דוח המרכז

349

Fairness in the Higher Education Admissions Procedure: The Psychometric Entrance Test in Arabic

Elliot Turvall
Shmuel Bronner
Tamar Kennet-Cohen
Carmel Oren

July 2008

מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
מיסודן של האוניברסיטאות בישראל
ת"ד 26015 ירושלים 91260, טל' 02-6759555

דוח מרכז 349

ISBN:965-502-143-2

Fairness in the Higher Education Admissions Procedure: The Psychometric Entrance Test in Arabic

Elliot Turvall, Shmuel Bronner, Tamar Kennet-Cohen and Carmel Oren
National Institute for Testing and Evaluation
Jerusalem, Israel

Paper presented at the annual meeting of the
Israeli Psychometric Association, Jerusalem, Israel, 2007

Abstract

The aim of the study was to examine whether the higher education admissions process in Israel discriminates against the Arab population. The specific measures examined were the criterion – the grade-point average at the end of the first year of university studies (FGPA) and six predictors, namely the high school matriculation certificate (HSM), the Psychometric Entrance Test (PET), its three subtests (Verbal, Quantitative and English) and the composite admission score (a combination of the HSM and the PET). The question of fairness was analyzed from two points of view, differential validity and differential prediction.

It was found that the validity of the predictors is considerably higher among the examinees in Hebrew than among the examinees in Arabic. Regarding the differential prediction it was found that in very few cases, the FGPA is under-predicted for the examinees in Arabic, whether the predictor is the HSM, the PET or the composite admission score. On the other hand, the criterion is over-predicted in quite a number of instances, especially when the HSM is used as the predictor. The effect size of the differences between the examinees in Arabic and in Hebrew was also calculated, separately for the academic departments with under- and over-prediction, as well as for the unbiased ones.

Table of Contents

Introduction

Recent research (Kennet-Cohen, Cohen and Oren, 2005) has indicated that members of the Arab sector perform less well than members of the Jewish sector on exams that measure learning achievements in general, and on higher education admissions tests in particular. In view of this fact, it has been claimed that higher education entrance exams discriminate against the Arab population. Moreover, since university studies are regarded as the key to success in modern society, a lot of public attention is directed at this question.

The conventional method of studying predictive bias is to consider the predictor in relation to the criterion for which it was created (Linn, 1984). This study examined the fairness of the university admissions process vis-à-vis Arabic-speaking examinees by relating the criterion of success in university studies to the predictors.

The question of bias was examined from two points of view (Linn, 1982):

1. Differential validity – analysis of the degree of agreement between the predictors and the criterion, across the two groups of examinees (Arabic- and Hebrew-speaking). Since Arabic-speaking examinees obtain considerably lower scores than Hebrew-speaking examinees, their scores might be expected to have lower reliability and, therefore, lower validity.
2. Differential prediction – which focuses on differences in the predicted values of the criterion, between the two types of candidates. The effect size of the differences was also calculated separately for academic departments with under- and over-prediction, as well as for the unbiased departments.

Fairness was examined, from both points of view, within academic departments grouped according to faculty and, in some cases, also according to academic year.

Method

Sample

The analyses were carried out on data for 41,314 first-year students in 374 academic departments from eight different faculties (Arts and Humanities, Social Sciences – Verbal[1], Law, Social Sciences – Quantitative[2], Natural Sciences, Engineering and Architecture, Medicine and Nursing), in Israel's six research universities over five academic years: 1995/1996, 1996/1997, 1997/1998, 1998/1999 and 2002/2003. (The data for the years 1999/2000, 2000/2001 and 2001/2002 were missing.)

Information for each student included the high school matriculation certificate average (HSM), the total score on the Psychometric Entrance Test (PET) and its components (Verbal Reasoning, Quantitative Reasoning and English as a second language), and the reported first year grade-point average from their respective universities (FGPA). The academic departments were selected on the condition that they included at least five students who had taken the PET in Hebrew and five who had taken the PET in Arabic. (This condition, along with the deletion of students with missing values, reduced the number of students from the original total of 148,667.)

Table 1 shows the distribution of students and academic departments, according to language, faculty and academic year.

[1] Sociology, Political Sciences, International Relations, Psychology, Education, etc.
[2] Economics, Statistics, Accounting, Business Administration, etc.

Table 1: Distribution of students and academic departments according to language, faculty and academic year

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	1,336* 344 20	1,085 414 19	1,109 430 22	959 503 22	770 298 19	**5,259** **1,989** **102**
Social Sciences – Verbal	2,797 376 18	2,742 466 18	2,478 435 19	2,570 465 20	2,669 788 21	**13,256** **2,530** **96**
Law	993 61 4	617 90 3	286 48 2	277 68 2	604 65 3	**2,777** **332** **14**
Social Sciences – Quantitative	758 99 5	661 88 5	627 65 6	678 81 6	592 95 6	**3,316** **428** **28**
Natural Sciences	487 148 6	826 159 9	785 145 10	743 166 11	559 219 9	**3,400** **837** **45**
Engineering and Architecture	487 88 6	577 107 5	540 90 4	611 100 6	333 50 4	**2,548** **435** **25**
Medicine	250 59 4	217 51 4	289 68 5	270 51 5	342 134 7	**1,368** **363** **25**
Nursing	149 59 5	303 57 6	335 94 7	428 132 8	645 274 13	**1,860** **616** **39**
Total	**7,257** **1,234** **68**	**7,028** **1,432** **69**	**6,449** **1,375** **75**	**6,536** **1,566** **80**	**6,514** **1,923** **82**	**33,784** **7,530** **374**

