. Although both of the alternatives proposed use 30 instead of a varying C in the α term, they are still prone to the individual effects because of the transformation based on the ratio $(OBP_i - C_j)/(B_j - C_j)$ in AOBP. For the example school presented in Table 3.10, if the transferred student is not detected as an outlier, then it will again affect all the students in the school, however this time it will increase all the AOBP's in the school. The effects are summarized in Appendix I; it is obvious that all the students in the school have benefited from the transfer of the student according to the first proposed method. Therefore the method used to determine the lower and upper limits, and hence C and B values, has a very crucial importance.

The method of ÖSYM used in handling the outlier students has also been subject to analysis, which is fixing the outside rate per observation (p : probability that a student in a school will be labeled as an outlier). Although fixing p is not a hypothesis test, it may be used as a rule to limit the CGPA distributions. However it is shown that, ÖSYM's method is not successful in handling different distributions, and hence cannot fix the p value for different schools (e.g., probability of labeling a student in a Science High School as an outlier is higher than the probability of labeling a student in a General High School)

As alternatives to ÖSYM's method, some tests based on the concepts of hypothesis testing are also considered. These alternative outlier detection rules aim to detect the outlying observation while fixing some-outside rate per sample (p_n : probability that in a school of size n one or more students will be labeled as outliers). Although these alternative outlier detection rules have similar errors because of the different distribution characteristics of the schools, their applications with a trimming policy have considerable performances (here it is assumed that ÖSYM uses a fixed p policy, because of its indifference between the students in the upper percentiles and between the students in the lower percentiles of schools in terms of educational assessment).

Since the rules fixing p_n aim to detect the real outliers, they declare very low numbers of outliers when compared to the rules applied with a policy fixing p . Especially GESD and Tukey Rules of Section 2.3 have performed better than the alternative Median Rule in the standardization rubric fixing p_n .

In the application of outlier detection methods (both ÖSYM and alternatives), normalization of the data ($X \rightarrow X^2$ transformation for left-skewed schools, and $X \rightarrow \ln X$ transformation for right-skewed schools) has improved performances of the rules considerably.

Although the main focus of this study is on the problem of assessing the academic achievement, it would be meaningful to point out some other problematic areas of the central placement system as areas for further research. Especially the weights of decision criteria were drastically changed as the new system gave up the use of the main component of the previous system, ÖYS. Without the achievement test of the previous system, the new central placement system can be considered as an aptitude test dominant system. Obviously the two tests are not measuring the same abilities and learning outcomes, although their results have been highly correlated. In this respect additional validity studies should be conducted to define the intended purposes of these two exams.

The policy of ÖSYM to balance the quality of education and local success is an act to increase the fairness of the system, however the balance is obtained at the expense of decreasing the discriminating power of AOBP, the higher the quality of education in a school, the narrower the $[\alpha, 80]$ range will be. The next question that should be answered is: whether AOBP is adequately squeezing the students in the higher quality schools, or whether it is adequately separating the students in the lower quality schools (i.e., the choice of $1/80$ as a coefficient of A in the α term). The study presented in this report has not tried to answer this question about the weights of the criteria, indeed it only aims to answer, whether the chosen criteria is serving for intended purposes or not. When ÖSYM's declared intention to increase the weight of past achievement in the placement process is considered, it is obvious that

determining these weights will be even more crucial. Aforementioned validity studies on exams may prove useful in future efforts to determine the weights of the criteria. Especially these kinds of studies have to consider the appropriateness of central examinations in assessing the abilities and learning outcomes obtained in vocational and technical schools.

As a concluding remark, comparing the CGPA's of different schools is a difficult task, and it probably is not possible to devise a 'fair' procedure for central assessment. Further research in this area may prove useful. However anything short of centrally administered tests (such as a test at the end of each high school year) is likely to leave a large number of students feeling that they are unfairly treated. On the other hand, increasing the number of tests is probably not a practical solution.



REFERENCES

- Barnett, V., and Lewis, T. (1984). *Outliers in Statistical Data*, 2nd Edition. John Wiley & Sons. Great Britain
- Brant, R. (1990). "Comparing Classical and Resistant Outlier Rules". *Journal of American Statistical Association*, 85(412), 1083-1090
- Carey, V. J., Walters, E. E., Wager, C. G., and Rosner, B. A. (1997). "Resistant and Test-Based Outlier Rejection: Effects on Gaussian One- and Two-Sample Inference". *Technometrics*, 39(3), 320-330
- Carling, K., (2000). "Resistant outlier rules and the non-Gaussian case". *Computational Statistics and Data Analysis*, 33, 249-258
- Chapman, D. W., and Snyder, C. W. Jr. (2000). "Can high stakes national testing improve instruction: reexamining the conventional wisdom". *International Journal of Educational Development*, 20, 457-474
- DİE- State Institute of Statistics (1995). *Turkish Population, 1923-1994, Demographic Structure and Development, Projections until mid 21st Century*. DİE Basımevi, Ankara
- DİE- State Institute of Statistics (1997). *Employment and Wage Structure, November 1994*. DİE Basımevi, Ankara
- DİE- State Institute of Statistics (2001). *Statistical Indicators, 1923-1998*. DİE Basımevi, Ankara
- Dundar, H., and Lewis, D. R. (1999). "Equity, quality and efficiency effects reform in Turkish higher education". *Higher Education Policy*, 12, 343-366
- ETS-Educational Testing Service. (2001). *Frequently asked questions for SAT I*. www.collegeboard.com/sat/html/students/mofaqt.txt, April 2001
- Frigge, M., Hoaglin, D. C., and Iglewicz, B. (1989). "Some implementations of boxplot". *The American Statistician*, 43(1), 50-54
- Hoaglin, D. C., and Iglewicz, B. (1987). "Fine-tuning some resistant rules for outlier labeling". *Journal of the American Statistical Association*, 82(400), 1147-1149

