Journal of Modern Education Review, ISSN 2155-7993, USA
February 2014, Volume 4, No. 2, pp. 119–141
Doi: 10.15341/jmer(2155-7993)/02.04.2014/006

http://www.academicstar.us

Matriculation and Psychometric Examinations in Mathematics and English, and the Relationship between Them

Shlomo Yitzhaki[1], Taina Pudalov[2], Aviel Krentzler[2]
(1. Department of Economics, Hebrew University, Jerusalem, Israel; 2. Central Bureau of Statistics, Israel)

Abstract: The contribution of this study lies in the research method, which enables to see whether two exams are examining the same abilities. The methodology can be described as reincarnation of variants of “Abstract Measurement Theory”, and “Item Characteristics Theory”. We examine the correlation between the scores of examinees on matriculation and psychometric exams. If the correlation is monotonic throughout the set of the scores, we can conclude that that the exams test the same abilities. However, if the sign of the local correlations change throughout the set of scores, we can conclude that the tests examine different fields of knowledge, or that the characteristics needed for success in one exam are different from those needed for success in the other exam.

Using this criterion, it can be concluded that the matriculation and psychometric examinations test indistinguishable abilities, but there is a large random error. The contributions of both exams to the total score for each student support the hypothesis that the existence of two examinations does not have a substantial impact. This conclusion applies all the more to the English exams, where the correlation between the matriculation and psychometric scores is higher than in the mathematics exams.

Key words: matriculation examination, psychometric examination, monotonicity

1. Introduction

For years now, Israel has used a combined system of matriculation and psychometric exams. These exams have different goals and are therefore in different formats. In more recent years, both exams have been used as criteria for acceptance to universities and institutions of higher learning. The double examination system using different testing formats results in a double burden on the examinees. Therefore, it may be valuable to evaluate the extent to which the fusion of the two exams into a single exam would decrease the burden on the examinees. That is, the goal of this work is to perform the preliminary stage of examining the cost versus benefit of the exam in order to evaluate whether the double examination is an excessive burden on the economy.

In order to evaluate the cost versus benefit, we must examine both the benefit of the existence of the double examination system and the financial burden on the economy. Measuring the cost to the economy is relatively

---

Shlomo Yitzhaki, Ph.D. in Economics, Professor Emeritus, Department of Economics, Hebrew University; research areas: public economics, econometrics. E-mail: shlomo.yitzhaki@huji.ac.il.
Taina Pudalov-Trahtenberg, MBA, Central Bureau of Statistics; research areas: economic research methods, education. E-mail: taina.pudalov@mail.huji.ac.il.
Aviel Kranzler, MBA, Central Bureau of Statistics; research areas: education. E-mail: avielk@cbs.gov.il.

simple, assuming that there exists data for the people employed in the field as well as estimates of time invested in preparation for these exams. It is more difficult, however, to measure the information buried within the existence of the double evaluation system,[1] since there is no simple way to differentiate between the existence of information in the abstract concept of knowledge.

In this article we concentrated on measuring the additional information on the examinees' abilities in the various subjects in the double examination system, which, as we mentioned, is the most difficult component to evaluate.

Our methodology is based on the following claim: we are examining the correlation between students' scores on their matriculation exams and on the psychometric exam, and if the result is that the correlation is monotonic for the entire range of scores, we will conclude that one of the tests is superfluous. If, however, the score ranking is not monotonic then it means that the tests check different areas of knowledge, or that the abilities that are necessary to succeed in one exam are different from those necessary to succeed in the other.

Evaluation of the correlation between the scores in the two tests causes problems that do not exist in the evaluation of the correlation between quantitative variables like, for example, the correlation between a person's weight and his or her height. This is due to the fact that there is no way to measure knowledge quantitatively, only indirectly such as the number of answers the examinee managed to answer correctly or alternatively by reading and evaluating the quality of the answers that the examinee provided. The number of questions that the examinee managed to answer depends both on the difficulty distribution of the questions on the exams and on the processing of the scores by the companies that administer the exams. In order to overcome this difficulty we will limit ourselves to checking the monotonicity of the correlation between the scores. The term "monotonicity of correlation" means that for the entire range of the matriculation exam scores (or psychometric test scores), an increase in matriculation exam scores (or psychometric test scores) brings about an increase in the average psychometric test scores (or matriculation exam scores) in the parallel exam. For this purpose we are using a relatively new statistical tool that aims to examine the monotonic correlation in areas where we have no method for measuring a variable directly.[2]

The structure of the paper is as follows: The first section contains a brief description of the research question. The second section deals with the significance that must be attributed to the skill evaluation, since we will not be analyzing the scores in the way we might analyze a quantitative variable like weight or height. The third section deals with presentation of the research method to examine the monotonicity of the correlation between English and Mathematics test subjects. The fourth section presents the data used and the analysis techniques while the fifth section presents the empirical findings.

The results of the empirical findings are that there is monotonic correlation between success on the matriculation examinations and success on the psychometric test and that this correlation is stronger in English than in Mathematics. At the same time, we found that there is white noise in the results of the different exams, which decreases the reliability of the exams in predicting abilities in that subject. In such a case one might argue that the reason for having two examination systems is to reduce the random noise and that this improvement requires the existence of two examinations in each subject. Section 6 seeks to answer this question that came up

[1] A study that was performed in 2007 by the National Center for Testing and Evaluation (http://www.nite.org.il/files/reports/342.pdf) found that the double system improves the validity of the prediction of success in freshman year of university. In section 6 we come back to this point.

[2] All methods used in this article are discussed in Yitzhaki and Schechtman (2013).

during the study and it uses the Gini index to measure the improvement resulting from the use of both scores instead of just one. This means that the evaluation was performed on the contribution of the use of two exams to improve the resilience of the score ranking. The conclusion from this section is that the additional information gained from using both examinations is less significant in the English evaluations than in Mathematics.

2. Short Description of the Research Question

The matriculation exams in the various core subjects are administered by the Ministry of Education in a national format and aim to rank students based on their level of specialization and their abilities in different subjects. The psychometric exam is based on the ranking by the National Institute for Testing and Evaluation “the psychometric test is a tool for predicting the chance of success in school and at institutions of higher learning”.[3] The score is used by those institutions to sort candidates into the various majors. The test allows for the ranking of all of the candidates on a common evaluation scale in comparison with alternative assessment scales it is less influenced by the examinees’ backgrounds or other subjective variables. The test is composed of three subjects: verbal reasoning, quantitative reasoning and English. In addition, the overall score is reported. In each of the subjects the assessment scale ranges from 50–150 points. The cumulative score on the test ranges from 200–800 points. The general score is calculated based on the scores of the examinee in the three subjects that make up the exam and is based on the weighted averages in which the scores in the different subjects receive different values.

The need to take two exams is a burden on the students who are required to prepare for two exams with different arrangements, as well as on the Israeli economy since there is a significant industry that would disappear in the event that only one exam is required. Since the matriculation exams came before the psychometric test, the question that remains to be answered is if the psychometric test evaluates some field of ability that is inherently different from that of the matriculation exams.

The argument that is used to justify the existence of the psychometric test is that the predictive capabilities of the test in terms of the success in academic studies is good and that the combination of the matriculation examinations along with the psychometric test score has even greater predictive success than each score separately. The claims regarding the predictive capabilities are not examined in this article, and that is because the students are accepted to study different majors.

The purpose of this article is to examine the more modest claim of the extent to which the psychometric tests measure a different range of skills than the matriculation exam on the same subject. That is, we will not be looking at the question of the cost-benefit analysis in terms of the economy or with the question of the predictive ability of the test but rather we will focus on the question of whether both tests, the matriculation exams and the psychometric test, evaluate indistinguishable skills.

The methodological innovation in this study is in the research method: using this method (which is applied in the paper) one can determine if two exams evaluate the same abilities. In other words, we can determine that the knowledge and abilities that allow for success in one exam are the same as the knowledge and abilities that are required for success in the second exam. In the next section we clarify the mathematical significance that should be applied to the results of the test and in the section after that we describe the tools that we used.

[3] https://www.nite.org.il/index.php/he/tests/psychometric/psychometric-about.htm.

3. Significance that Should be Attributed to A Test of Abilities[4]

For arguments sake, we will make a number of simplifying assumptions that we will later remove. In the first stage let's assume that we are able to define the difference between different abilities, for example that English skills are different and require different abilities than mathematics. We will call this variable that represents a person's ability in a particular subject that we wish to examine, "a".

The second assumption is that ability is a continuous variable such that if Nili is better than Guy then the probability that she will answer a question correctly within the framework of the test is greater than the probability that Guy will answer correctly. The significance of this assumption is that if there are two persons with a_1 and a_2 abilities and we know that $a_1 < a_2$, meaning that the person 2 ability is greater than person 1 ability then if a question with a difficulty level of q is asked, then the probability that the subject with ability a_2 will answer correctly is greater than the probability of the subject with ability a_1 to answer the question correctly.