*First row: Number of examinees in Hebrew
Second row: Number of examinees in Arabic
Third row: Number of academic departments

Variables

Predictors

Six predictors, which are briefly discussed below, were used in this study:

1. Composite admission score (Composite)

   This score consists of equally weighted high school grades (HSM) and Psychometric Entrance Test (PET) scores. (These weights were applied at the level of the candidates for each university.) This is the means whereby in general candidates are admitted to higher education in Israel. (It ranges from 0 to 100.)

2. High school matriculation certificate average (HSM)

High school graduates receive a high school matriculation certificate, which is based on a combination of high school grades and scores on national tests in several subjects. (It ranges from 50 to 120.)

3. The Psychometric Entrance Test (PET)

The PET is designed to assess various cognitive and scholastic abilities and predict success in future academic studies. It ranges from 200 to 800, with a mean of 500 and a standard deviation of 100. It includes three subsections of multiple-choice questions. In the total PET score the subtests are weighted as follows: Verbal 40%, Quantitative 40% and English 20%. For those students who took the PET several times, the highest score obtained before the start of the academic year was used in this study.

4. The verbal reasoning subsection of the Psychometric Entrance Test (Verbal)

The verbal subtest of the PET includes 60 items focusing on verbal skills and abilities needed for higher education: the ability to analyze and understand complex written texts, the ability to think systematically and logically and the ability to perceive subtle distinctions and nuances among words and concepts. It ranges from 50 to 150, with a mean of 100 and a standard deviation of 20.

5. The quantitative reasoning subsection of the Psychometric Entrance Test (Quantitative)

The quantitative subtest of the PET includes 50 items focusing on the ability to use numbers and mathematical concepts to solve quantitative problems as well as the ability to analyze information presented in the form of graphs, tables and charts. It ranges from 50 to 150, with a mean of 100 and a standard deviation of 20.

6. English as a second language subsection of the Psychometric Entrance Test (English)

The English subtest of the PET includes 54 items focusing on the ability to read and understand English texts at an academic level. It ranges from 50 to 150, with a mean of 100 and a standard deviation of 20.

Criterion

The criterion used in this study was the grade-point average at the end of the first year of university studies (FGPA). (It ranges from 10 to 100.)

In tables 2A and 2B, the mean and standard deviation of the predictors and the criterion are shown, according to language and faculty. (The statistics were computed within academic departments, weighted by the number of students in that department and averaged across departments.)

Table 2A: Mean and standard deviation of the criterion and the three main predictors according to language and faculty

Faculty	**FGPA**		**Composite**		**HSM**		**PET**	
	Mean	***Sd***	**Mean**	***Sd***	**Mean**	***Sd***	**Mean**	***Sd***
Arts and	80.5*	*8.9*	47.1	*9.1*	88.1	*8.3*	543.1	*74.8*
Humanities	67.3	*11.3*	41.8	*6.4*	89.1	*7.4*	457.3	*48.2*
Social Sciences –	83.2	*6.9*	50.0	*7.0*	90.1	*7.1*	561.0	*62.4*
Verbal	74.3	*8.8*	48.0	*5.6*	94.3	*6.5*	494.6	*45.9*
Law	80.6	*6.4*	64.6	*4.3*	102.2	*5.3*	683.8	*39.8*
	70.6	*8.2*	62.8	*3.6*	106.9	*4.2*	615.6	*36.9*
Social Sciences –	77.1	*10.4*	59.7	*5.0*	97.1	*6.2*	659.0	*39.6*
Quantitative	70.0	*13.0*	60.3	*3.7*	105.1	*4.6*	596.1	*37.7*
Natural Sciences	75.1	*13.0*	57.4	*6.1*	96.9	*6.4*	635.0	*53.6*
	65.1	*13.7*	54.4	*4.8*	102.0	*5.3*	551.6	*43.0*
Engineering and	75.3	*8.4*	57.3	*4.2*	101.5	*5.4*	659.6	*42.1*
Architecture	68.8	*10.6*	54.9	*3.4*	107.0	*3.8*	589.8	*33.8*
Medicine	83.9	*6.7*	65.4	*3.6*	105.2	*5.1*	706.9	*27.3*
	82.3	*6.9*	66.3	*2.3*	111.6	*2.8*	666.8	*26.5*
Nursing	82.8	*6.0*	51.7	*5.7*	92.8	*6.3*	582.9	*49.7*
	77.8	*6.8*	53.1	*4.0*	99.8	*5.2*	540.1	*34.0*
Total	**80.6**	***8.4***	**53.5**	***6.7***	**93.6**	***6.8***	**598.1**	***57.7***
	71.6	***10.2***	**51.4**	***5.1***	**98.0**	***5.9***	**529.0**	***42.6***