Hoaglin, D. C., Iglewicz, B., and Tukey, J. W. (1986). "Performance of some resistant rules for outlier labeling". *Journal of the American Statistical Association*, 81(396), 991-999

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis*. John Wiley & Sons, USA

Law, A. M., and Kelton, W. D. (1991). *Simulation Modeling and Analysis*, 2nd Edition. McGraw-Hill, USA

Linn, R. L., and Gronlund, N. E. (1995). *Measurement and Assessment in Teaching*. Prentice-Hall, New Jersey

OECD- Organization for Economic Co-operation and Development / Center for Educational Research and Innovation. (1997). *Education at a Glance, OECD Indicators 1997*. OECD, Paris.

ÖSYM- Student Selection and Placement Center. (1981). *1981 Üniversitelerarası Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1982). *1982 Üniversitelerarası Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1983). *1983 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1984). *1984 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1985). *1985 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1986). *1986 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1987). *1987 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1988). *1988 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1989). *1989 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1990). *1990 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1991). *1991 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1992). *1992 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1993). *1993 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1994). *1994 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1995). *1995 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1996). *1996 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1997). *1997 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1998). *1998 Öğrenci Seçme ve Yerleştirme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (1999). *1999 ÖSS, Öğrenci Seçme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (2000a). *Selection and placement of students in higher education institutions in Turkey*. www.osym.gov.tr/eng/eng.html, August 2000

ÖSYM- Student Selection and Placement Center. (2000b). *2000 ÖSS, Öğrenci Seçme Sınavı Kılavuzu*. ÖSYM Yayınları, Ankara

ÖSYM- Student Selection and Placement Center. (2001). *Yüksek Öğrenime Girişte Okul Türü ve Öğrenim Durumuna Göre Başvuran-Yerleşen Aday Sayıları*. ÖSYM Yayınları, Ankara

Quesenberry, C. P., and David, H. A. (1961). "Some test for Outliers". *Biometrika*, 48, 379-390

Ramberg, J. S., Dudewicz, E. J., Tadikamalla, P. R., and Mykytka, E. F. (1979). "A Probability Distribution and Its Uses in Fitting Data". *Technometrics*, 21(2), 201-209

Rosner, B. (1983). "Percentage points for a Generalized ESD Many-Outlier procedure". *Technometrics*, 25(2), 165-172

Stephens, M. A. (1974). “EDF Statistics for Goodness of Fit and Some Comparisons”. *Journal of the American Statistical Association*, 69, 730-737

Tukey, J. W. (1977). *Exploratory Data Analysis*, Reading. Addison-Wesley, Massachusetts

ÜSYM-Inter-university Selection and Placement Center. (1979). *Yükseköğretime girişte fırsat eşitliği*. ÜSYM Yayınları, Ankara

ÜSYM- Student Selection and Placement Center. (1980). *1980 Üniversitelerarası Öğrenci Seçme Sınavı Kılavuzu*. ÜSYM Yayınları, Ankara

YÖK-Higher Education Council. (1999). “Yükseköğretime giriş sınavı- Geçmiş yıllarla karşılaştırma ve değerlendirme”. Yükseköğretim Kurulu, Ankara



APPENDIX A

DERIVATION OF AOBP FORMULATION

$AOBP_i^k$: Weighted secondary education academic achievement score of applicant i weighted according to composite ÖSS score category k

A_j^k : Arithmetic mean of the valid ÖSS Scores of students in composite score category k ($k = 1, 2, 3$) in school j

B_j : Highest OBP in the school j

C_j : Lowest OBP in the school j

ÖSYM's original formulation is:

Let

$$D_j^k = \frac{80 - \left[\left(\frac{A_j^k}{80} C_j \right) - \left(\frac{A_j^k - 80}{10} \right) \right]}{\left(\frac{A_j^k}{80} B_j \right) - \left(\frac{A_j^k}{80} C_j \right)}, \text{ then}$$

$$AOBP_i^k = \left[\left(\frac{A_j^k}{80} C_j \right) - \left(\frac{A_j^k - 80}{10} \right) \right] + \left[\left(OBP_i \frac{A_j^k}{80} \right) - \left(\frac{A_j^k}{80} C_j \right) \right] D_j^k$$

Note that the defined α term can be rewritten as follows:

$$\alpha_j^k = \left[\frac{A_j^k}{80} \times (C_j - 8) \right] + 8 = \left[\left(\frac{A_j^k}{80} C_j \right) - \left(\frac{A_j^k - 80}{10} \right) \right]$$

then

$$D_j^k = \frac{80 - \alpha_j^k}{\frac{A_j^k}{80} (B_j - C_j)}$$

The corresponding AOBP formulation of ÖSYM will be:

$$AOBP_i^k = \alpha_j^k + \frac{A_j^k}{80} (OBP_i - C_j) \frac{80 - \alpha_j^k}{\frac{A_j^k}{80} (B_j - C_j)}$$

which corresponds to the derived AOBP formulation (2.4) upon the elimination of $A_j^k/80$ terms

$$AOBP_i^k = \alpha_j^k + \left[(80 - \alpha_j^k) \frac{OBP_i - C_j}{B_j - C_j} \right]$$

APPENDIX B

FORMULA FOR TUKEY'S RULE

Let n denote the sample size to which the Tukey's Rule will be applied.