The third assumption is that there is no randomness, meaning that each question has a critical difficulty level, such that every examinee with an ability level that is greater than the critical required ability for providing the correct answer will answer the question correctly, and anyone whose ability to answer the question is lower than the critical ability level, the probability that they will answer this particular question correctly is zero. This assumption will be omitted later.

Using these three assumptions, we can define what an exam as:

Definition: An exam is a set of questions on a particular subject, with varying degrees of difficulty that are intended to find out the abilities of the examinees.

Now we will move on to differentiate between the measurement of the height of a group of people and the measurement of their knowledge under the above assumptions. In the height assessment, there is significance to the differences in centimeters between people. In the knowledge measurement the differences in the test results are a function of the difficulty distribution of the questions on the exam. Essentially we could claim that in the test for knowledge assessment on a particular subject, the parallel in the height example would be if the examinees stood behind a screen and the examiner asked questions such as "who is at least 150 cm tall? Who is at least 160 cm tall?" and all of the people in the group who are taller than 150 cm or 160 cm answer according to their respective heights, whether or not they are the exact height that was asked. At this point it becomes clear that there is no difference between the scores received by the people who are 151 cm tall and those who are 159 cm tall and there is a difference between the people who are 159 cm tall and those who are 161 cm tall. This means that the examiner, by determining the difficulty distribution of the questions in the questionnaire, applies a monotonic non-decreasing transformation to ability. Changing the difficulty distribution of the questions is equivalent to applying a different monotonic non-decreasing transformation.

The argument so far intends to explain that scores in an exam are an ordinal variable. This fact is already known and used by "Abstract Measurement Theory" as forwarded by several authors including Cliff (1977, 1992), which was not adopted by the profession (Cliff, 1992).

We now move to discuss of the assumptions assumed so far.

The first assumption we will eliminate is the third, which claims that there is absolute certainty in the translation of ability into score. Let's assume that the probability that one answers a question at a particular level

4 For further information on this topic, see Yitzhaki and Eisenstaedt (2003) and Schechtman and Yitzhaki (2009).

of difficulty correctly is a non-decreasing function of ability that can be randomly disrupted. In such a case, we will claim that on average, the answers of the examinees with high-level abilities will be higher scoring than the answers of examinees will lower ability levels. That is, we will assume that the chances of answering a question correctly are a random variable with a conditional expected value that increases along with an increase in the examinees' abilities.

An additional possibility is that the results of the exam do not examine a single range of capabilities but rather examine a mix of abilities in a number of fields. For example: the success on the test depends on innate ability and diligence and these two traits are distributed differently in the population. In such a case, this does not require that the score be a non-decreasing monotonic transformation of a particular ability that is checked in it but rather it will be dependent on the distribution of the traits in the population and the correlation between them.[5] In this case, we may find non-monotonic correlation between the scores of the exams.[6] Our claim is that if the score on the psychometric exam is an increasing monotonic function of the matriculation score and the matriculation score increases as a function of the psychometric test score, we will conclude that due to the limitations that exist in the measurement of knowledge, the tests evaluate indistinguishable abilities. We refer to it as the same ability. Since there could be a disturbance in the monotonicity of the connection between the two scores, because of the random disruption, we will ignore the small disruptions in the correlation between the scores and we will address only the large disruptions, where the probability that they are the result of random disturbances is low.

4. Monotonic Correlation

The method employed enables us to check whether a monotonic non-decreasing transformation of a random variable enables the change in the sign of a correlation or regression coefficient. This is equivalent to asking whether there exists an alternative exam, with a different difficulty distribution of the questions, that can change the sign of the correlation between the scores of the two exams. The method to identify the monotonicity of the correlation is based on the LMA curve which is an acronym for the longer name: Line of independence Minus the Absolute concentration curve.

The LMA curve allows for identification of whether or not the signs of the local correlation coefficients between two variables do not change along the distribution of one of the variables. If the conclusion that is reached is that the sign of the correlation coefficient does not change then the conclusion is that there is no non-decreasing monotonic transformation that can change the sign of the correlation coefficient and in this case we conclude that there is no way to differentiate between two subjects of knowledge and as a result, as far as we are concerned, the exams check the same type of abilities.[7] If, however, the sign of the local correlation coefficients change along the distribution, then the conclusion is that there is a non-decreasing monotonic transformation, which can change the sign of the correlation coefficient. This implies that one can write an alternative exam that can change the sign of the correlation coefficient. The alternative legitimate exam that can do the sign change is based on a change in the level of difficulty of the questions. The significance that should be

[5] See Yitzhaki, Itzhaki and Pudalov (2012).

[6] In general, the average slope of the regression curve of the variable Y on the variable X does not have to have the same sign as the average slope of the regression curve of X on the variable Y. The sign of the average of the slopes is equal on in the event that the definition of the examinee between the variables is symmetrical in the variables: only in the event that statement "what you see from here is what you see from there" is true.

[7] In the event that these are different areas of skill, all we can say is that the tests cannot differentiate between them. That is because that we can examine their ability only through the exam results.

attributed to such a finding is that the exams evaluate different abilities and that by constructing the proper exams one can differentiate between them.

The next section explains the tools available for analysis and will perform a demonstration by comparing the results of psychometric tests and matriculation examinations. Since the use of this tool is relatively simple relative to the required proofs, we refer the interested reader to the relevant sources.[8]

5. Method for Data Analysis

The population is composed of those who were examined in the mathematics and English matriculation exams in 1999–2000 and participated in the psychometric test. An additional limitation is that the examinees met the acceptance requirements for university, meaning, among other things, were examined at a level of at least 4 units in English and at least 3 units in Mathematics.

There are a number of difficulties in comparison of scores from the two exams: one is that the examinee in the matriculation exam can choose the level he or she wishes to be examined by selecting the number of units on the test, while the psychometric test is uniform. That means that matriculation exams are actually made up of a number of smaller exams correlating with the number of units on the test while the psychometric test is just one exam. The second difficulty in comparison is that the scale of scoring of the different exams is different and the question is how one can bring them to a common denominator for the purpose of comparison.

We tackle the first problem by checking every level of the matriculation exam separately and comparing each one to the psychometric test, and subsequently integrate all of the matriculation exams according to the weight they are given by the universities for the different levels of the exam and by doing this we will be able to compare all of the students. In terms of the second complication which is the fact that there are a number of ways to perform normalization where the most common method is the assumption that the distribution is normal and therefore it is acceptable to normalize using the standard deviation. However as we will see later on, the distribution of the test scores is not normal and therefore the normalization methods that we have chosen to use is based on the following domain:[9]

We left the matriculation scores as they were. We normalized the psychometric test scores such that the domain of the scores in which it is identical to the matriculation scores: Let X be the original psychometric score and A, B are the minimum and maximum matriculation scores, respectively and C and D are the minimum and maximum original psychometric scores respectively. Let Y be the normalized psychometric score.

The transformation formula is as follows:

$$Y = \alpha + \beta(X - C) \quad \text{where} \quad \beta = \frac{B-A}{D-C} \quad \text{and} \quad \alpha = A$$

The result of the normalization is that the domain of the matriculation exam scores and the domain of the psychometric scores are identical.

An additional difference is that while the psychometric exam is identical for all of the examinees, the matriculation exams in mathematics are divided into three levels in accordance with the number of units upon which the student is being tested. That means that there is self-selection of the student in the matriculation

[8] See Yitzhaki and Schechtman (2013).
[9] It is important to note that the normalization method does not affect the results. Its main effect is on defining the conversion rate between one point on one exam and one point on the other, which is relevant only for the last section of the paper.

examinations in accordance with the number of units chosen. At this stage we are limiting ourselves to students being tested on 5 units. Afterwards we will add the other groups of students to whole sample.

For the purpose of the monotonic test, we use two drawings of Absolute Concentration Curves that allow for analysis of the monotonicity of the correlation.

The drawing of the first curve is under the assumption that the results of the two exams are statistically independent.

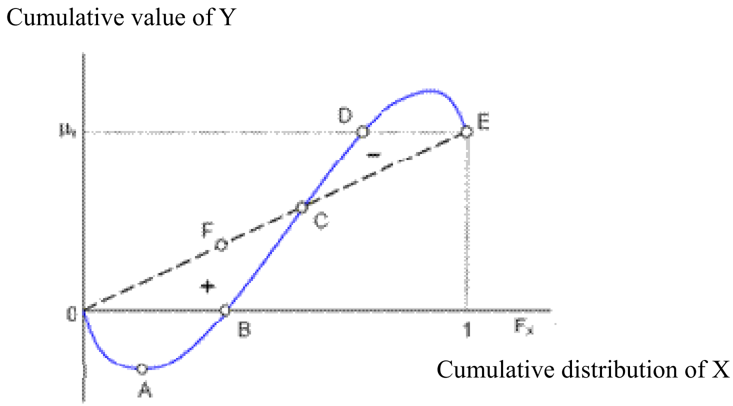

Figure 1 Line of Independence and the Absolute Concentration Curve

On the horizontal axis is the cumulative distribution of the scores that came from X. On the vertical axis is the cumulative value of the scores of the Y exam, assuming that the scores are statistically independent. The graph in this case is a straight line that starts at the origin and ended at point (1, mean score of Y).