*First row: Examinees in Hebrew
Second row: Examinees in Arabic

Table 2B: Mean and standard deviation of the three subsections of the PET according to language and faculty

Faculty	**Verbal**		**Quantitative**		**English**	
	Mean	***Sd***	**Mean**	***Sd***	**Mean**	***Sd***
Arts and Humanities	109.5* 93.9	*16.0* *11.0*	103.6 92.5	*15.2* *12.5*	111.4 88.8	*17.6* *11.9*
Social Sciences – Verbal	112.7 99.0	*13.6* *11.1*	108.2 102.0	*13.6* *11.8*	111.6 92.7	*16.6* *11.8*
Law	132.5 117.0	*9.4* *10.8*	131.5 126.3	*9.6* *10.2*	133.5 115.0	*11.2* *12.5*
Social Sciences – Quantitative	126.3 113.0	*10.1* *11.1*	130.2 125.8	*8.8* *9.3*	127.0 106.9	*13.1* *13.7*
Natural Sciences	121.0 106.4	*12.7* *11.0*	126.4 116.6	*10.6* *10.8*	123.7 99.0	*15.4* *13.1*
Engineering and Architecture	123.4 111.3	*11.4* *10.3*	133.0 124.4	*8.2* *8.3*	127.5 107.3	*13.3* *14.5*
Medicine	135.3 127.0	*7.9* *8.8*	136.6 134.1	*7.1* *7.2*	137.8 124.4	*8.4* *10.9*
Nursing	115.4 106.2	*12.0* *9.5*	112.5 111.3	*11.4* *9.9*	116.8 99.9	*15.3* *12.4*
Total	**117.6** **104.0**	***13.0*** ***10.8***	**116.2** **109.3**	***12.3*** ***11.0***	**118.4** **98.4**	***15.4*** ***12.4***

*First row: Examinees in Hebrew
Second row: Examinees in Arabic

Differential Validity

The differential validity was investigated using the correlation coefficient between each of the six predictors and the criterion (FGPA). It was computed separately for the two languages, within academic departments, then weighted by the number of students in that department and averaged across departments. As a point of reference, the validity for all the 91,584 first-year students (Hebrew- and Arabic-speaking) was also calculated, as well as for the 83,108 Hebrew-speaking first-year students separately.

The validity was corrected for range restriction, since it was computed on the selected group of examinees only. This adjustment requires using the standard deviation of the explicit selection variable in an unselected sample. The composite admission score was treated as the explicit selection variable. Its standard deviation in an unselected sample was estimated by a weighted average of its standard deviation among applicants to an academic department (by university and academic year). These estimates were based on data for applicants to all Israeli universities during the

academic years of 1991/1992 and 1992/1993. (A detailed description of the correction method can be found in Kennet-Cohen, Bronner, & Oren, 1999.)

Differential Prediction

A second aspect of bias is differential prediction, which was used in order to detect departments with a systematic under- or over-prediction of the criterion for the examinees in Arabic. In this study the technique used was the boundary conditions analysis suggested by Birnbaum (1979, 1981; Linn 1984), in order to detect differential prediction. This method neutralizes the risk of reporting bias due to statistical artifacts that may result from use of Cleary's regression model (Cleary, 1968).

Principles of Boundary Conditions

Systematic bias, or the absence thereof, can be depicted in the form of a path diagram (Birnbaum, 1979, 1981). Figure 1 shows a variation of this illustration (Linn, 1984).

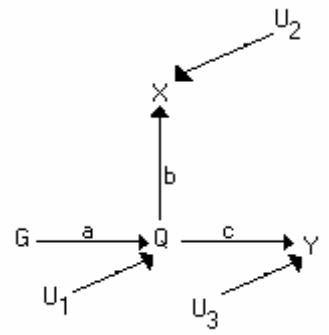

Figure 1: Birnbaum's path diagram illustrating a condition of no bias (Linn, 1984)

In figure 1, the observable variables are G, a dichotomous variable denoting membership in a group (in our case examinees in Hebrew or Arabic), X, a predictor and Y, the criterion. The unobservable variables are Q, an idealized underlying ability, as well as U_1, U_2 and U_3, representing mutually uncorrelated disturbance factors. The relevant path coefficients are called a, b and c.

The illustration in figure 1 shows an unbiased situation. The two groups are allowed to differ in the latent ability, Q, and as a result of this also in the predictor, X,

and the criterion, Y. However, the differences between the groups in the observable variables are due only to differences in the underlying true ability, Q. In other words, there are no direct, non-zero paths from G to X or from G to Y. Had there been non-zero paths from G to X or from G to Y, this would imply systematic bias attributable to belonging to a specific group.