Carey et al. (1997) proposes the following formulas for k_T (where $\log(n)^k$ is used to denote the exponentiation of the quantity $\log(n)$):

For $p_n=0.05$:

- i. For $10 \leq n < 300$, if $n \bmod 4 = 0$ or 1 ,

$$k_T \approx 2.308486 - 0.08364704 \log(n) + 0.01044393 \log(n)^2 + 0.0009883642 \log(n)^3$$

if $n \bmod 4 = 2$ or 3 ,

$$k_T \approx 2.668116 - 0.2723008 \log(n) + 0.04326603 \log(n)^2 - 0.0009040389 \log(n)^3$$

- ii. For $300 \leq n < 2000$

$$k_T \approx 1.521267 + 0.1458931 \log(n)$$

For $p_n=0.01$ and $10 \leq n < 2000$:

$$k_T \approx 7.085162 - 2.692155 \log(n) + 0.5950019 \log(n)^2 - 0.05763641 \log(n)^3 \\ + 0.002164973 \log(n)^4$$

APPENDIX C

SUMMARY STATISTICS FOR SAMPLE SCHOOLS

Table C.1: Average, minimum and maximum values, and histogram summaries of arithmetic mean CGPA (\bar{X}_{CGPA}) for 12 school categories an the whole sample

Category	Number of Schools	Average X_{CGPA}	Minimum X_{CGPA}	Maximum X_{CGPA}	Histogram Summaries of Schools				
					$2.0 \leq X_{CGPA} < 3.0$	$3.0 \leq X_{CGPA} < 3.5$	$3.5 \leq X_{CGPA} < 4.0$	$4.0 \leq X_{CGPA} < 4.5$	$4.5 \leq X_{CGPA} < 5.0$
1	203	3.143	2.508	4.233	58	126	16	3	
2	73	4.044	3.404	4.663		2	26	42	3
3	34	3.970	3.093	4.381		1	17	16	
4	40	3.711	2.296	4.654	3	10	14	10	3
5	39	4.511	3.497	4.858		1		15	23
6	10	3.999	3.440	4.442		1	3	6	
7	56	3.508	2.794	4.517	2	26	24	3	1
8	72	3.283	2.420	4.181	12	45	14	1	
9	40	3.577	2.568	4.385	4	12	21	3	
10	94	3.138	2.415	4.161	39	43	10	2	
11	70	3.728	3.133	4.469		11	50	9	
12	67	3.551	2.291	4.204	7	12	46	2	
<i>Overall₁*</i>	798	3.512	2.291	4.858	125	290	241	112	30
<i>Overall₂*</i>	763	3.466	2.291	4.766	125	289	241	99	9

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

Table C.2: Average, minimum and maximum values, and histogram summaries of standard deviation CGPA (s_{CGPA}) for 12 school categories an the whole sample

Category	Number of Schools	Average s_{CGPA}	Minimum s_{CGPA}	Maximum s_{CGPA}	Histogram Summaries of Schools				
					$0.0 \leq s_{CGPA} < 0.2$	$0.2 \leq s_{CGPA} < 0.4$	$0.4 \leq s_{CGPA} < 0.6$	$0.6 \leq s_{CGPA} < 0.8$	$0.8 \leq s_{CGPA} < 1.2$
1	203	0.653	0.275	1.080		1	70	111	21
2	73	0.491	0.309	0.724		9	53	11	
3	34	0.566	0.393	0.750		1	19	14	
4	40	0.578	0.190	0.852	1	3	17	17	2
5	39	0.297	0.141	0.483	9	23	7		
6	10	0.527	0.377	0.749		1	6	3	
7	56	0.697	0.270	0.998		2	11	31	12
8	72	0.594	0.228	1.355		4	47	15	6
9	40	0.470	0.297	0.728		13	21	6	
10	94	0.522	0.212	0.769		11	62	21	
11	70	0.539	0.171	0.948	1	7	43	17	2
12	67	0.528	0.278	0.929		7	43	15	2
Overall ₁ *	798	0.564	0.141	1.355	11	82	399	261	45
Overall ₂ *	763	0.576	0.171	1.355	3	62	392	261	45

*: Overall₁: Whole sample population, additional 35 schools in category 5 are included, Overall₂: Whole sample population, additional 35 schools in category 5 are excluded.

Table C.3: Average, minimum and maximum values, and histogram summaries of sample skewness (ν^j) for 12 school categories an the whole sample

Category	Number of Schools	$\nu^j < 0$	$\nu^j > 0$	Average ν^j	Histogram Summaries of Schools				
					$\nu^j < -0.5$	$-0.5 \leq \nu^j < 0.0$	$0.0 \leq \nu^j < 0.5$	$0.5 \leq \nu^j < 1.0$	$1.0 \leq \nu^j$
1	203	16	187	0.526	3	13	73	95	19
2	73	50	23	-0.210	15	35	22		1
3	34	28	6	-0.239	6	22	5	1	
4	40	21	19	0.057	4	17	14	2	3
5	39	36	3	-0.812	24	12	3		
6	10	8	2	-0.234	3	5	2		
7	56	14	42	0.207	2	12	29	13	
8	72	12	60	0.397	2	10	32	25	3
9	40	13	27	0.215	3	10	13	12	2
10	94	14	80	0.408	3	11	39	33	8
11	70	31	39	-0.007	6	25	33	5	1
12	67	15	52	0.228	2	13	36	12	4
Overall ₁ *	798	258	540	0.192	73	185	301	198	41
Overall ₂ *	763	226	537	0.239	52	174	298	198	41

*: Overall₁: Whole sample population, additional 35 schools in category 5 are included, Overall₂: Whole sample population, additional 35 schools in category 5 are excluded.