The drawing of the second curve is the actual cumulative value of the scores in the Y exam. This curve can take on different shapes and all we can determine for sure is the starting and ending points. The curve starts at the origin and ends at point (1, mean score of Y), meaning at the same point as the first curve.[10] The vertical difference between the two curves is the LMA curve.

The LMA curve has the following properties:

(1) Does the graph cross the horizontal axis? In the event that it does, then the relationship between the X and Y variables is not monotonic, where X explains Y. Meaning, there is a range of X in which the covariance is negative, the range below the horizontal axis, and there is a range of X in which the covariance is positive, where the graph is above the horizontal axis. In such a case, non-decreasing monotonic transformation applied to X can change the sign of Pearson's correlation coefficient between the two variables. That is because the monotonic transformation of X can decrease the size of the undesired correlation area and increase the size of the desired correlation area. The significance of this for our work is that the exams do not examine the same type of abilities.

(2) Is there a region in which the graph changes its curve from concave to convex or vice versa. In this case, as well, the significance is that the correlation is not monotonic. If the segment is concave then the correlation in that segment is positive and the convexity indicates negative correlation in that segment. Assuming that the graph does not cross the horizontal axis, is that although one cannot change the sign of the correlation between the two variables (i.e., by using a different exam on X), one can change the sign of the correlation by disallowing some of the student to participate in the X exam). That is, if we remove some of the students whose

[10] The cumulative value is the value sum of the variable up to a certain point. X is arranged in increasing order although this does not

scores are a part of undesired sections then there is another legitimate exam that will swap the score ranking between the two exams. In this case, as well, the conclusion is that the exams do not check the same ability.

In the implementation of the evaluation of the matriculation and psychometric exams, our claim will be that as long as the correlation between the results of one test increases along with the scores of the other exam, as far as we are concerned, the two exams measure the same ability. However, if the correlation is not monotonic, we will determine that the exams check different abilities. It should be clear that it is possible that the exams examine different abilities but one cannot distinguish between them.

The area contained between the LMA graph and the x-axis is equal to cov(Y, F(X)) which is the equivalent of the covariance if we measure variability by the Gini index. If we normalize the vertical axis by dividing by cov(X, F(X)) the area contained between the curve and the horizontal axis is the regression coefficient based on the Gini index of the variable Y on the variable X. The normalized graph is called NLMA where the N represents normalization.[11]

Note, that as opposed to the covariance that is based on the variance that is symmetric between the variables, (resulting from the fact that cov(X,Y) = cov(Y,X), the covariance based on the Gini index is not symmetric and as a result one should verify that what is seen from one side is the same as what is seen from the other, meaning that increasing monotonic transformation does not affect the X variable the way the same transformation affects Y.

This concludes our review of the features relevant to the field of education evaluation.

6. Empirical Assessment

The data set is composed of matriculation-eligible students who were tested in mathematics and English in 1999-2000 and participated in the psychometric examination afterwards.[12] That is, we limit ourselves to the same population to whom the university selection process is open.

In total there were 23,069 examinees (passers) in the mathematics and English matriculation exams in 1999–2000, of whom 20,235 were eligible for a Matriculation certificate. (Students eligible for matriculation are those that met the requirements of the Ministry of Education to receive a matriculation certificate. This includes, among other things, 20 credits, and completion of the tests in the subjects defined as required). Of those students, those who met the University requirements were 17,867 (among others, those who passed the mathematics matriculation with at least 3 units and also passed the English matriculation with at least 4 units). The size of the sample (essentially the examinees in a particular year) is relatively large such that almost every deviation in estimation of the parameter would be significant.

The first test was performed on the students who took the 5 unit English matriculation exam.

6.1 The Examinees in 5 Units of English

Table 1 summarizes the central parameters that characterize the score distribution and the correlation between them. As one can see, the (normalized) psychometric examination score, the mean and the median are higher than in the matriculation exam which is a result that can be explained by the fact that the psychometric test is just a single exam while the matriculation exam is administered differently according to the number of units completed. The standard deviation and the Gini coefficient of the matriculation exam are higher than the standard

require that the value of Y meet the requirements of X.

[11] The features of the Gini index can be found in Yitzhaki and Schechtman (2013), Yitzhaki (2003).

[12] Each student's highest psychometric test score was selected.

deviation and the Gini coefficient in the English psychometric examination, again a result of the self selection in the matriculation exam. The final four columns describe the correlation coefficients: Pearson's index which is an index based on linear correlation, the Spearman index and Gini's two correlation coefficients,[13] and one can see that the high correlation between the results of the exams reaches 0.7.

Table 1 English Tests Parameters of 5 Units' Examinees*

N = 12,100	Average	Median	Min	Max	St. Dev.	Gini	Pearson	Spearman	Gini Correlation Γyx	Gini Correlation Γxy
Matriculation	83.51	85.00	55	100	9.17	0.062	0.682 (0.000)	0.687 (0.000)	0.693	0.693
Psychometric exam	87.28	89.00	55	100	8.55	0.055				

* In order to compare standard deviation with the Gini coefficient the latter should be multiplied on average. For example, in English 5 units exam absolute Gini index is 0.062*83.51 = 5.18, and 0.055*87.22 = 4.80 in English part of psychometric exam.

From Table 1 one can learn that the average and the median of the normalized psychometric test scores are higher than the matriculation scores indicating that in the psychometric test there are more high scores than in the 5-unit matriculation exam. The standard deviation and the Gini coefficient indicate that the variance in the matriculation in the scores is higher than in the psychometric while the various correlation coefficients show the same range of size of correlation, which is relatively low, since we are looking at the scores received by a student in two exams that are meant to test capabilities in the same subject.

Figure 2 presents the distribution of matriculation scores in comparison with normalized psychometric scores. It is clear from the illustration that the psychometric test is better for discerning in the lower range of abilities while the matriculation tests are concentrated in the higher ability regions, a result that can be explained by the clustering that exists in the matriculations and in the fact that the psychometric examination is a single test that is identical for everyone. In addition, it is easy to see that the scores are not Normally distributed and that is because their distribution is asymmetric.

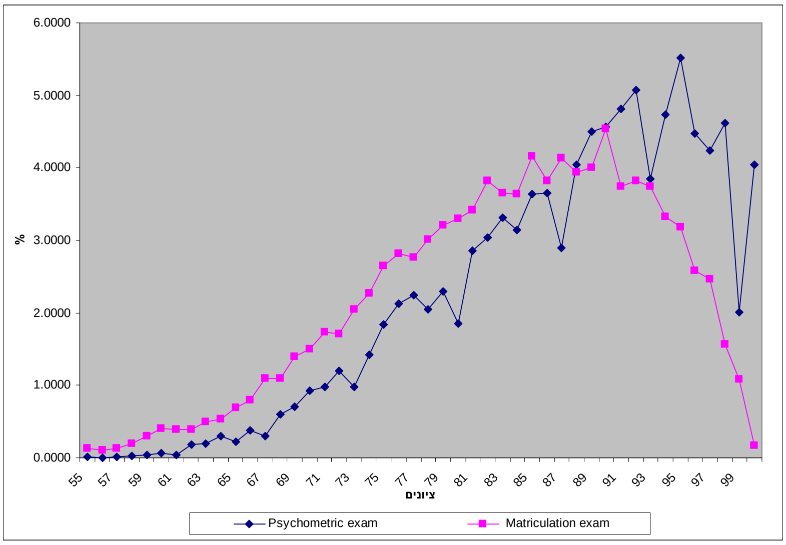

Figure 2 Distribution of 5 Units English Examinees*
* The observations were connected with straight lines.

[13] The Gini index has two equal correlation coefficients where the common distribution is symmetrical in both variables. The sample error of the correlation coefficient of Gini requires the construction of a special program. Since all of the correlation coefficients are similar to one another and since the sample is large, we did not calculate the sample error.

The next question examined is the level to which the two tests check the same abilities. For this purpose we use the monotonic correlation test.

Figure 3 examines the extent to which the psychometric test exhibits monotonic correlation with the matriculation exam scores. The features of the illustrated graph are as follows: if the graph increases (decreases), it's a sign that the scores in that subject are lower (higher) than the average of the scores on that test. If the graph is concave (convex), that is an indication that the scores in that subject are increasing (decreasing). The area between the graph and the horizontal axis is equal to the Gini correlation coefficient. If, however, the graph crosses the horizontal axis it is an indication that there is an additional legitimate exam[14] that can flip the sign of the Gini and Pearson's correlation coefficients between the scores of the two exams.

