

High Stake Testing At The Entrance To Higher Education In Turkey

Ali Baykal

*Bahcesehir University,
Istanbul*

ali.baykal@bahcesehir.edu.tr

ABSTRACT

In Turkey the imbalance between supply and demand for tertiary education is the main reason for having the Student Selection Examination (SSPE) at the entrance. SSPE is being criticized by everyone but it is acknowledged by the whole society for some thirty years. The resources allocated to higher education are very low. Therefore it is obligatory to select the most capable students to make the investment most efficient. Student success at school of course is a very important factor, but there is some evidence that it is not as predictive as SSPE. The reliability of SSPE battery is quite high. A selection test is not supposed to cover the whole domain of all subject matter areas. The main purpose of SSPE is not to appraise present competencies but to predict the future performance. SSPE must be maintained as a fair balance for justice rather than a jack of all trades.

Keywords:

INTRODUCTION

The purpose of this study is to delineate some critical aspects of Student Selection Examination (SSPE) at the entrance to higher education (osym.gov.tr). Selecting and sorting is very intricate and delicate task for everybody but especially for educators. No matter what the reason is it is very hard to discriminate among people. To begin with we can admit that such an examination is a merciless, a cruel elimination mechanism. It perpetuates the inequality in education favoring the rich over the poor, the male over the female, the urban over the rural. It is a fact that those who possess the resources become more successful in this exam. It is a fact that there are gaps between schools, towns, regions, with regard to the performance in the exam. The aim of the exam is not to make them even, but to distinguish between the equals as well.

Reasons for Student Selection

The exam is a means of assessment and the aim of the assessment is not to change the reality, but to depict it as it is! The success or the failure at the exam is not the reason of inequality, but its consequence. The reasons behind the inequality are factors that have been created, and ongoing for years before the exam.

Why are we then still obliged to maintain this apparatus?

What are the compelling forces for making such tough discriminations between young people every year?

In general, there are two reasons for the system of selection:

1. There may be prerequisites or minimum entry requirements for the educational program. For example, no conservatory would admit someone with a hearing disability, no matter what the quota is. Some visual capabilities may be required at the entrance to the schools where they train aircraft captains. Shortly whenever or wherever there are minimum requirements for admittance then we have to have a selection program.

2. Another reason why we have to select is because there is imbalance between the supply and demand. No one can accommodate in a small room a larger audience than its capacity. Each year, Turkey can admit three-to-four hundred thousands school graduates to tertiary education. The demand is around one and a half million.

Facts and Figures from Turkey

Some indicators of the resources that Turkey allocates to education can be seen in Figure 1.

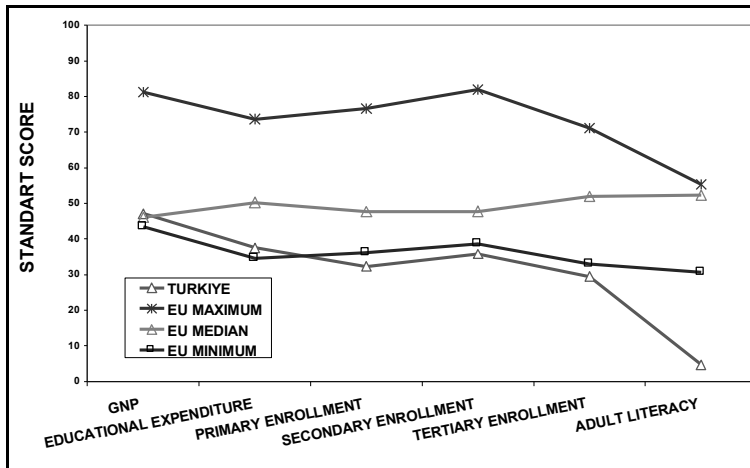


Figure 1: Turkey and EU: Educational Indicators (Baykal,)

The money allocated to education can be seen here. We can see basic education schooling, secondary education, schooling at tertiary education level, and finally adult literacy in Turkey. As can be seen in the graph Turkey is around the minimum level in each indicator within the European countries. Whereas Turkey is among the first 20 big economies of the world in terms of GDP (PWIF 2003). Obviously Turkey fails to supply sufficient resources for tertiary education to meet the demand. Figure 2 illustrates the student inflow at the entrance to higher education in Turkey until recently.