Table C.4: Average, minimum and maximum values, and histogram summaries of ÖSS score average (A^{EW}) for 12 school categories an the whole sample

Category	Number of Schools	Average A^{EA}	Histogram Summaries of Schools				
			$A^{EA} < 100$	$100 \leq A^{EA} < 120$	$120 \leq A^{EA} < 140$	$140 \leq A^{EA} < 160$	$140 \leq A^{EA}$
1	203	105.495	26	175	1	1	
2	73	127.780		10	61	2	
3	34	135.740		2	18	14	
4	40	126.388	1	10	22	7	
5	39	154.407		1	3	25	10
6	10	125.726		3	7		
7	56	108.079	4	46	6		
8	72	101.496	40	30	2		
9	40	106.165	7	30	3		
10	94	96.993	79	14	1		
11	70	101.505	34	36			
12	67	105.561	10	56	1		
Overall ₁ *	798	111.022	201	413	125	49	10
Overall ₂ *	763	109.054	201	412	122	27	1

*: Overall₁: Whole sample population, additional 35 schools in category 5 are included, Overall₂: Whole sample population, additional 35 schools in category 5 are excluded.

APPENDIX D

DISTRIBUTION FITTING RESULTS

Table D.1: Distribution fitting results (for $\alpha = 0.05$ case) in four separate categories for the 12 school categories and the whole sample in numbers of schools and numbers of students

Category	Number of Schools	Histogram Summaries of Schools				Histogram Summaries of Students			
		<i>Only Normal</i>	<i>Only Lognormal</i>	<i>Normal and Lognormal</i>	<i>Neither</i>	<i>Only Normal</i>	<i>Only Lognormal</i>	<i>Normal and Lognormal</i>	<i>Neither</i>
1	203	4	49	97	53	360	7853	3777	12072
2	73	9	2	52	10	883	145	2416	1107
3	34	3	1	18	12	290	61	1137	1504
4	40	2	2	26	10	55	116	798	459
5	39	4	0	14	21	351	0	1007	1554
6	10	1	0	7	2	32	0	473	98
7	56	2	5	40	9	170	409	1409	1570
8	72	2	12	47	11	284	839	1793	1610
9	40	0	1	37	2	0	29	1026	51
10	94	5	19	56	14	384	2478	2895	2377
11	70	2	2	60	6	40	124	1995	161
12	67	4	6	54	3	293	130	2637	118
<i>Overall₁*</i>	798	38	99	508	153	3142	12184	21363	22681
<i>Overall₂*</i>	763	35	99	495	134	2881	12184	20450	21295

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

Table D.2: Distribution fitting results (for $\alpha = 0.01$ case) in four separate categories for the 12 school categories and the whole sample in numbers of schools and numbers of students

Category	Number of Schools	Histogram Summaries of Schools				Histogram Summaries of Students			
		<i>Only Normal</i>	<i>Only Lognormal</i>	<i>Normal and Lognormal</i>	<i>Neither</i>	<i>Only Normal</i>	<i>Only Lognormal</i>	<i>Normal and Lognormal</i>	<i>Neither</i>
1	203	3	48	118	34	135	9948	5714	8265
2	73	4	0	62	7	583	0	3235	733
3	34	6	1	20	7	647	61	1250	1034
4	40	0	1	35	4	0	102	1199	127
5	39	3	0	21	15	243	0	1603	1066
6	10	2	0	8	0	98	0	505	0
7	56	0	1	46	9	0	194	1794	1570
8	72	3	6	58	5	252	512	2395	1367
9	40	0	1	39	0	0	28	1078	0
10	94	1	14	72	7	165	2325	4101	1543
11	70	0	1	67	2	0	22	2216	82
12	67	2	1	62	2	220	43	2840	75
<i>Overall₁*</i>	798	24	74	608	92	2343	13235	27930	15862
<i>Overall₂*</i>	763	23	74	589	77	2268	13235	26511	14796

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

APPENDIX E

COMPARISON OF OUTLIER DETECTION RULES

Table E.1: Average of p values (\bar{p}_i) and error rates for ÖSYM's method and two alternative outlier detection rules for the transformed data