From illustration 3 we learn that the graph increases initially and subsequently decreases. In terms of the curve of the graph we are looking at a concave graph, allowing us to conclude that the psychometric scores produce a monotonic correlation with the matriculation exam results.

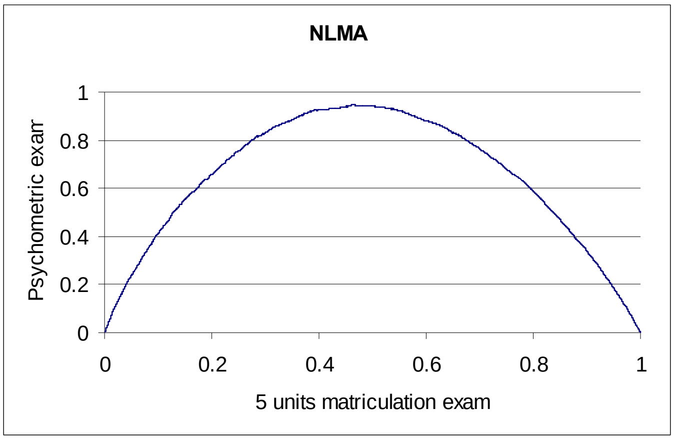

Figure 3 Psychometric Score as A Function of 5 Units English Exam Score

In order to check if the score of the matriculation exam produces symmetrical monotonic correlation with psychometric exam scores, we go back to Figure 3 and flip the axes. On the horizontal axis we present the cumulative distribution of the psychometric score and on the vertical axis we see the matriculation score. Figure 4 shows the NLMA graph of the matriculation scores as a function of the cumulative distribution of the psychometric exam. The curve that emerges brings us to the same conclusion as before the matriculation scores correlate monotonically with the psychometric scores. That means that the correlation is symmetrical and monotonic and therefore, at least in terms of this group, the tests are substitutes (though imperfect) for one another. We note that they are imperfect because of the relatively low level of correlation between the scores of the two exams.

In order to check the symmetry in the correlation between the scores of the exams in a non-parametric manner, we draw a transition matrix. Table 2 presents the transition matrix between the deciles of the scores. On the horizontal axis we have the deciles of the 5-unit English matriculation exam and on the vertical axis we have the psychometric scores for those same examinees. The diagonal shows the percentage of students whose scores are in the same decile in both tests. Roughly 50 percent of the lower decile remain in the lower decile in both the matriculation exam and in the psychometric exam and close to fifty percent of the highest decile remain in the

[14] When we use the phrase "another legitimate test" the intention is a different test with different distribution of difficulty of the questions.

highest decile. Only five percent of the lowest decile in the psychometric examination are above the median in the matriculation exams and a similar percentage of the lowest decile in the matriculation exams is above the median score in the psychometric test.

The significance that can be attributed to the transition matrix is that the exams check the same skill though there is noise in the results.

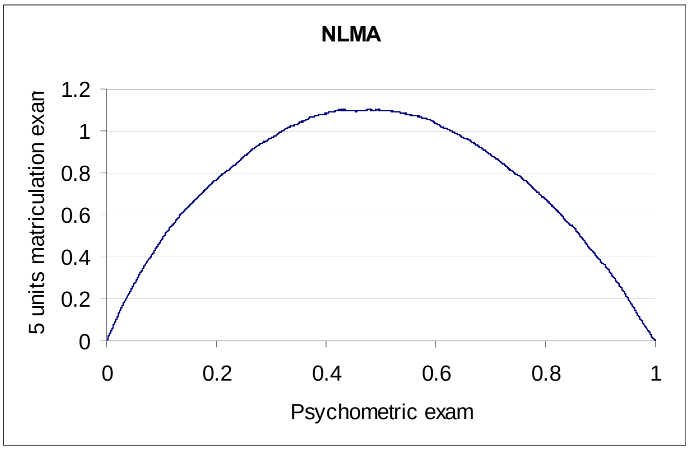

Figure 4 5 Units English Exam Score as A Function of Psychometric Score

Table 2 Transition Matrix: Deciles of Scores — 5 Units English Exam and English Part of Psychometric Exam Scores

	Matriculation Exam										
		1	2	3	4	5	6	7	8	9	10	Total
Psychometric Exam	1	4.8	2.0	1.6	0.7	0.5	0.2	0.1	0.03	0.01		10
	2	2.0	2.2	2.4	1.5	0.8	0.6	0.3	0.1	0.03		10
	3	1.3	1.8	1.9	1.7	1.0	1.1	0.8	0.2	0.2	0.02	10
	4	0.9	1.6	1.4	1.8	1.4	1.3	0.5	0.8	0.3	0.05	10
	5	0.6	1.1	0.9	1.1	1.5	1.7	1.1	1.2	0.7	0.1	10
	6	0.2	0.6	0.8	1.4	1.7	1.6	1.1	1.7	0.8	0.2	10
	7	0.1	0.3	0.5	0.7	1.1	1.2	2.0	2.0	1.5	0.6	10
	8	0.1	0.3	0.3	0.5	1.0	1.0	1.7	1.8	2.1	1.2	10
	9	0.1	0.2	0.3	0.3	0.7	0.8	1.5	1.2	2.0	3.1	10
	10		0.1	0.1	0.2	0.3	0.4	0.9	1.0	2.4	4.7	10
	Total	10	10	10	10	10	10	10	10	10	10	100

The existence of "noise" in the results should be troubling because it means that there are examinees who go from the highest decile to the second to lowest in their scores and the transfers appear to be symmetrical such that we cannot attribute them to a problem with the score system of one of the exams. The conclusion that results is that both exams evaluate the same subject but that the number of errors in the ranking of each exam in comparison with the other is high and therefore the mean of the two exams can be used as a better indication than simply relying on the results of one exam only.

6.2 Examinees in Mathematics, 5 Unit Matriculation Exam

In this section we compare between the results of the mathematics matriculation exam and the quantitative reasoning of the psychometric test among students who took the Mathematics matriculations for five units of study. The number of students for whom we have scores for both of the exams is 7,059. It is important to note that the number of students who took 5 units of mathematics is about 40 percent lower than the number of students who took the English 5-unit matriculation examination (roughly 7000 versus 12000). The rest of the evaluation is structurally similar to the assessment for the English matriculation exam and therefore we have chosen to focus only on the findings.

Table 3 Math Exams Parameters of 5 Units' Examinees

N = 7059	Average	Median	Min	Max	St. Dev.	Gini	Pearson	Spearman	Gini Correlation Γyx	Gini Correlation Γxy
Matriculation exam	85.9	88.0	53	100	10.79	0.070	0.384 (0.000)	0.391 (0.000)	0.409	0.392
Psychometric exam	87.9	89.0	53	100	7.94	0.049				

As one can see in Table 3, the mean score and the median (both normalized) of the psychometric examination are higher than the mean and median scores on the matriculation examination (5 units). However, the two variability measures that we used, the standard deviation and the Gini coefficient are higher in the matriculation examination than in the psychometric test. The significance is that the matriculation scores have greater variability than the psychometric scores meaning that relative to the psychometric test they are more concentrated at the edges of the score distributions. This type of situation can result from normalization of the scores for the students that are tested on five units in matriculation, normalization that is done independently of the other groups in mathematics *(diff. number of units).* That means that since the psychometric examination is a single exam, it is normal that someone who decides to take the 5-unit mathematics exam will be in the higher range of the psychometrics scores.[15] The surprising finding here is the relatively low correlation between the two exams that were given to the same people, that have the most extensive education in mathematics. All of the correlation coefficients between the results of the two exams do not surpass 0.4 while in English they stand at 0.7. Another surprising finding is that much like in the results of the English exam, here there is also symmetry in the distribution of the scores in the matriculation and the psychometric tests. Yet another surprising finding is the similarity between the two Gini correlation coefficients, such that "what you see from here (from the matriculation exam) is similar to what you see there (in the psychometric test)", meaning that a chance in the basis for comparison does not change the finding in comparison with coefficient. The symmetry between the distributions and the lower correlation in the mathematics matriculation exams indicate that the amount of random noise in mathematics is greater than the amount of random noise in the English tests.

Figure 5 presents the matriculation exam results and the normalized psychometric test scores for students who were tested in 5-units on their mathematics matriculation exam. It is clear from the figure that the psychometric test, relative to the 5-unit matriculation exam, concentrates in discerning between the highest and lowest ranges of the abilities while the matriculation exam focuses on the intermediate range of abilities. We can see this in the fact that the density of the examinees is greater in the psychometric in the center while in the 5-unit matriculation the density is greater in the two extremes of the distribution.[16] Since there is only one psychometric test while the matriculation exam is presented in three clusters, it makes sense that the matriculation exams will have a lower discernment capability in the highest and lowest regions. However as we mentioned in our analysis

[15]The hidden claim behind this sentence can be tested by testing the hypothesis that the examinees at the different levels of matriculation exams are stratified groups in the psychometric examination. This hypothesis is not tested in this study.