Hence, in the unbiased case the partial regression coefficients of X and Y on G should be zero, when Q is held constant, i.e.

$$\beta_{\mathrm{XG.Q}} = 0 \tag{1}$$

and

$$\beta_{YG.Q} = 0\,. \tag{2}$$

The problem is that Q is unobservable, making it impossible to estimate the regression coefficients in equations (1) and (2). However, Birnbaum (1979, 1981) showed that his model in the unbiased case implies certain constraints on regression coefficients involving only observable variables. Consequently we can define boundary conditions in terms of these regression coefficients. If the conditions are not satisfied, we can conclude that the situation is biased, either with under- or over-prediction towards the group of interest (in our case, the examinees in Arabic).

To begin with, to detect underestimation of the true ability of the group of interest, expressed in under-prediction of the criterion of this group in relation to its predictor, we would study the regression of Y on X and G. This is actually Cleary's regression model of investigating bias (Cleary, 1968). The partial regression coefficient for G, under the assumption of nonbias, in terms of the unknown path coefficients, a, b and c, from figure 1, is

$$\beta_{\mathrm{YG.X}} = \frac{\mathrm{ac}(1-\mathrm{b}^2)}{1-(\mathrm{ab})^2}\,. \tag{3}$$

If we assume parallel within-group slopes, the regression coefficient in equation (3) is proportional to the difference between the within-group intercepts. Thus, this regression coefficient should equal 0 in the unbiased case. However, this will be true only in extreme and unrealistic cases, when a=0 (The latent ability is the same in both groups.), c=0 (The underlying ability does not influence the criterion.) and b=1 (The predictor is a perfectly valid and reliable measure of the true ability.).

However, in a realistic case, it is assumed that both b and c are positive. (Examinees with higher latent ability get better results, both on the predictor and on

the criterion.) The remaining path coefficient, a, may be either positive or negative, depending on how the dichotomous group variable, G, is coded. If we assume that G is coded so that a is positive, then the regression coefficient in equation (3) is also positive. Thus, if we are in an unbiased situation regarding the latent variables, such as that depicted in figure 1, then the regression coefficient $\beta_{YG.X} > 0$.

This means that $\beta_{YG.X} > 0$ might indicate bias, but does not necessarily. On the other hand, if we find that $\beta_{YG.X} < 0$, we can conclude that the situation is biased and that the criterion for the group which has lower mean scores on the predictor is underestimated.

To detect overestimation of the criterion for the group that has lower mean scores on the predictor, we look at the partial regression of X on Y and G. This regression model is a special case of Cole's (1973) conditional-probability model. The regression coefficient of interest, under the assumption of nonbias, as in figure 1, expressed in the path coefficients, will be

$$\beta_{\mathrm{XG.Y}} = \frac{\mathrm{ab}(1-\mathrm{c}^2)}{1-(\mathrm{ac})^2}. \qquad (4)$$

If we again assume that a, b and c are positive and smaller than 1, then $\beta_{XG.Y} > 0$. This means that $\beta_{XG.Y} > 0$ might indicate bias, but does not necessarily. On the other hand, if we find that $\beta_{XG.Y} < 0$, we can conclude that the situation is biased and that the criterion for the group that has lower mean scores on the predictor is overestimated.

To sum up, if we find that $\beta_{YG.X} < 0$ in a specific academic department, we can conclude that the group which has lower mean scores on the predictor is discriminated against. Finding that $\beta_{XG.Y} < 0$ implies that the criterion for the group with lower mean scores on the predictor is overestimated. If neither of these two equations is fulfilled, we conclude that the situation is unbiased.

Implementing the Boundary Conditions

In order to implement the differential prediction, two multivariate regression equations were calculated: the first one with the criterion as the dependent variable and each predictor, membership in a group (0=Arabic and 1=Hebrew), and the interaction between the latter two, as independent variables; the second one with each predictor as the dependent variable and the criterion, membership in a group (0=Arabic and 1=Hebrew), and the interaction between the latter two, as independent variables.

With regard to the first regression equation, where the criterion is the dependent variable, we calculated a correlation matrix with the observed correlations between the criterion, the predictor, the membership in a group and the interaction between the latter two. These correlations were thereafter corrected for range restriction as in the case of corrected validity (see above). Next, the regression analysis was implemented using the corrected correlations. The conclusions were drawn from the regression in the following way: When the coefficient for the variable describing membership in a group is positive (meaning that the regression line for the examinees in Hebrew is above the regression line for the examinees in Arabic), we conclude that there is over-prediction for the examinees in Arabic, if the coefficient of the interaction term is smaller, in absolute values, than the coefficient for the variable describing membership in a group. Similarly, when the coefficient for the variable describing membership in a group is negative (meaning that the regression line for the examinees in Arabic is above the regression line for the examinees in Hebrew), we conclude that there is under-prediction for the examinees in Arabic, if the coefficient of the interaction term is smaller, in absolute values, than the coefficient for the variable describing membership in a group.