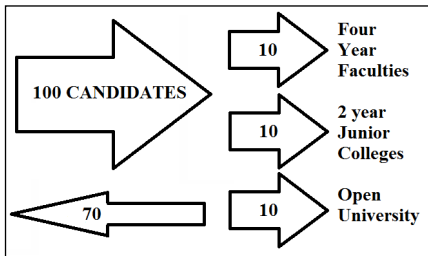


Figure 2: Student Inflow at the Entrance to Higher Education in Turkey

Does this demand cause a cram at the entrance to the higher education institutions? First of all cram is a misleading portrayal. There is not a big demand for higher education in Turkey. When we look at the potential population and the numbers of the applicants in Table 1 we can see that this is not the case. The percentage of potential in total population is 7.6% between ages 19 to 22. Percentage of applicants is only 2.4.

Table 1: Potential Population at the Entrance to Higher Education in Turkey

Years	Age 19-22	Applicants	Ratio
2010	5450	1513	0,28
2011	5390	1692	0,31
2012	5372	1895	0,35

Recently more than 5 million people are at an age to enter the university. And we have applications of 1,5 million. Almost half of them are trying their chances repeatedly every year. So, there is no such crowding in the demand. Only 1 to 1.5 million apply to enter. But since the facilities are limited, so is the possibility to get through. Out of 1000 students who finish the elementary school 70 only are able to complete the university in the next ten years. The enrollment at formal higher educational institutions is above 50 % in highly evolved countries and 20% in developing countries. In Turkey, the supply for education in general and higher education in particular, is so limited that there is a crowd of young population before the higher education. In any case, it is better for them than to form a crowd at soccer stadiums, in smoky cafes, unskilled labor recruitment markets, noisy discos, or extremist sects. Enigmatically, as if this were a negative attitude, there are talks of restricting the demand for higher education.

EXPECTED vs. OBSERVED PHENOMENA

Even if selection is mandatory aren't there alternatives to the exam? There are some as follows:

- a. If the future of the country and of the society can be gambled applicants might be selected by draws, or lotteries,
- b. If there was no risk of power struggle between the ruling class in the society and the academic administration applicants can be admitted by recommendation letters,
- c. if it doesn't hurt the feelings of justice quotas at the universities can be sold at auctions,
- d. if there was no fear of corruption, and bribery the students could have been admitted on the basis of grade point average attained at the secondary level.

All of the above have been tested and renounced in a variety of situations. We ended up to have an examination after each disappointment.

What are the desirable attributes of an selection and placement examinations? To what extent SSPE satisfies the criteria to be met?

In general there are three qualities for the accountability of a testing procedure: Reliability, validity and practicality. By reliability we do not just refer to the honesty of the people who take the exam, and of the people who organize the exam, but especially to the extent that measurement is free from random error. There are different types of reliability. Within the scope of this paper key reliability, scorer reliability, and the internal reliability of the test will be emphasized. Key reliability is the consistency between the answer keys prepared by the experts. This can be assured before the multiple choice exams. Scorer reliability is the consistency between the scores given by different scorers. This is also perfect in multiple choice exams. In essay exams however if we have 150 experts who evaluate the questions that we prepare, we'd have different answers from every single one of them. In scoring essays "halo effect" is an unavoidable source of error. Halo effect is something to do with the perception of people. It is not something that has to do with dishonesty. Each perceived stimulation usually affects the next one. For example, when we push our hand in hot water and then take it out and put it into warm water, we perceive it to be colder than it actually is. Halo effect applies to scoring essays. Every paper scored affects the score of the next paper. An average paper looks better after a poor one, or the same average response looks worse than it really is if it had been scored after a good one. To sum up scorer reliability is very low in scoring in written and oral exams. It is also very expensive, time taking and difficult to score essays and performance.

Objectivity

Objectivity in scoring examinations is very important in Turkey. Why? The Economist Magazine published a Pocket World in Figures 2003. One of the indicators about development and quality of life. is corruption perception. Unfortunately, Turkey ranks very bad in the perception of corruption, whether or not we have corruption, that's something to be questioned. But in Turkey, when we carry out, for example, a written examination, there has always been a sense, a perception that there is some kind of negative discrimination.