[16] It is important to note that since the score distribution depends on the difficulty distribution of the questions, there is no way to

of the English test, this can result from the normalization that was performed on the matriculation tests separately for the scores of those who took the 5-unit exam.

The next question examined is whether the two exams check the same type of abilities. For this we use the test of monotonicity of the correlation. Figure 6 presents the NLMA curve of Psychometric exam as a function of the cumulative distribution of the matriculation's scores. The graph is less smooth than it was in the analysis of the English exams, which could stem from the fact that the number of people taking the mathematics exam is smaller than the number of people who took the English exam, where the large number of exam takers could cause the smoothing of the distribution and elimination of the random fluctuations.

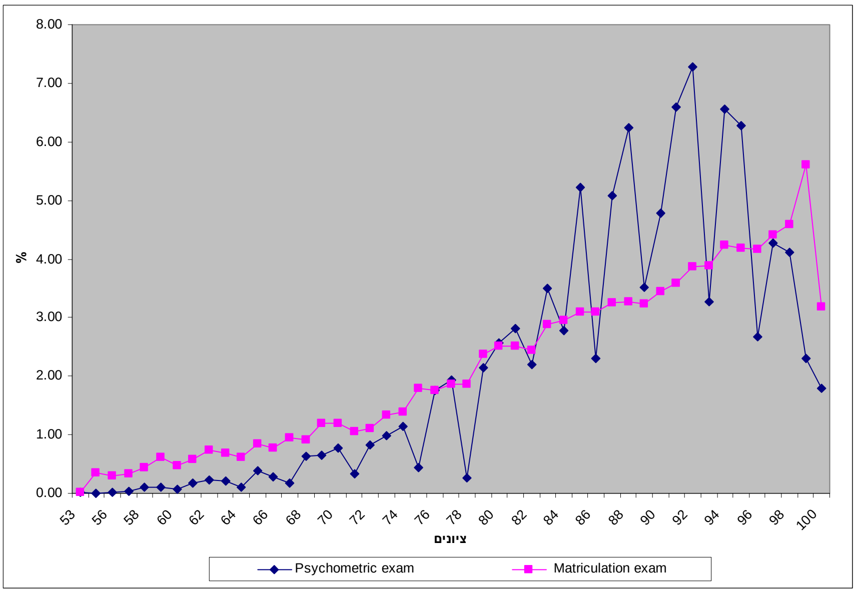

Figure 5 Distribution of 5 Units Math Examinees*

*In the mathematics matriculation of 5 units there are 10 questions total: 6 questions worth 16 2/3 points (3 units) and 4 questions worth 25 points (supplement to reach 5 units). Afterwards the results are weighted. In the psychometric test in mathematics there are 50 questions (two parts with 25 questions each). The absence of a lot of extremely high marks on the psychometric test scores seem odd to us but those were the results that we received.

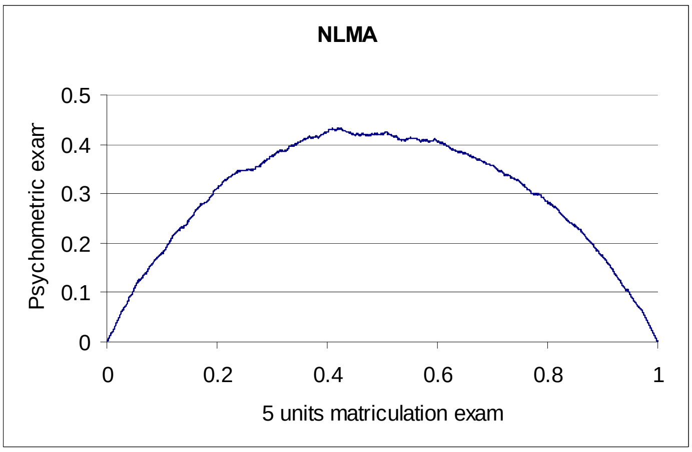

Figure 6 Psychometric Score as A Function of 5 Units Math Exam Score

---

evaluate the true distribution of the population.

Figure 7 presents the NLMA curve of the matriculation exam scores as a function of the cumulative distribution of the psychometric test. Much like in the English exams, in the mathematics domain the curve that emerges leads to the same conclusion as the one that we saw when we looked at it in the opposite way. The matriculation scores exhibit monotonic correlation with the psychometric scores, aside from random fluctuations in small ranges.

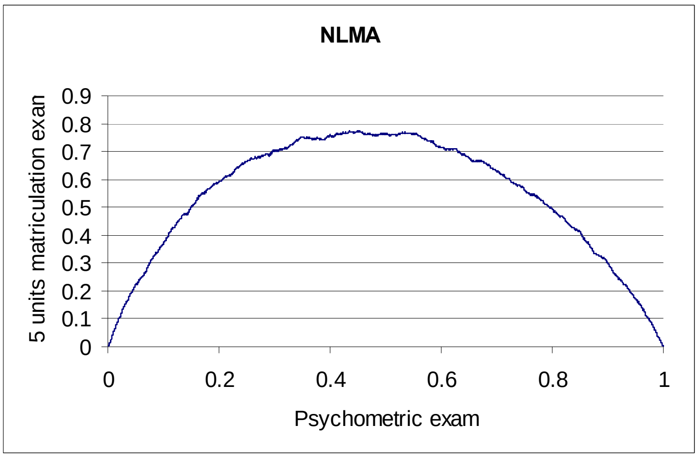

Figure 7 5 Units Math Exam Score as A Function of Psychometric Score

In this respect, the findings that despite the fact that the correlations are monotonic, the correlation coefficients are low, is somewhat unsettling. It is important to understand the significance of Table 4 transition matrix between the score deciles. On the horizontal axis, there are score deciles in the 5-unit mathematics matriculation exam and on the vertical axis are the psychometric exam data. The diagonal represents the percentage of students that remained in the same decile. If we check the lowest decile in both exams we will find that only 2.7 percent of the students are in the lowest decile in the matriculation exams and in the psychometric test. Meaning that in terms of the lowest decile, there is agreement between the two exams only within the lowest 2.7 percent of students (which are 27 percent of that group). The surprising thing is that roughly 18 percent of the group of the lowest decile of the psychometric test are above the median in the matriculation exam (0.1+0.3+0.3+0.4+0.7) and roughly 25 percent of the lowest decile of the matriculation scores are above the median in the psychometric exam. A similar finding occurs if one focus only on the highest decile in which just 29 percent of the students there remain in the highest decile in both exams. We focused on the highest deciles because there the probability that the students will remain in the same decile in both exams is the highest. That is because the density of the distribution is higher in the center than in the margins of the graph, which causes the chance of staying in the same decile in the center of the graph lower than in the margins. The transition matrix appears to be almost symmetrical indicating that the distribution includes white noise. The obvious conclusion is that the predictive capacity of success on one of the exams on the success in the other exam is low and is summarized in the correlation coefficient of about 40 percent. The conclusion we reached is that the mathematics exam has random noise that weigh on the predictive ability of success in one of the exams versus success on the other.

The transition matrix in mathematics shows the increase in "noise" in comparison with the exams in English. The conclusion is that there is a higher correlation between the results of the exams in English than between the results of the exams in Mathematics. Therefore, the damage to the economy that will result from cancelling the burden of two English examinations is smaller than the damage that will result from candidate selection in mathematics.

Table 4 Transition Matrix: Deciles of Scores — 5 Units Math Exam and Math Part of Psychometric Exam Scores

		Matriculation Exam										
		1	2	3	4	5	6	7	8	9	10	Total
Psychometric Exam	1	2.7	1.9	1.5	1.2	0.9	0.7	0.4	0.3	0.3	0.1	10
	2	1.8	1.9	1.4	1.2	0.7	0.9	0.8	0.6	0.5	0.2	10
	3	1.4	1.2	0.9	1.2	1.2	1.2	0.9	0.9	0.8	0.4	10
	4	0.7	1.3	1.1	1.3	1.2	1.0	1.2	1.1	0.7	0.4	10
	5	0.8	1.1	1.0	1.3	1.1	1.0	0.9	0.9	1.0	0.9	10
	6	0.7	1.0	1.4	1.1	1.2	1.1	1.0	1.0	0.8	0.8	10
	7	0.7	0.7	0.9	0.8	1.1	0.8	1.3	1.3	1.3	1.2	10
	8	0.6	0.5	0.8	0.8	1.3	1.3	1.3	1.2	1.3	1.2	10
	9	0.3	0.4	0.6	0.8	0.8	1.0	1.4	1.3	1.5	2.0	10
	10	0.2	0.3	0.3	0.4	0.7	0.9	0.9	1.4	2.0	2.9	10
	Total	10	10	10	10	10	10	10	10	10	10	100

6.3 Examinations in English, 4 and 5 Units

In this section we will repeat the analysis that was performed in the previous section but this time we will include the students who were examined on both 4 and 5 units in their English matriculation examinations, meaning all of the students who are eligible for acceptance into university. The weighting of the matriculation exam results was performed according to the rules at the universities when students are accepted.[17] Our observations are derived from the data from 17,867 examinees that were tested from 1999–2000 who met the university entrance requirements, among whom 5,767 completed 4 units and 12,100 completed 5 units. As in the previous section, we are normalizing the scores such that the domain of scores in the two exams will be identical.