The second regression, where the predictor is the dependent variable, is implemented in parallel. Conclusions are drawn as above, except that they are inverted. When the coefficient for the variable describing membership in a group is positive (meaning that the regression line for the examinees in Hebrew is above the regression line for the examinees in Arabic) and if the coefficient of the interaction term is smaller, in absolute values, than the coefficient for the variable describing membership in a group, we conclude that there is over-prediction of the predictor for the examinees in Arabic. This means that there is under-prediction for the criterion,

which is of key interest to us. Similarly, when the coefficient for the variable describing membership in a group is negative (meaning that the regression line for the examinees in Arabic is above the regression line for the examinees in Hebrew), we conclude that there is over-prediction of the criterion for the examinees in Arabic, if the coefficient of the interaction term is smaller, in absolute values, than the coefficient for the variable describing membership in a group.

If both regressions show bias in the same direction (under- or over-prediction), we conclude that there is bias in this direction. Otherwise we conclude that the situation is unbiased. (A detailed description of this method can be found in Kennet-Cohen, 2001.)

Effect Size

Effect size was computed, according to Cohen (1988), as the difference between the means for the two groups, divided by a pooled estimate of the standard deviation. This was done separately for the academic departments with under- and over-prediction, as well as for the unbiased ones. The effect size was computed within academic departments, weighted by the number of students in that department and averaged across departments. The effect size for all academic departments was also computed as a point of reference.

The effect size was also corrected for range restriction, via corrected correlations between a variable denoting membership in a group and the predictor/criterion of interest. (A detailed description of this method can be found in Kennet-Cohen, 2001.)

Results

Differential Validity

The differential validity was investigated by means of the correlation coefficient between each of the six predictors and the criterion (FGPA). In table 3 the corrected validity for the predictors is shown, according to language and faculty.

Table 3: Corrected validity for the predictors, according to language and faculty

Faculty	Composite	HSM	PET	Verbal	Quantitative	English
Arts and Humanities	0.44* 0.36	0.39 0.33	0.36 0.28	0.32 0.23	0.29 0.17	0.27 0.21
Social Sciences – Verbal	0.42 0.31	0.37 0.28	0.36 0.25	0.31 0.20	0.30 0.17	0.24 0.16
Law	0.65 0.25	0.63 0.24	0.50 0.18	0.41 0.17	0.42 0.12	0.31 0.11
Social Sciences – Quantitative	0.46 0.48	0.42 0.52	0.36 0.32	0.23 0.21	0.38 0.26	0.20 0.20
Natural Sciences	0.59 0.40	0.52 0.37	0.50 0.33	0.39 0.16	0.48 0.36	0.33 0.19
Engineering and Architecture	0.59 0.51	0.53 0.43	0.46 0.46	0.35 0.25	0.48 0.45	0.30 0.23
Medicine	0.58 0.56	0.56 0.60	0.45 0.45	0.29 0.38	0.40 0.25	0.31 0.22
Nursing	0.52 0.49	0.46 0.46	0.43 0.30	0.32 0.11	0.36 0.26	0.30 0.23
Total	**0.49** **0.38**	**0.44** **0.35**	**0.40** **0.29**	**0.32** **0.21**	**0.35** **0.22**	**0.26** **0.19**

*First row: Examinees in Hebrew
Second row: Examinees in Arabic

Overall, the validity for PET examinees in Hebrew is much higher than for examinees in Arabic. This is true for all predictors and almost all faculties.

As a point of reference, the corrected validity for all the 91,584 first-year students (Hebrew- and Arabic-speaking) was also calculated as well as for the 83,108 Hebrew-speaking first-year students separately. In table 4 the corrected validity for the predictors is shown, according to faculty.

Table 4: Corrected validity for the predictors among all the first-year students, according to faculty, as well as for the Hebrew-speaking first-year students separately

Faculty	Composite	HSM	PET	Verbal	Quantitative	English
Arts and Humanities	0.44*	0.35	0.41	0.36	0.31	0.32
	0.44	0.39	0.36	0.32	0.29	0.26
Social Sciences – Verbal	0.43	0.31	0.42	0.37	0.33	0.32
	0.45	0.40	0.38	0.32	0.32	0.25
Law	0.59	0.51	0.53	0.47	0.41	0.40
	0.62	0.60	0.49	0.41	0.40	0.32
Social Sciences – Quantitative	0.49	0.43	0.40	0.27	0.40	0.26
	0.50	0.47	0.38	0.25	0.39	0.23
Natural Sciences	0.54	0.45	0.47	0.36	0.47	0.32
	0.54	0.49	0.45	0.34	0.45	0.29
Engineering and Architecture	0.55	0.48	0.44	0.34	0.45	0.32
	0.54	0.51	0.42	0.32	0.42	0.28
Medicine	0.50	0.45	0.37	0.26	0.35	0.24
	0.49	0.48	0.38	0.25	0.33	0.27
Nursing	0.46	0.35	0.44	0.34	0.34	0.34
	0.52	0.46	0.43	0.33	0.36	0.28
Total	**0.48**	**0.39**	**0.43**	**0.35**	**0.37**	**0.32**
	0.49	**0.45**	**0.40**	**0.32**	**0.36**	**0.27**

*First row: All the 91,584 first-year students (Hebrew- and Arabic-speaking)
Second row: All the 83,108 Hebrew-speaking first-year students

Differential Prediction

The differential prediction was calculated as specified above. Table 5 shows the distribution of academic departments with under- and over-prediction in the criterion, as well as the academic departments without bias, according to predictor and faculty, for examinees in Arabic.