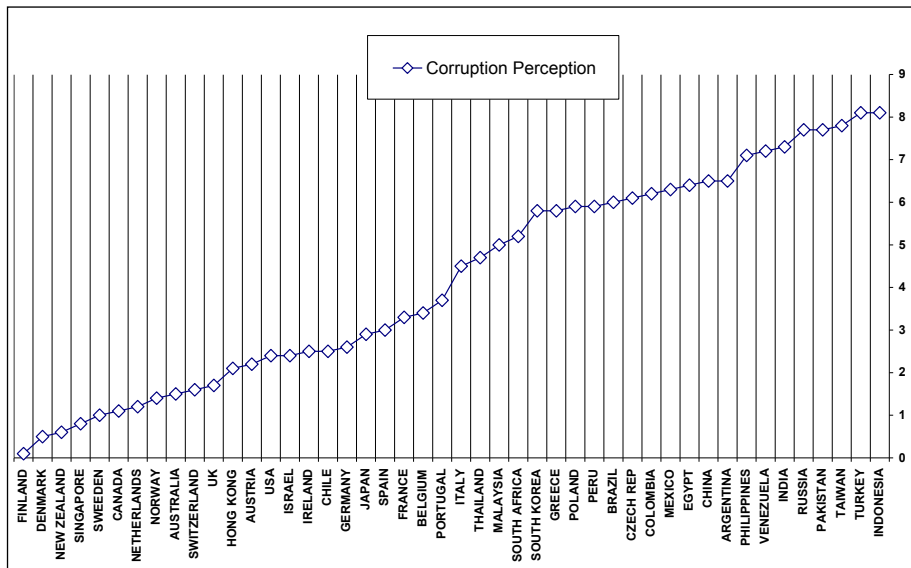


Figure 3: Corruption Perception Index in the Biggest Economies of the World (PWIF; 2003)

intersubject reliability

There are different methods and formulas to compute the internal consistency of scores obtained with a multiple choice test. Inter-subject reliability is something that ranks between 1 and 0. If it is zero it means that the scores are totally random, but the closer it gets to 1 it becomes more and more free from randomness, haphazardness. An internal consistency formula is known as KR-21 developed by Kuder and Richardson (Thorndike; 1988).

Table 2: KR-21 Reliability Coefficients of SSPE

Components Given in Different Years

Tests / Years	2010	2011	2012
Turkish	0,852	0,820	0,860
Hist/Geog/Psych	0,804	0,723	0,738
Math	0,815	0,715	0,733
Phys/Chem/Bio	0,657	0,622	0,567
English	0,940	0,963	0,969

Most of the reliabilities are quite high and all of them are satisfactory. Even the one which is 0.567 in natural sciences for the year 2012 is acceptable. This lower value can be explained by the attitude to sciences especially towards biology which is beyond our scope today.

Construct Validity

Another important test quality is construct validity. What should SSPE measure? What is SSPE really

measuring? The closer we get what is intended to be measured, the higher the construct validity will be. For example, I want to measure intelligence but I am measuring factual information; I want to measure creativity but I am actually measuring disobedience. For example, in Mathematics tests, there are implications of language, because you want to measure the ways of thinking, I mean the abstraction abilities of students, and you also have to measure the ability to understand verbal expressions. Figure 4 illustrates the construct validity.

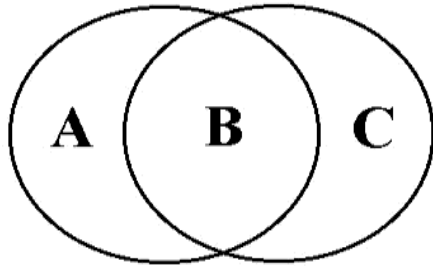


Figure 4: Relevancy Between What is Intended to be Measured and What is Really Measured

In Figure 4 area A+B is the set of competencies which are purported to be measured. Area C+B is the set of elements which have been measured in reality. In other words we have elements which couldn't have been measured although we wanted to (A), and there are some undesired leaks (C) which should not have been included. Area C is the extent to which we measure what we want to measure (construct validity). In real practice we can not claim that SSPE measures creativity. Creativity means originality, flexibility, and divergence. These competencies cannot be measured with multiple choice items which eventually converge upon predetermined keyed responses. We can neither say that SSPE measures the ability to synthesize and the ability to make deductions. Recently the learning outcomes, products of learning are described by the behaviors of the students as a part of Bologna process. This is something we borrowed from the US, and since 1964 we have been pursuing as well. We are looking at the level of education and trying to measure education based on the behavioral outputs of the students. And that is the basis on which Student Selection and Placement Center (SSPC) prepares the examinations. Therefore, as I said before, analysis-synthesis, comprehension, skills to find out a preset answers can be measured to some extent. Open ended questions such as "what can you do with a pencil" or "what can you do with a single brick" are divergent questions that require original answers can not be measured by this examination. We have to accept this. But luckily most of the competencies to be measured are correlated with the ones which cannot be measured. Table 4 displays a classification of examples of relevant and irrelevant measures in SSPE.