Category	ÖSYM			Median			GESD		
	\bar{p}_i	$MSD_{\bar{p}_i}$	$MSD_{0.02}$	\bar{p}_i	$MSD_{\bar{p}_i}$	$MSD_{0.02}$	\bar{p}_i	$MSD_{\bar{p}_i}$	$MSD_{0.02}$
1	0.0056	0.0001	0.0004	0.0070	0.0003	0.0006	0.0163	0.0017	0.0017
2	0.0223	0.0004	0.0004	0.0093	0.0005	0.0007	0.0185	0.0015	0.0015
3	0.0162	0.0002	0.0002	0.0029	0.0001	0.0005	0.0018	0.0001	0.0006
4	0.0061	0.0001	0.0005	0.0044	0.0003	0.0007	0.0231	0.0026	0.0026
5	0.0334	0.0003	0.0003	0.0200	0.0004	0.0004	0.0452	0.0028	0.0033
6	0.0163	0.0002	0.0003	0.0042	0.0001	0.0005	0.0082	0.0001	0.0004
7	0.0043	0.0001	0.0005	0.0036	0.0002	0.0006	0.0164	0.0024	0.0025
8	0.0052	0.0001	0.0005	0.0093	0.0005	0.0008	0.0273	0.0026	0.0026
9	0.0101	0.0003	0.0005	0.0111	0.0009	0.0010	0.0330	0.0030	0.0031
10	0.0090	0.0002	0.0005	0.0180	0.0015	0.0015	0.0315	0.0029	0.0029
11	0.0156	0.0005	0.0006	0.0211	0.0018	0.0018	0.0337	0.0039	0.0040
12	0.0080	0.0002	0.0005	0.0092	0.0004	0.0007	0.0234	0.0021	0.0021
<i>Overall₁*</i>	0.0107	0.0003	0.0005	0.0104	0.0007	0.0009	0.0233	0.0023	0.0023
<i>Overall₂*</i>	0.0097	0.0002	0.0005	0.0100	0.0007	0.0009	0.0222	0.0023	0.0023

*: *Overall₁*: Whole sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole sample population, additional 35 schools in category 5 are excluded.

APPENDIX F

LOWER-OUTLIERS ACCORDING TO ÖSYM

Table F.1: Statistics for the size of schools (n) (average, minimum, and maximum school size values) and histogram summaries for the 12 school categories and the whole sample, categorized according to the existence of lower-outliers by ÖSYM's method

	Category	Number of Schools	Number of Students	Average n	Minimum n	Maximum n	Histogram Summaries Schools						
							$5 \leq n < 50$	$50 \leq n < 100$	$100 \leq n < 150$	$150 \leq n < 200$	$200 \leq n < 250$	$250 \leq n < 300$	$300 \leq n$
Lower-outlier(s) is detected	1	40	9227	230,675	11	544	4	7	3	5	4	5	12
	2	51	3926	76,980	11	238	16	22	7	5	1		
	3	26	2574	99,000	34	267	7	9	5	3	1	1	
	4	10	460	46,000	19	102	8	1	1				
	5	37	2722	73,568	32	94	8	29					
	6	6	433	72,167	32	111	2	3	1				
	7	9	1363	151,444	26	744	5		2	1			1
	8	12	1070	89,167	32	225	7	1	1	1	2		
	9	13	541	41,615	11	124	11	1	1				
	10	31	5317	171,516	11	607	9	2	5	4	2	4	5
	11	29	1372	47,310	11	139	18	9	2				
	12	18	1569	87,167	22	248	6	6	4		2		
	Overall ₁ *	282	30574	108,418	11	744	101	90	32	19	12	10	18
	Overall ₂ *	249	28204	113,269	11	744	93	65	32	19	12	10	18
Lower-outlier(s) is not detected	1	163	14835	91,012	9	616	88	27	18	10	3	7	10
	2	22	625	28,409	9	77	20	2					
	3	8	418	52,250	22	112	4	3	1				
	4	30	968	32,267	7	97	25	5					
	5	2	190	95,000	44	146	1		1				
	6	4	170	42,500	35	62	3	1					
	7	47	2195	46,702	10	359	33	12			1		1
	8	60	3456	57,600	9	398	42	9	4	2		1	2
	9	27	565	20,926	7	55	26	1					
	10	63	2817	44,714	6	247	48	6	6	2	1		
	11	41	948	23,122	5	98	38	3					
	12	49	1609	32,837	7	114	43	5	1				
	Overall ₁ *	516	28796	55,806	5	616	371	74	31	14	5	8	13
	Overall ₂ *	514	28606	55,654	5	616	370	74	30	14	5	8	13

*: Overall₁: Whole sample population, additional 35 schools in category 5 are included, Overall₂: Whole sample population, additional 35 schools in category 5 are excluded.

APPENDIX G

COEFFICIENT OF CORRELATION ESTIMATES

Table G.1: Coefficient of correlation (r) estimates between students' CGPA and ÖSS scores in Quantitative category summarized for 12 categories and the whole sample ($CGPA_i$ vs. \ddot{OSS}_i^Q)

Category	Number of Schools	Average r	Minimum r	Maximum r	Histogram Summaries of Schools						
					$-1.0 \leq r < -0.2$	$-0.2 \leq r < 0.0$	$0.0 \leq r < 0.2$	$0.2 \leq r < 0.4$	$0.4 \leq r < 0.6$	$0.6 \leq r < 0.8$	$0.8 \leq r < 1.0$
1	203	0,518	-0,147	0,856		2	5	39	93	59	5
2	73	0,497	-0,105	0,858		1	4	19	20	27	2
3	34	0,604	0,227	0,805				1	13	19	1
4	40	0,527	-0,209	0,901	1	2	3	3	13	12	6
5	39	0,630	0,423	0,843					13	24	2
6	10	0,346	-0,138	0,697		1	2	3	2	2	
7	56	0,502	0,150	0,858			2	13	25	13	3
8	72	0,344	-0,059	0,774		3	16	27	16	10	
9	40	0,372	-0,480	0,908	2		8	10	12	6	2
10	94	0,253	-0,089	0,833		7	34	38	8	5	2
11	70	0,384	-0,186	0,705		5	10	18	24	13	
12	67	0,405	-0,095	0,814		1	8	23	29	4	2
Overall ₁ *	798	0,447	-0,480	0,908	3	22	92	194	268	194	25
Overall ₂ *	763	0,439	-0,480	0,908	3	22	92	194	256	173	23

*: *Overall₁*: Whole Sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole Sample population, additional 35 schools in category 5 are excluded.