Table 5: Distribution of academic departments with under- and over-prediction in the criterion, as well as the academic departments without bias, according to predictor and faculty, for examinees in Arabic

Faculty	Composite	HSM	PET	Verbal	Quantitative	English
Arts and Humanities	2*	0	5	0	1	2
	78	48	97	102	92	99
	22	54	0	0	9	1
Social Sciences – Verbal	0	0	0	0	1	0
	70	37	95	95	88	95
	26	59	1	1	7	1
Law	0	0	0	0	0	0
	10	7	14	14	14	14
	4	7	0	0	0	0
Social Sciences – Quantitative	0	0	1	0	0	0
	23	16	27	28	27	28
	5	12	0	0	1	0
Natural Sciences	0	0	5	1	2	3
	39	23	40	43	43	42
	6	22	0	1	0	0
Engineering and Architecture	0	0	1	1	0	2
	25	19	24	24	25	23
	0	6	0	0	0	0
Medicine	0	0	2	1	0	4
	24	20	23	24	25	21
	1	5	0	0	0	0
Nursing	2	0	1	1	0	3
	20	14	38	38	27	36
	17	25	0	0	12	0
Total	**4**	**0**	**15**	**4**	**4**	**14**
	289	**184**	**358**	**368**	**341**	**358**
	81	**190**	**1**	**2**	**29**	**2**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic

As table 5 shows, the composite admission score under-predicts the FGPA for the examinees in Arabic in only four academic departments. However, it over-predicts the criterion in about 22% (81/374) of the academic departments, in all faculties except Engineering and Architecture (no academic departments) and Medicine (one academic department).

The HSM does not under-predict the FGPA for examinees in Arabic in any academic departments. However, it over-predicts the criterion in a little more than half of the academic departments. This phenomenon is especially important in the faculties of Arts and Humanities, Social Sciences – Verbal, Law and Nursing.

Regarding the PET, we conclude that it over-predicts the FGPA for the examinees in Arabic in only one academic department. However, it under-predicts the criterion in about 4% (15/374) of the academic departments, mainly in the faculties of Arts and Humanities and Natural Sciences.

The verbal subtest of the PET is the most precise of all predictors, with roughly 98% (368/374) of the academic departments in the group without bias for examinees in Arabic. The quantitative subtest of the PET under-predicts the criterion for examinees in Arabic in four academic departments, whereas it over-predicts the FGPA in almost 8% (29/374) of the academic departments, mainly in the faculties of Arts and Humanities, Social Sciences – Verbal and Nursing. The English subtest of the PET shows a pattern similar to that of the entire PET (see above).

The number of academic departments with under- and over-prediction in the criterion, as well as the academic departments without bias, regarding examinees in Arabic, is consistent over the academic years included in this study, as is evident from tables 7, 8, 9, 10, 11 and 12 in the appendix.

Effect Size

Table 6 shows the corrected effect size (Arabic-Hebrew) in the academic departments with under- and over-prediction, as well as in the unbiased ones, where the composite admission score was used to divide the academic departments into the different groups, according to faculty and predictor. As a point of reference, the corrected effect size for all academic departments is also included. The number of academic departments in each combination can be found in table 5.

Table 6: Corrected effect size (Arabic-Hebrew) in the academic departments with under- and over-prediction, as well as in the unbiased ones, and in all departments, where the composite admission score was used for the division, according to faculty and predictor