Comparison of the differences between the scores using the parallel table of 5-units shows that the mean and median in each distribution are almost identical, though the standard deviation of each test and the Gini inequality coefficient have increased, a natural result of the fact that we expanded the population of exam takers. Figure 8 displays the distribution of frequencies. One can see that the distribution of the matriculation exam scores is more continuous than the distribution of the psychometric scores. If one "smooths out" the psychometric scores by averaging the adjacent observations within the results one gets almost identical distributions. Moreover, comparison of the correlation coefficients shows that all of the correlation coefficients increased significantly such that the quality of correlation between the two scoring systems increases significantly from when we compared the psychometric scores to the students who were tested on 5 units alone.

Table 5 English Tests Parameters

N = 17867	Average	Median	Min	Max	St. Dev.	Gini	Pearson	Spearman	Gini Correlation Γyx	Gini Correlation Γxy
Matriculation exam	101.5	103	68	125	13.72	0.077	0.82 (0.000)	0.82 (0.000)	0.83	0.83
Psychometric exam	102.3	104	68	125	14.41	0.081				

[17] 12.5 points are added for English and Mathematics that were learned at a level of 4 units, and 25 points are added if 5 units were studied.

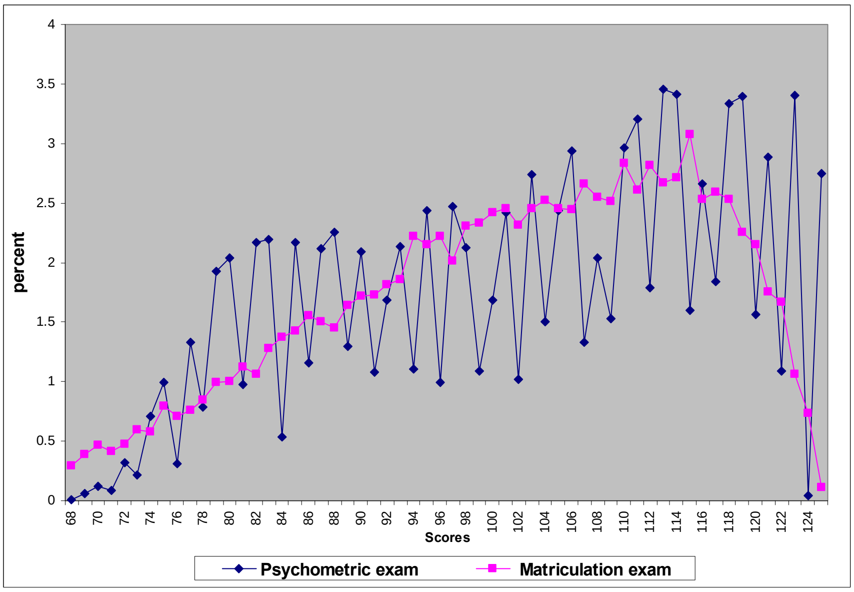

Figure 8 Distribution of English Examinees

Figure 9 presents an NLMA curve of the psychometric score as a function of the matriculation scores. As one can see from the figure the graph is concave along its entire length and therefore, even in this case we find that the psychometric scores serve as an increasing monotonic function of the matriculation exam scores, indicating that we are dealing with the same domain.

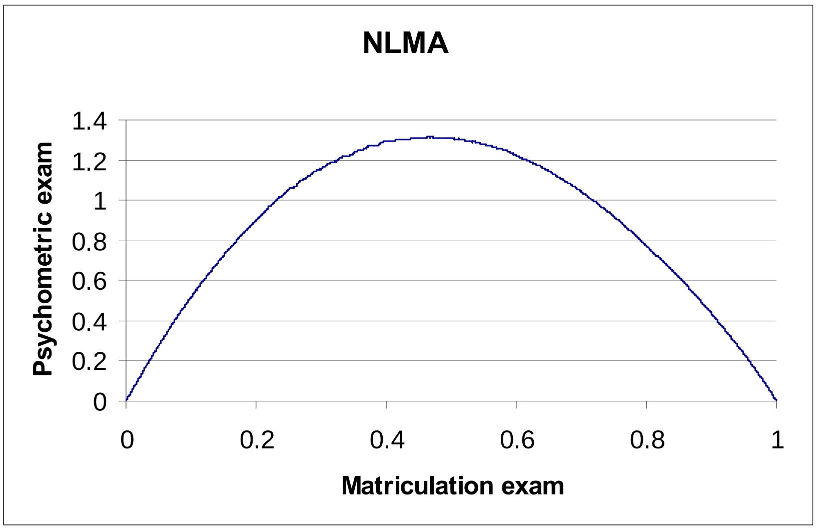

Figure 9 Psychometric Score as A Function of English Matriculation Exam Score

Just like in the previous section, here too we flipped the dependent variable and the independent variable and present the NLMA curve of the matriculation score as a function of the psychometric test scores. Here as well, the graph is concave and smooth indicating a monotonic correlation between the scores. The Gini correlation coefficients between the two variables, which are of similar size, indicate the same.

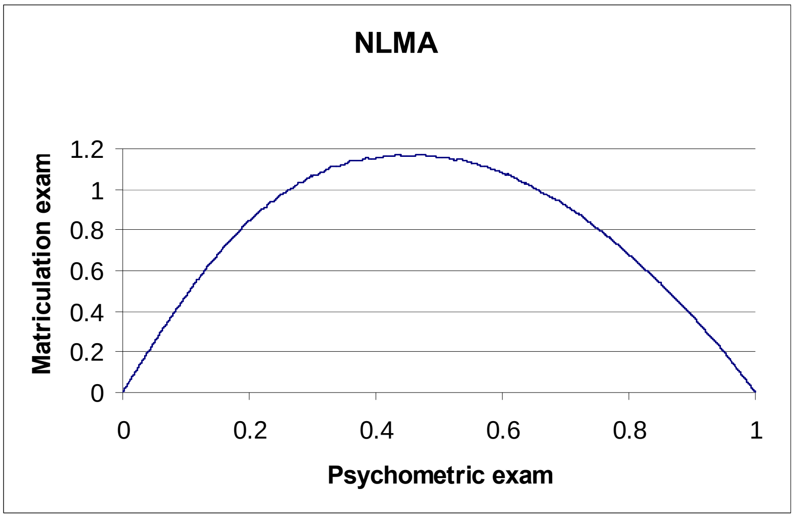

Figure 10　English Matriculation Exam Score as A Function of Psychometric Score

To sum up the comparison, Table 6 presents the transition matrix between the psychometric test score and the matriculation exam score.

Table 6　Transition Matrix: Deciles of Scores — Math Matriculation and Psychometric Exams' Scores

		Matriculation Exam										
		1	2	3	4	5	6	7	8	9	10	Total
Psychometric Exam	1	5.2	2.6	1.3	0.5	0.3	0.1	0.1				10
	2	3.1	3.1	2.1	1.2	0.4	0.2	0.1				10
	3	1.1	2.2	2.5	2.1	1.2	0.6	0.2	0.1			10
	4	0.4	1	1.6	1.9	2.1	1.6	0.9	0.4	0.1		10
	5	0.1	0.6	1.3	1.8	1.9	1.9	1.3	0.7	0.3	0.1	10
	6	0.1	0.3	0.7	1.3	1.7	1.9	1.8	1.3	0.8	0.2	10
	7		0.2	0.3	0.8	1.3	1.9	1.8	1.7	1.4	0.6	10
	8			0.1	0.2	0.7	1	1.7	2.4	2.4	1.5	10
	9			0.1	0.1	0.4	0.7	1.4	2	2.6	2.6	10
	10				0.1	0.1	0.3	0.8	1.4	2.3	5	10
	Total	10	10	10	10	10	10	10	10	10	10	100

About 28.45 percent of the population of examinees remains in the same decile of scores, roughly 35 percent are below the diagonal and about 36 percent are above the diagonal. The increase in correlation between the scoring systems can be seen through the fact that as opposed to the previous section where examinees migrated from the highest decile in one of their scores to the lowest decile in their other score because the cells far from the diagonal remained empty. From this we can determine that when we combine the students who completed 5 units with those who completed 4 units, the correlation increases between the two sets of scores.

The variance of the sum or mean of two variables is lower the lower the correlation between the two variables. The high correlation between the two scoring systems is indicative of the fact that there is not a large gain from averaging of the scores and that is because the gain from the averaging is higher the lower the correlation between the scores.