Table 4: Summative Analysis of Construct Validity in SSPE

	UNDESIRABLE	DESIRABLE
Dimensions have been measured	A: Income, gender, chance success	B: Knowledge, comprehension, Analysis
Dimensions could not have been measured	D: Weight, height, eye color	C: Creativity, synthesis, physical skills

Content validity

Content validity is the extent to which the items sampled in the test represents the total population of items in the content domain. Content domain has two dimensions: The construct or the sub-constructs such as comprehension, applications, analysis, creativity etc. constitute the first dimension. Second one is the subject matter area such as physics, chemistry and sub-topics in any of those... Desired competencies can be plotted within the cells of the crosstabulation of these two dimensions. We can also set the competencies to be measured on this grid. It is practically impossible to include all of the items covering all of the performance levels and the whole domain of

subject matter area. Due to so many constraints limited number of questions can be asked in SSPE. First of all affective traits, psycho-motor skills, and some mental abilities such as creativity, synthesis cannot be measured within the practical limits of SSPE. Memory skills are not relevant to a selection test. Some higher level mental constructs (e.g. comprehension, application, analysis, evaluation etc.) which are converging to a keyed response are being tested. Subject matter areas covered in SSPE are the classical course content such as Turkish, maths, physical and social sciences, and humanities. Shortly we cannot say that the content validity of SSPE is high. As a matter of fact it is not supposed to be high. It is not possible to maximize all of the desirable attributes simultaneously. What is essential is to maximize the quality which is relevant to the main purpose. The main purpose of SSPE is not to appraise present competencies but to predict the future performance.

Curricular validity

A special kind of content validity is the curricular validity. To what extent is the content of SSPE compatible with the high school curriculum? How coherent is it with the curriculum to be set up for the future? Theoretically SSPE is neither bound with high school curriculum nor it can be in accordance with the gigantic curricula in higher education. Instead some potential aptitudes and/or attitudes can be measured. But when such tool constructs are used as the criterion of selection they turn out to be the aims. Some years ago the general ability batteries had been used in the university entrance examinations. Right after this practice private tutorial organizations (dersane) appeared all over the country to train for "general ability". Inequality among applicants increased and interest in curricular topics decreased. Since so many students cannot afford individualized training for non-school themes it is rational to keep SSPE content within the boundaries of high-school curriculum. Therefore items in SSPE must be constructed in such a way to measure higher level cognitive competencies relevant to the most recent curricular content. That's what is being done by SSPE rather inelegantly though.

Face validity

Face validity is the degree of acceptability of an exam by the people who are involved in it. No matter how reliable, valid and practical in essence an exam must look appropriate to its audience. Although SSPE is being criticized by everyone it is acknowledged by the whole society for some thirty years. Nowadays some keep accusing the practice of measuring a three-year high school education in three hours? First of all, the length of time of the evaluation process does not necessarily have to be proportional to the dimensions of the entity to be measured. A baby takes nine months to grow before birth, but its weight and size can be determined in seconds, its color and sex can be told immediately. Inflation develops during the whole year but you can evaluate it in a very short period of time. The preparation of a dish may take hours, but the taste is manifest in the first few bites. The tasting of the aged wine does not take years!

There is no rule such as the longer the period of the exam the more accurate the measurements will be. Measurement is reduction of data. The simpler and the shorter it is the more precise it can be. The validity and reliability of the measurement depends on the system and tools, and not on the duration. Secondly, SSPE is not a final exam or a curriculum evaluation for the high school education. The aim is not to measure the past performance, but to estimate the future. This is the most important feature that SSPE should have, and it is called predictive validity.

Predictive validity

Since the resources allocated to education are low we are obliged to select the most capable students in order to make the investment in tertiary education most efficient. This can be done through an examination or through another process if it had been possible. The statistical correlation between predictor (SSPE) and the criterion (e.g. GPA at the university) is a simple measure of "predictive validity". Figure 5 illustrates the correlational thinking graphically. The points which constitute the ellipsis are individual applicants. The abscissa and the ordinate of any point are the measures of SSPE score and GPA of a particular applicant respectively. .