Faculty	FGPA	Composite	HSM	PET	Verbal	Quantitative	English
Arts and Humanities	-0.33*	**-1.39**	-0.91	-1.75	-0.83	-1.30	-2.74
	-1.76	**-0.89**	0.06	-1.71	-1.42	-1.04	-1.81
	-2.37	**-0.30**	0.69	-1.49	-1.35	-0.73	-1.83
	-1.83	**-0.81**	0.14	-1.67	-1.40	-0.99	-1.83
Social Sciences – Verbal	-	-	-	-	-	-	-
	-1.43	**-0.61**	0.29	-1.39	-1.29	-0.76	-1.43
	-1.65	**0.45**	1.38	-0.73	-0.79	-0.08	-1.11
	-1.49	**-0.33**	0.59	-1.21	-1.15	-0.58	-1.35
Law	-	-	-	-	-	-	-
	-1.43	**-0.66**	0.10	-1.62	-1.64	-0.74	-1.73
	-1.43	**0.34**	1.15	-0.96	-1.00	-0.25	-1.11
	-1.43	**-0.45**	0.32	-1.48	-1.50	-0.64	-1.60
Social Sciences – Quantitative	-	-	-	-	-	-	-
	-0.65	**0.32**	1.20	-1.27	-1.14	-0.42	-1.37
	-0.72	**0.41**	1.18	-1.30	-1.04	-0.48	-1.20
	-0.66	**0.34**	1.20	-1.27	-1.13	-0.43	-1.34
Natural Sciences	-	-	-	-	-	-	-
	-1.09	**-0.89**	0.44	-2.05	-1.64	-1.32	-2.17
	-0.95	**0.25**	1.38	-1.24	-0.84	-0.66	-1.67
	-1.06	**-0.67**	0.62	-1.89	-1.49	-1.19	-2.08
Engineering and Architecture	-	-	-	-	-	-	-
	-1.28	**-1.30**	0.19	-2.21	-1.69	-1.61	-2.16
	-	-	-	-	-	-	-
	-1.28	**-1.30**	0.19	-2.21	-1.69	-1.61	-2.16
Medicine	-	-	-	-	-	-	-
	-0.01	**0.77**	1.85	-1.19	-1.08	-0.16	-1.92
	0.76	**2.17**	2.80	-0.68	-0.03	-0.19	-1.32
	0.03	**0.85**	1.90	-1.17	-1.02	-0.16	-1.88
Nursing	0.02	**-1.24**	-0.06	-2.52	-2.59	-0.27	-2.38
	-1.11	**-0.05**	0.94	-1.42	-1.24	-0.51	-1.82
	-0.83	**1.48**	2.55	-0.55	-0.82	0.60	-1.26
	-0.94	**0.72**	1.76	-0.99	-1.05	0.08	-1.54
Total	**-0.23**	**-1.35**	**-0.66**	**-1.98**	**-1.34**	**-1.00**	**-2.64**
	-1.28	**-0.58**	**0.43**	**-1.59**	**-1.39**	**-0.88**	**-1.70**
	-1.45	**0.49**	**1.44**	**-0.91**	**-0.90**	**-0.16**	**-1.29**
	-1.31	**-0.35**	**0.64**	**-1.45**	**-1.28**	**-0.72**	**-1.62**

*First row: Academic departments with under-prediction for examinees in Arabic, according to the composite admission score

Second row: Academic departments without bias for examinees in Arabic, according to the composite admission score

Third row: Academic departments with over-prediction for examinees in Arabic, according to the composite admission score

Fourth row: All academic departments

As expected, the corrected effect size for the composite admission score becomes greater (positive) as we move from academic departments with under-prediction for examinees in Arabic, to academic departments without bias for examinees in Arabic and then to academic departments with over-prediction for examinees in Arabic. This pattern is true for most other combinations of faculty and predictor as well.

When studying the effect size for all academic departments, we conclude that the effect size for the composite admission score is higher than that of the FGPA, indicating that the criterion is overestimated for the examinees in Arabic. If we compare the HSM and the PET, the effect size for the PET is a lot closer to that of the FGPA, although somewhat smaller. The effect size for the HSM, however, indicates a considerable overestimation of the criterion for the examinees in Arabic. The results for the three subtests of the PET show that the examinees in Arabic, compared to the examinees in Hebrew, get the highest mean scores on the quantitative section, followed by the verbal and the English sections.

Discussion

The purpose of the present study was to investigate whether the Arab sector is discriminated against by the higher education admissions process in Israel, from the perspective of predictive validity. The criterion was the grade-point average at the end of the first year of university studies (FGPA) and the predictors were the high school matriculation certificate (HSM), the Psychometric Entrance Test (PET), its three subtests (Verbal, Quantitative and English) and the composite admission score (a combination of the HSM and the PET). The data came from 41,314 first-year students in 374 academic departments in eight different faculties at six Israeli universities during five academic years.

The question of bias was analyzed from two points of view, differential validity and differential prediction. The results show that the validity is higher for examinees in Hebrew than it is for examinees in Arabic. This is true for all predictors and almost all faculties. This might be interpreted as indicating that the admissions process is biased against the examinees in Arabic. However, one explanation of this phenomenon is that it results from the lower reliability for examinees in Arabic, compared to examinees in Hebrew.

The analysis of the differential prediction shows that the composite admission score (which is the means whereby candidates are admitted to higher education in Israel) over-predicts the FGPA for examinees in Arabic in about 22% (81/374) of the academic departments. It under-predicts the FGPA in four academic departments.

The HSM over-predicts the FGPA for examinees in Arabic in a little more than half of the academic departments, whereas it does not under-predict the FGPA in any academic department. The over-prediction is especially relevant in the faculties of Arts and Humanities, Social Sciences – Verbal, Law and Nursing.

The PET under-predicts the FGPA for examinees in Arabic in about 4% (15/374) of the academic departments, mainly in the faculties of Arts and Humanities and Natural Sciences. However, it only over-predicts the FGPA in one academic department. The verbal subtest of the PET is the most precise of all predictors, with roughly 98% (368/374) of the academic departments in the group without bias for examinees in Arabic. The quantitative subtest of the PET under-predicts the criterion for examinees in Arabic in four academic departments, whereas it over-predicts the FGPA in almost 8% (29/374) of the academic departments, mainly in the faculties of

Arts and Humanities, Social Sciences – Verbal and Nursing. The English subtest of the PET shows a pattern similar to that of the entire PET (see above).

When compared with previous studies, the findings in the present one regarding validity are very similar to the corresponding results found by Bronner (2004). However, it should be mentioned that roughly half the students in the current study also took part in that research.