Table G.2: Coefficient of correlation (r) estimates between students' CGPA and ÖSS scores in Verbal category summarized for 12 categories and the whole sample ($CGPA_i$ vs. $\ddot{O}SS_i^V$)

Category	Number of Schools	Average r	Minimum r	Maximum r	Histogram Summaries of Schools						
					$-1.0 \leq r < -0.2$	$-0.2 \leq r < 0.0$	$0.0 \leq r < 0.2$	$0.2 \leq r < 0.4$	$0.4 \leq r < 0.6$	$0.6 \leq r < 0.8$	$0.8 \leq r < 1.0$
1	203	0,477	-0,014	0,862		1	7	49	105	35	6
2	73	0,491	0,185	0,804			2	12	45	13	1
3	34	0,523	0,106	0,798			2	5	17	10	
4	40	0,502	-0,321	0,810	1	3	2	2	13	18	1
5	39	0,545	0,244	0,762				4	23	12	
6	10	0,545	0,391	0,712				2	5	3	
7	56	0,581	-0,213	0,873	1			7	17	28	3
8	72	0,419	0,038	0,820			8	26	26	10	2
9	40	0,166	-0,609	0,713	4	2	15	15	3	1	
10	94	0,240	-0,115	0,785		7	39	33	11	4	
11	70	0,457	-0,082	0,785		2	9	15	25	19	
12	67	0,439	-0,026	0,781		1	5	22	27	12	
Overall ₁ *	798	0,439	-0,609	0,873	6	16	89	192	317	165	13
Overall ₂ *	763	0,434	-0,609	0,873	6	16	89	189	295	155	13

*: *Overall₁*: Whole Sample population, additional 35 schools in category 5 are included, *Overall₂*: Whole Sample population, additional 35 schools in category 5 are excluded.

APPENDIX H

REAL AND NET DIFFERENCES FOR ALTERNATIVES

Table H.1: Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the first alternative method, summarized for 12 school categories and the whole sample

Category*	Number of Schools C>30 C=30			Average Δ	Minimum Δ	Maximum Δ	Number of Students	Histogram Summaries of Students			
								$\Delta < -10$	$-10 \leq \Delta < -5$	$-5 \leq \Delta < 0$	$\Delta = 0$
1	203	163	40	-1,395	-13,538	0,000	24062	106	1542	12987	9427
2	73	22	51	-0,274	-12,387	0,000	4551	8	50	544	3949
3	34	8	26	-0,155	-6,030	0,000	2992		4	406	2582
4	40	30	10	-1,961	-12,938	0,000	1428	9	168	756	495
5	39	2	37	-0,082	-5,868	0,000	2912		6	182	2724
6	10	4	6	-0,468	-5,135	0,000	603		1	164	438
7	56	47	9	-1,634	-12,230	0,000	3558	62	377	1698	1421
8	72	60	12	-1,646	-10,120	0,000	4526	2	350	3040	1134
9	40	27	13	-2,137	-14,306	0,000	1106	45	148	344	569
10	94	63	31	-0,794	-14,430	0,000	8134	34	345	2371	5384
11	70	41	29	-0,910	-13,287	0,000	2320	13	84	810	1413
12	67	49	18	-1,325	-13,949	0,000	3178	33	217	1310	1618
Overall ₁ *	798	516	282	-1,129	-14,430	0,000	59370	312	3292	24612	31154
Overall ₂ *	758	514	249	-1,175	-14,430	0,000	56810	312	3286	24430	28782

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

Table H.2: Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the second alternative method (outlier detection according to Tukey's Rule), summarized for 12 school categories and the whole sample

Category	Number of Schools	C>30				C=30				Average Δ	Minimum Δ	Maximum Δ	Histogram Summaries of Students							
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4				$\Delta < -15$	$-15 \leq \Delta < -10$	$-10 \leq \Delta < -5$	$-5 \leq \Delta < 0$	$0 \leq \Delta < 5$	$5 \leq \Delta < 10$	$10 \leq \Delta < 15$	$15 \leq \Delta < 20$
1	203	0	0	88	75	8	0	0	32	-1,356	-13,538	9,863	101	2048	14465	7187	261			
2	73	0	0	19	3	47	0	3	1	0,869	-12,387	12,717	4	42	706	3778	19	2		
3	34	0	0	4	4	23	0	2	1	0,373	-6,030	4,632		4	792	2196				
4	40	0	0	26	4	8	0	0	2	-1,544	-12,938	17,661	9	185	736	474	19	3	2	
5	39	0	0	1	1	31	0	5	1	1,114	-5,868	6,261		6	417	2459	30			
6	10	0	0	3	1	5	0	0	1	0,433	-5,135	5,708		1	175	424	3			
7	56	0	0	36	11	6	0	0	3	-1,664	-12,230	4,609	62	373	2082	1041				
8	72	0	0	41	19	8	0	0	4	-1,405	-23,697	7,639	6	5	475	2632	1338	70		
9	40	0	0	25	2	13	0	0	0	-1,244	-14,306	8,979	45	148	320	586	7			
10	94	0	0	44	19	11	0	1	19	-0,069	-15,935	17,042	1	34	484	3589	3542	470	11	3
11	70	0	0	38	3	27	0	0	2	-0,216	-13,287	8,025	13	99	841	1357	10			
12	67	0	0	41	8	13	0	0	5	-0,353	-13,949	19,759	27	218	1414	1276	242			1
Overall ₁ *	798	0	0	366	150	200	0	11	71	-0,709	-23,697	19,759	7	180	2903	22426	31981	1845	18	9
Overall ₂ *	763	0	0	365	149	173	0	6	70	-0,787	-23,697	19,759	7	180	2902	22216	29704	1773	18	9