6.4 Mathematics — 3, 4 and 5 Matriculation Units

In this section we examine the relationship between scores of the two exams for all students who were examined in 3, 4 and 5 units, meaning that we are integrating the scores of all of the students who meet the mathematics requirements for acceptance to university. The number of students that were tested in both exams and who meet the requirements for university acceptance was 17,867 examinees from 1999–2000, of whom 4,092 completed 3 units, 6,716 completed 4 units and 7,059 completed 5 units. The weighting of the scores was according to the weighting applied by the universities.

Table 7 Math Tests Parameters

N = 17867	Average	Median	Min	Max	St. Dev.	Gini	Pearson	Spearman	Gini Correlation Γyx	Gini Correlation Γxy
Matriculation exam	100.3	100.5	55	125	14.14	0.082	0.59 (0.000)	0.60 (0.000)	0.6	0.6
Psychometric exam	98.7	101	55	125	15.13	0.088				

Comparison of the scores (Table 7) from the matriculation examinations and the psychometric test shows that the matriculation exam results are higher on average by 1.5 points, though this finding is not significant because it depends on the type of normalization performed. The standard deviation and the Gini coefficient show that the score distribution is much larger in the psychometric exam while the correlation coefficients show a smaller level of correlation between the scores than in the English exams. The Gini correlation coefficients are equal indicating that we can expect symmetry in the distribution of the scores in both exams.

Figure 11 points to higher frequency in the high matriculation exam scores and the fact that the matriculation exam results are distributed more continuously than the psychometric test scores.

Figure 12 presents the NMLA curve of the mathematics psychometric test score as a function of the matriculation exam result. One can see that the graph is concave and smooth indicating a strong monotonic correlation between the psychometric scores as a function of the matriculation exam scores.

Figure 13 presents the NLMA curve of the mathematic matriculation examination result as a function of the psychometric score. As expected, we found that there is a stable monotonic correlation the significance of which is that both of the tests evaluate the same subject.

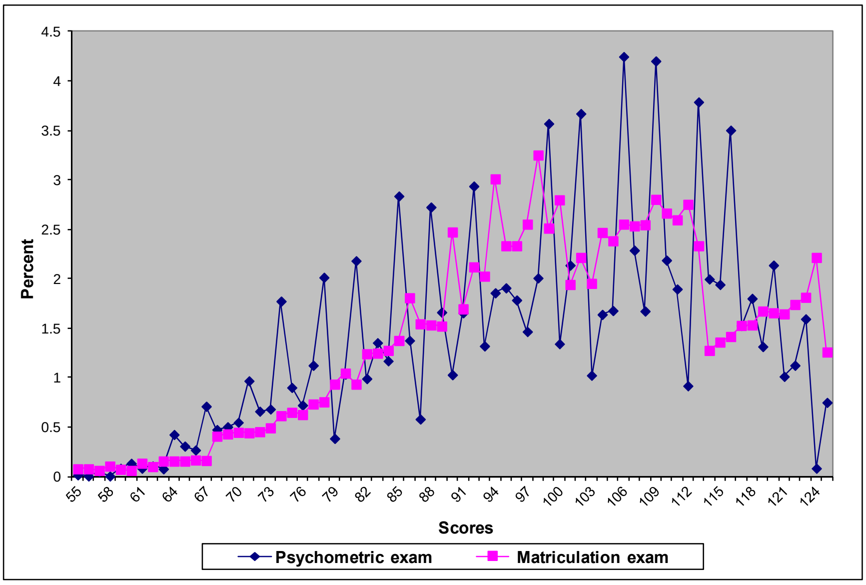

Figure 11 Distribution of Scores of Math Examinees

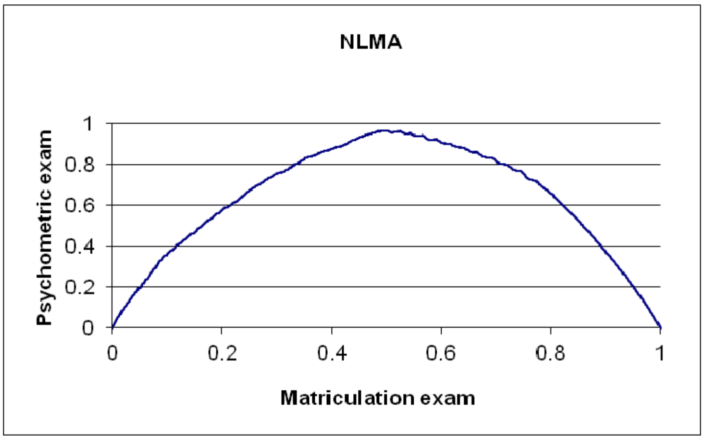

Figure 12 Psychometric Score as A Function of Math Matriculation Exam Score

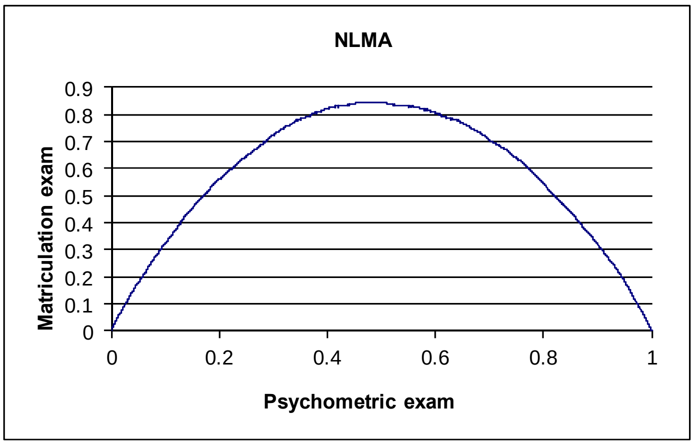

Figure 13 Math Matriculation Exam Score as A Function of Psychometric Score

Table 8 presents the transition matrix from one scoring system to the other. One can see that only 20.2 percent of the exam takers remained on the diagonal meaning that only 20 percent of the students remain in the same decile of scores. Roughly 39 percent of the students were below the diagonal and 41 percent were above the diagonal. One can see that the distribution of the students outside the diagonal is higher in mathematics than in English. However it is important to note that one might get the impression that the number of random errors in each exam is relatively high. Take for example the fact that only 35 percent of the students who were in the highest decile in one of the exams are in the highest decile in the other as well, but that only 40 percent of the students who were in the lowest decile in one of the exams remain in the lowest decile in the other.

6.5 Level of Improvement in Robustness of the Score

In the previous sections we showed, subject to the constraints that grades are an ordinal variable, that the matriculation and psychometric exams in the subjects of mathematics and English, check the same types of abilities. Our conclusion was based on the fact that the scores in the exams form a monotonic transformation of one another: meaning that the higher the scores on the matriculation exams, the higher the psychometric scores and vice versa — the higher the psychometric scores the higher we can expect the matriculation scores to be. In this section we wish to examine the claim that the combination of the two scores from the two exams — the psychometric and matriculation examinations — improves the robustness of the score and that the two exams are better than just one.

Table 8 Transition Matrix: Deciles of Scores — Math Matriculation and Psychometric Exams' Scores

Psychometric Exam	Matriculation Exam 1	2	3	4	5	6	7	8	9	10	Total
1	3.5	2	1.6	1.1	0.9	0.4	0.3	0.2			10
2	2.3	1.6	1.7	1.5	1.3	0.6	0.5	0.4	0.1		10
3	1.3	1.8	1.5	1.5	1.4	0.9	0.7	0.6	0.3	0.1	10
4	1.1	1.5	1.4	1.3	1.2	1.2	1.1	0.7	0.3	0.2	10
5	0.6	0.9	1.1	1.3	1.3	1.3	1.3	1.1	0.6	0.4	10
6	0.4	0.7	0.9	1	1.3	1.3	1.3	1.3	1.1	0.6	10
7	0.4	0.7	0.8	1.1	1.3	1.6	1.1	1.2	1.1	0.8	10
8	0.2	0.4	0.4	0.5	0.5	1.1	1.8	2.1	1.7	1.3	10
9	0.2	0.3	0.4	0.5	0.5	0.9	0.9	1.3	2.5	2.7	10
10	0.1	0.2	0.2	0.3	0.3	0.7	0.8	1.2	2.3	4	10
Total	10	10	10	10	10	10	10	10	10	10	100

To check this claim, we have to abandon the assumption that scores are an ordinal variable and to treat them as a quantitative variable. This assumption enables us to check the contribution to the average scores of each exam. However, to mitigate this strong assumption we checked the effect on the Gini inequality index, that as argued by Lambert and Decoster (2005) "reveals more" than other measures of variability. Note that if the results of the exams are statistically independent of one another than the averaging of the scores will decrease the dispersion (variance or Gini) of the average of the scores to roughly half of the initial variance. However, assuming that the correlation between the scores of the two exams is one then there is no combination of test scores that will decrease distribution of the average score but will rather leave it at the same level. Since the correlations that we found between the score the student received on the psychometric test and the score that he or she received on the matriculation exam are relatively high and greater in English than in mathematics, our expectation is that the contribution of the combination of the exams in English will decrease the distribution of the average score of the student in the two exams less so than the combination of the mathematics scores.