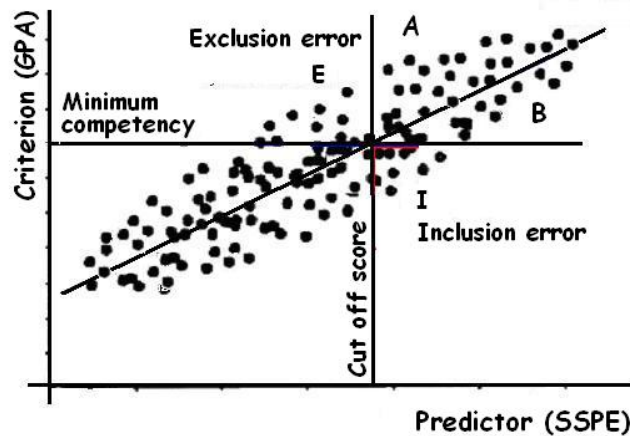


Figure 5: Hypothetical Illustration of Predictive Validity of SSPE

(Turgut, 1975; Guilford&Fruchter, 1978)

g

In practice cut off score is not predetermined by SSPE but it is the minimum score qualified for entrance to the last slot in the quota. The applicants below the “cut off score” are eliminated in SSPE. Therefore the GPA scores of those on the left side of the vertical line are not known. Area E are the students who would have been likely to be successful at the university but eliminated in SSPE. Therefore it represents the degree of exclusion error. Area I represents the inclusion error. Because these are the ones who were thought to have been successful in higher education but they disappointed SSPE. In area A there are the students who are more successful than expected; and in area B there are the ones who are less successful than expected. In case of perfect prediction the ellipsis collapses along the inclined line which means that the correlation is +1.00. In case of zero correlation prediction line disappears and elliptical distribution becomes a circular scattergram which indicates sheer randomness. Negative correlation implies significant prediction but in the opposite direction. There are many observations which has shown that SSPE predicts more significantly than expected. Due to the statistical nature of correlation coefficient such studies should be continuously carried out separately for each particular department.

School success of course is very important factor, but there is some evidence that it is not as predictive as SSPE. For instance, the best students of high-schools are not as successful as their schoolmates neither in SSPE nor at the university. We see that the correlation between SSPE and the college GPA is greater than the correlation between high-school GPA and the college GPA. So the high-school GPA is not a very good predictor even in the presence of SSPE. In the absence of SSPE as a sole criterion for entrance it is apt to corrupt abruptly.

Consequential validity

Consequential validity is the attribute expected to have been verified in a selection and placement program. Consequential validity can be inferred from the long term effects of the selection procedures. Most obvious evidence for the consequential validity is the contributions of graduates who had been selected years ago. There are many academicians among us. Many of them are younger than me. Administrators, experts, teachers had been selected by SSPE and had been educated in universities. Those people who are selected by this system have been educating the students, curing the patients, directing the institutions, managing the firms in Turkey for about thirty years. So, despite all of the downfalls, we can say that the Student Selection and Placement Centre is one of the most essential organizations for Turkish society.

CONCLUSION

The selection exam is an evaluation tool like any other tool, it can be improved. It can be made more efficient, more reliable, more valid (Ferrara, 2007). Most of the professionals in this field wish they were able to recommend a better system. But no one was able to come up with a better solution so far. Undoubtedly there are certain corrections that should be made in SSPE. First of all we must stop exploiting SSPE as if it were a Swiss knife. We are

trying to use SSPE to make every kind of correction in our educational system. We still hope to make up inequality in education by SSPE. We are trying to control secondary education by SSPE. Since these are the tasks impossible to be accomplished with examinations, SSPE is being judged very severely. SSPE must be maintained as a fair balance for justice rather than a jack of all trades.

REFERENCES

- Baykal, A. "Variables Inseparable: Indicators of Human Development and Information, Communication Technologies in Turkey". In Peköz, B., Bratton, N. (Eds.). *KSEE2005: Proceedings of the Knowledge Society Economy and Education*. North Cyprus: Girne American University. March 24-26, 2005. pp. 106-115.
- Ferrara, S. *Educational Measurement: Issues and Practice Toward a Psychology of Large-Scale Educational Achievement Testing: Some Features and Capabilities*, Springer London, 2007.
- Guilford, J.P., B. Fruchter. *Fundamental Statistics in Psychology and Education. (Sixth Edition)*. New York: McGraw-Hill, 1978, pp. 437-438.
- <http://osym.gov.tr/Genel/BelgeGoster.aspx?> (Accessed on December 2, 2013)
- PWIF 2003. Pocket World of Figures, 2003.
- Thorndike, R.L. "Reliability", Keeves, J.P. (Ed.). *Educational Research, Methodology, and Measurement*. New York: Pergamon, 1988. pp. 330-344.
- Turgut, M.F. "Theories of Error and Estimating the Errors of Measurement", *Hacettepe Bulletin of Social Sciences and Humanities*. 1975, 7:1-2, pp. 1-20.