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

Table H.3: Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the second alternative method (outlier detection according to Median Rule), summarized for 12 school categories and the whole sample

Category	Number of Schools	C>30				C=30				Average Δ	Minimum Δ	Maximum Δ	Histogram Summaries of Students							
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4				Δ<-15	-15≤Δ<-10	-10≤Δ<-5	-5≤Δ<-0	0≤Δ<5	5≤Δ<10	10≤Δ<15	15≤Δ<20
1	203	0	0	85	78	7	0	1	32	-1,291	-15,241	10,801	1	106	2060	14124	7428	342	1	
2	73	0	0	19	3	40	0	10	1	0,446	-22,491	7,956	4	10	54	961	3517	5		
3	34	0	0	4	4	22	0	3	1	0,287	-6,030	2,798			4	835	2153			
4	40	0	0	27	3	7	0	1	2	-1,717	-12,938	8,257	9	190	771	444	14			
5	39	0	0	1	1	23	0	11	3	0,344	-5,868	5,940		7	785	2114	6			
6	10	0	0	3	1	4	0	1	1	0,274	-5,135	5,708		1	205	395	2			
7	56	0	0	36	11	5	0	1	3	-1,702	-12,230	3,559	62	367	2107	1022				
8	72	0	0	36	24	8	0	0	4	-1,164	-23,697	16,160	6	5	459	2544	1374	130	7	
9	40	0	0	25	2	11	0	1	1	-1,661	-17,149	8,979	2	46	154	382	517	5		
10	94	0	0	42	21	9	0	2	20	-0,097	-15,935	16,305	1	40	459	3622	3609	387	13	
11	70	0	0	35	6	19	0	5	5	-0,604	-16,597	13,562	2	14	107	1076	1113	6	2	
12	67	0	0	38	11	11	0	0	7	-0,790	-13,949	19,759	27	211	1549	1374	16		1	
Overall ₁ *	798	0	0	351	165	166	0	36	80	-0,797	-23,697	19,759	16	319	4073	28961	25060	913	23	5
Overall ₂ *	763	0	0	350	164	146	0	26	77	-0,848	-23,697	19,759	16	319	4066	28239	23235	907	23	5

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

Table H.4: Average, minimum and maximum values, and histogram summaries of the real difference (Δ) of students according to the second alternative method (outlier detection according to GESD Rule), summarized for 12 school categories and the whole sample

Category	Number of Schools	C>30				C=30				Average Δ	Minimum Δ	Maximum Δ	Histogram Summaries of Students							
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4				Δ<-15	-15≤Δ<-10	-10≤Δ<-5	-5≤Δ<-0	0≤Δ<5	5≤Δ<10	10≤Δ<15	15≤Δ<20
1	203	0	0	83	80	8	0	0	32	-1,418	-13,538	18,340		75	2071	14848	6812	234	18	4
2	73	0	0	19	3	46	0	4	1	0,762	-22,491	12,717	4	5	43	772	3711	14	2	
3	34	0	0	4	4	23	0	2	1	0,373	-6,030	4,632			4	792	2196			
4	40	0	0	25	5	8	0	0	2	-1,347	-12,938	17,661		9	162	744	479	26	6	2
5	39	0	0	1	1	29	0	7	1	0,937	-5,868	5,940			6	565	2321	20		
6	10	0	0	3	1	5	0	0	1	0,433	-5,135	5,708			1	175	424	3		
7	56	0	0	36	11	6	0	0	3	-1,664	-12,230	4,609		62	373	2082	1041			
8	72	0	0	40	20	8	0	0	4	-1,320	-23,697	16,160	6	5	471	2604	1347	85	7	1
9	40	0	0	24	3	13	0	0	0	-1,148	-14,306	11,108		33	143	347	577	4	2	
10	94	0	0	42	21	11	0	1	19	-0,106	-14,430	17,042		28	433	3616	3688	351	14	4
11	70	0	0	37	4	24	0	1	4	-0,233	-16,597	21,400	2	13	84	924	1273	17	5	2
12	67	0	0	39	10	12	0	0	6	-0,360	-15,032	19,759	1	34	211	1434	1241	252	2	3
Overall ₁ *	798	0	0	353	163	193	0	15	74	-0,744	-23,697	21,400	13	264	4002	28903	25110	1006	56	16
Overall ₂ *	763	0	0	352	162	168	0	8	73	-0,814	-23,697	21,400	13	264	3996	28338	23141	986	56	16