The formula that can be used to decompose the relative Gini index includes as a special case the structure of the decomposition of the coefficient of variation. The formula is as follows:[18]

Let $Y = X_1+X_2$ and $\delta_k = \mu_k/\mu_Y$ (k = 1,2) the relative contribution of the score of the test to the sum of the scores of both of the tests (the inclusive score) then

$$G_Y^2 - G_Y\sum_{k=1}^{2}\delta_k D_{kY} G_k = \sum_{k=1}^{2}\delta_k^2 G_k^2 + \delta_1\delta_2 G_k G_j(\Gamma_{12}+\Gamma_{21}) \qquad (1)$$

when $D_{kY} = \Gamma_{kY} - \Gamma_{Yk}$ and $\Gamma_{12} = \dfrac{\operatorname{cov}(X_1, F(X_2))}{\operatorname{cov}(X_1, F(X_1))}$ is the correlation coefficient for the Gini index.

If $D_{kY}=0$ and $\Gamma_{21} = \Gamma_{12}$ then the resulting equation is

$$G_Y^2 = \sum_{k=1}^{2}\delta_k^2 G_k^2 + 2\,\delta_1\delta_2 G_1 G_j \Gamma_{12} \qquad (2)$$

Where the Formula (2) is parallel to the formula that we would get if we were to ask to see the contribution

[18] Development of the formula can be found in Yitzhaki and Schechtman (2013).

of each component to the coefficient of variation of the total score.[19] Formula (1) fits the Gini index and includes Formula (2) as a special case and the advantage of this formula is that it does not assume that the correlation coefficient is symmetrical between the two score distributions.

Table 9 presents the components of the Gini index in the total score of both tests in English. As one can see, the Gini index for the psychometric test is greater than that of the matriculation examination (0.082 versus 0.077) while the average score is slightly higher in the psychometric than in the matriculation. The index of the weighted score is only slightly lower than that of the matriculation examinations, a result that can be explained by the high correlation that was found between the scores of the two exams.

Table 9 Gini Decomposition — English Matriculation and Psychometric Exams

	English Matriculation Exam	English Psychometric Exam	Total
G	0.077	0.082	0.076
δ	0.498	0.502	
Γ_{ij}	English Matriculation Exam	English Psychometric Exam	Total
English Matriculation Exam		0.827	0.953
English Psychometric Exam	0.834		0.962
סה"כ	0.953	0.595	
G_0^2	$G_0\Sigma\delta_i D_{i0} G_i$	$\delta_i^2 G_i^2$	$\Sigma\Sigma\delta_i\delta_j G_i G_j\ \Gamma_{ij}$
0.00578	-0.00001	0.00316	0.00262

The second part in the table presents the correlation coefficients between the score the matriculation examinee received and the same student's psychometric score. The difference between the correlations is relatively small indicating that there is symmetry in the distribution.[20] The correlation with the average score appears to be similar in both scoring systems and therefore we get that since the average score is based on high correlations between the individual scores there is not much gain from the existence of two exams which are similar in the ranking results of the examinees.

The last row presents the contribution of each component to the breakdown of the Gini index of the average score. The contribution can be attributed roughly 60 percent to the independent components of the score and 40 percent to the correlation between the scores. The correlation between the results of the two exams affects the result of the total score.

Table 10 checks the Gini index of the average score in the matriculation exam in mathematics in comparison with the matriculation score in English. As can be seen, inequality in the mathematics scores is higher than the inequality in the English scores and the inequality in the sum of the results of the exams is smaller than in the inequality in their scores in each exam. An explanation of this result can be derived from the relatively low correlation between the scores (0.5). The symmetry of the correlation indicates that the distribution was normalized in both exams in a similar manner. There is also symmetry of the correlation between each score and the sum of the score in both exams. The relatively low correlation means that it makes more sense to look at the average of the scores in mathematics relative to English exams, since they contribute different independent

[19] The modifying coefficient is the standard deviation divided by the variable average.

[20] This result can stem from the fact that the two sets of scores go through a similar normalization process.

information about the student. The relatively low correlation also explains why there is a decrease in inequality in the sum of the scores in comparison with each score individually.

Table 10 Gini Decomposition — English and Math Matriculation Exams

	English Matriculation Exam	Math Matriculation Exam	Total
G	0.077	0.082	0.068
δ	0.503	0.497	
Γ_{ij}	English Matriculation Exam	Math Matriculation Exam	Total
English Matriculation Exam		0.496	0.861
Math Matriculation Exam	0.496		0.874
Total	0.853	0.871	
G_0^2	$G_0\Sigma\delta_i D_{i0} G_i$	$\delta_i^2 G_i^2$	$\Sigma\Sigma\delta_i\delta_j G_i G_j \Gamma_{ij}$
0.00468	-0.00003	0.00315	0.00156

7. Conclusions

This study serves as preliminary research for the broader question of whether or not we should concern the policy makers responsible for education: to what extent is the existence of two exams on the same subjects an excessive burden on the Israeli economy? The answer to that question can be derived from a cost-benefit analysis of the existence of both exams. In this study we examined a preliminary question, one we must ask before we perform cost-benefit analysis — we asked “to what extent do the tests examine the same types of abilities?”

The answer we received from the study is that the exams do in fact check the same abilities though they do possess a significant random error. The random error in the mathematics evaluations is greater than the in the exams in English and therefore the cost-benefit analysis is crucial for the exam in English because the overlap between those two exams is greater.

The conclusion at which we arrived from the contribution of the two exams to the total score of the student strengthen the claim that the contribution of the double examination does not contribute greatly and therefore doubts are raised regarding the contribution of the existence of two exams to evaluate skills in the same subject. The above conclusion is stronger in the case of English where the correlation between the matriculation exam score and the psychometric score is greater than in mathematics.

It should be emphasized that there is no intent in this study to decide regarding the following question: suppose we received the results of the study and we found that there is in fact an unnecessary burden on the population, which exam should be abandoned? An answer to this question requires a check of the goals and the uses of the exams, the needs of the system of higher education and additional considerations. This study determines that the matriculation and psychometric examinations in parallel subjects (English and Mathematics) do not evaluate different abilities between which we were able to differentiate, and that is based on analysis of the results of the exams.

References

Cliff N. (1977). “A theory of consistency of ordering generalizable to tailored testing”, *Psychometrika*, Vol. 42, pp. 375–401.

Cliff N. (1979). “Test theory without true scores?”, *Psychometrika*, Vol. 44, No. 4, pp. 373–392.

Cliff N. (1992). "Abstract measurement theory and the revolution that never happened", *Psychological Science*, Vol. 3, No. 3, pp. 186–190.

Heene Moritz (2013). "Additive conjoint measurement and the resistance toward falsifiability in psychology, frontiers in psychology", *Quantitative Psychology and Measurement*, Vol. 4, p. 246, available online at: http://www.frontiersin.org.

Lambert P. J. and Decoster A. (2005). "The Gini coefficient reveals more", *Metron*, Vol. LXIII, No. 3, pp. 373–400.

Lord F. M. and Novick M. R. (1968). *Statistical Theories of Mental Test Scores*, Readind, Massachusetts: Adison-Wesley.

Markus K. A. and Borsboom D. (2012). "The cat came back: Evaluating arguments against psychological measurement", *Theory Psychol.*, Vol. 22, pp. 452–466.

Narens Louis and R. D. Luce (1993). "Further comments on the 'non revolution' arising from axiomatic measurement theory", *Psychological Science*, Vol. 4, No. 2, pp. 127–130.

Schechtman E., Yitzhaki S. and Yevgeny Artsev (2008). "Who does not respond in the household expenditure survey: An exercise in extended Gini regressions", *Journal of Business & Economic Statistics*, Vol. 26, No. 3, pp. 329–344.

Schechtman E. and Yitzhaki S. (2009). "Ranking groups' abilities — Is it always Reliable?", *International Journal of Testing*, Vol. 9, No. 3, pp. 195–214.

Yitzhaki S. (2003). "Gini's mean difference: A superior measure of variability for non-normal distributions", *Metron*, Vol. LXI, No. 2, pp. 285–316.

Yitzhaki S. and Maggie Eisenstaedt (2003). "Ranking groups' versus individuals' ranking", in: Amiel, Yoram and John A. Bishop (Eds.), *Fiscal Policy, Inequality, and Welfare: Research on Economic Inequality*, Vol. 10, pp. 101–123.

Yitzhaki S. and E. Schechtman (2012). "Identifying monotonic and non-monotonic relationship", *Economics Letters*, Vol. 116, pp. 23–25.

Yitzhaki S., Rinat Itzhaki and Taina Pudalov (2012). "A nonparametric ICC using the Gini's mean difference approach", unpublished, available online at: http://ssrn.com.

Yitzhaki S. and E. Schechtman (2013). *The Gini Methodology: A Primer on a Statistical Methodology*, Springer: N.Y. Forthcoming.