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

Table H.5: Average, minimum and maximum values, and histogram summaries of the net difference (Δ) of students according to the second alternative method (outlier detection according to Median Rule), summarized for 12 school categories and the whole sample

Category	Number of Schools	C>30				C=30				Average Δ	Minimum Δ	Maximum Δ	Histogram Summaries of Students							
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4				$\Delta < -15$	$-15 \leq \Delta < -10$	$-10 \leq \Delta < -5$	$-5 \leq \Delta < 0$	$0 \leq \Delta < 5$	$5 \leq \Delta < 10$	$10 \leq \Delta < 15$	$15 \leq \Delta < 20$
1	203	0	0	85	78	7	0	1	32	-0,443	-14,393	11,649	49	1385	12101	9830	694	3		
2	73	0	0	19	3	40	0	10	1	1,294	-21,643	8,804	4	6	42	436	4049	14		
3	34	0	0	4	4	22	0	3	1	1,135	-5,182	3,646			1	208	2783			
4	40	0	0	27	3	7	0	1	2	-0,870	-12,090	9,105	6	133	583	683	23			
5	39	0	0	1	1	23	0	11	3	1,192	-5,020	6,788			1	306	2590	15		
6	10	0	0	3	1	4	0	1	1	1,122	-4,287	6,556				100	498	5		
7	56	0	0	36	11	5	0	1	3	-0,855	-11,382	4,407	43	299	1484	1732				
8	72	0	0	36	24	8	0	0	4	-0,316	-22,849	17,007	6	3	314	2149	1852	188	13	1
9	40	0	0	25	2	11	0	1	1	-0,813	-16,301	9,827	2	26	132	353	584	9		
10	94	0	0	42	21	9	0	2	20	0,751	-15,087	17,153	1	27	274	2855	4302	653	15	7
11	70	0	0	35	6	19	0	5	5	0,243	-15,749	14,410	2	12	83	724	1487	10	2	
12	67	0	0	38	11	11	0	0	7	0,057	-13,101	20,607	19	158	1105	1843	52			1
Overall ₁ *	798	0	0	351	165	166	0	36	80	0,051	-22,849	20,607	15	191	2822	22404	32233	1663	33	9
Overall ₂ *	763	0	0	350	164	146	0	26	77	0,000	-22,849	20,607	15	191	2821	22145	29948	1648	33	9

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

Table H.6: Average, minimum and maximum values, and histogram summaries of the net difference (Δ) of students according to the second alternative method (outlier detection according to GESD Rule), summarized for 12 school categories and the whole sample

Category	Number of Schools	C>30				C=30				Average Δ	Minimum Δ	Maximum Δ	Histogram Summaries of Students							
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4				Δ<-15	-15≤Δ<-10	-10≤Δ<-5	-5≤Δ<-0	0≤Δ<5	5≤Δ<10	10≤Δ<15	15≤Δ<20
1	203	0	0	83	80	8	0	0	32	-0,604	-12,724	19,154		28	1361	12743	9437	463	22	8
2	73	0	0	19	3	46	0	4	1	1,577	-21,677	13,532	4	4	33	363	4086	59	2	
3	34	0	0	4	4	23	0	2	1	1,187	-5,216	5,446			1	210	2773	8		
4	40	0	0	25	5	8	0	0	2	-0,533	-12,124	18,475		6	111	565	702	35	7	2
5	39	0	0	1	1	29	0	7	1	1,751	-5,053	6,754			1	204	2650	57		
6	10	0	0	3	1	5	0	0	1	1,248	-4,321	6,522				104	491	8		
7	56	0	0	36	11	6	0	0	3	-0,850	-11,416	5,423		44	305	1485	1716	8		
8	72	0	0	40	20	8	0	0	4	-0,506	-22,883	16,974	6	3	327	2239	1801	138	11	1
9	40	0	0	24	3	13	0	0	0	-0,334	-13,492	11,922		16	122	323	627	16	2	
10	94	0	0	42	21	11	0	1	19	0,708	-13,615	17,856		19	264	2845	4394	588	15	9
11	70	0	0	37	4	24	0	1	4	0,581	-15,782	22,215	2	12	65	637	1572	24	6	
12	67	0	0	39	10	12	0	0	6	0,454	-14,218	20,573		22	160	1084	1628	276	5	2
Overall ₁ *	798	0	0	353	163	193	0	15	74	0,070	-22,883	22,215	12	154	2750	22802	31877	1680	70	22
Overall ₂ *	763	0	0	352	162	168	0	8	73	0,000	-22,883	22,215	12	154	2749	22598	29576	1626	70	22

*: Overall₁: Whole Sample population, additional 35 schools in category 5 are included, Overall₂: Whole Sample population, additional 35 schools in category 5 are excluded.

APPENDIX I

EFFECTS OF AN INDIVIDUAL STUDENT FOR THE PROPOSED METHOD

Table I.1: The effects of an individual student for the first proposed alternative

Students	OBP		AOBP		AOBP (alternative 1)	
	School Before Transfer	School After Transfer	School Before Transfer	School After Transfer	School Before Transfer	School After Transfer
1	80.000	80.000	80.000	80.000	80.000	80.000
2	77.000	77.000	77.675	77.495	77.217	77.495
n-1	37.000	37.000	46.675	44.095	40.106	44.095
n	35.000	35.000	45.125	42.425	38.250	42.425
n+1		30.000		38.250		38.250

THE UNIVERSITY OF
MILWAUKEE