However, the current research shows greater differences in validity between the two population sectors, both regarding the corrected validity (where different methods of correcting for range restriction were used in the different studies) and the observed validity, than an earlier study conducted by Bronner, Allalouf and Oren (1996). Yet, the differential prediction and the effect size are similar in both studies.

An issue that deserves to be discussed here is the sampling method. As was explained above, the academic departments were selected provided they included at least five students who took the PET in Hebrew and five who took it in Arabic. This condition dramatically reduced the number of academic departments in a manner that was not always representative. In practice, most of the excluded academic departments did not have enough students from the Arab sector. This means that the sample is not representative regarding the examinees in Hebrew, but with regard to examinees in Arabic, our population of interest, the sample is indeed representative.

References

Birnbaum, M. H. (1979). Procedures for detection and correction of salary inequity. In T. R. Pezzullo & B. F. Birtingham (Eds.), *Salary equity* (pp. 121-144). Lexington, MA: Lexington Books.

Birnbaum, M. H. (1981). Reply to McLaughlin: Proper path models for theoretical partialling. *American Psychologist*, *36*, 1193-1195.

Bronner, S., Allalouf, A. & Oren C. (1996).

הוגנות מערכת המיון לאוניברסיטאות עבור נבחני מכפ"ל בערבית – השפעת כיול המרכיב המילולי.

[Fairness of the Selection Process to Universities for PET Examinees in Arabic – The Effect of Equating the Verbal Subtest.] Research report no. 223. Jerusalem, Israel: NITE.

Bronner, S. (2004).

השלכות שילוב מכפ"ל או מצרף במערכת המיון להשכלה הגבוהה על תוקף הניבוי וההוגנות של המערכת.

[The Effect of Integrating the PET or the Aggregate Score on Predictive Validity and Fairness in the Higher Education Selection Process.] Research report no. 152. Jerusalem, Israel: NITE.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, *5*, 115-124.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cole, N. S. (1973). Bias in selection. *Journal of Educational Measurement*, *10*, 237-255.

Kennet-Cohen, T., Bronner, S., & Oren, C. (1999). *The Predictive Validity of the Components of the Process of Selection of Candidates for Higher Education in Israel*. Research report no. 264. Jerusalem, Israel: NITE.

Kennet-Cohen, T. (2001)

ניבוי דיפרנציאלי ותוקף דיפרנציאלי של מערכת המיון לאוניברסיטאות על-פי מיצב חברתי-כלכלי של המועמדים.

[Differential Prediction and Differential Validity of the University Admission Procedure according to the Socio-Economic Status of the Candidates.] Research report no. 285. Jerusalem, Israel: NITE.

Kennet-Cohen, T., Cohen Y. & Oren C. (2005).

השוואת הישגים במגזר היהודי ובמגזר הערבי בשלבים שונים של מערכת החינוך – לקט ממצאים. *[A Comparison of Achievements in the Jewish and the Arab Sector at Different Stages in the Educational System – A Survey of Research Findings.]* Research report no. 327. Jerusalem, Israel: NITE.

Linn, R. L. (1982). Ability testing: Individual differences and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences and controversies*, Part II (pp. 335-388). Washington, DC: National Academy Press.

Linn, R. L. (1984). Selection Bias: Multiple meanings. *Journal of Educational Measurement, 21,* 33-47.

Appendix

Differential Prediction according to Academic Year

Table 7: Distribution of academic departments with under- and over-prediction in the criterion, with the composite admission score as predictor, as well as the academic departments without bias, according to faculty and academic year, regarding the examinees in Arabic

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	0*	1	1	0	0	**2**
	16	14	15	17	16	**78**
	4	4	6	5	3	**22**
Social Sciences – Verbal	0	0	0	0	0	**0**
	16	14	14	14	12	**70**
	2	4	5	6	9	**26**
Law	0	0	0	0	0	**0**
	3	3	1	2	1	**10**
	1	0	1	0	2	**4**
Social Sciences – Quantitative	0	0	0	0	0	**0**
	4	3	6	4	6	**23**
	1	2	0	2	0	**5**
Natural Sciences	0	0	0	0	0	**0**
	4	8	10	11	6	**39**
	2	1	0	0	3	**6**
Engineering and Architecture	0	0	0	0	0	**0**
	6	5	4	6	4	**25**
	0	0	0	0	0	**0**
Medicine	0	0	0	0	0	**0**
	4	4	4	5	7	**24**
	0	0	1	0	0	**1**
Nursing	0	0	1	1	0	**2**
	3	2	5	4	6	**20**
	2	4	1	3	7	**17**
Total	**0**	**1**	**2**	**1**	**0**	**4**
	56	**53**	**59**	**63**	**58**	**289**
	12	**15**	**14**	**16**	**24**	**81**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic

Table 8: Distribution of academic departments with under- and over-prediction in the criterion, with the HSM as predictor, as well as the academic departments without bias, according to faculty and academic year, regarding the examinees in Arabic

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	0*	0	0	0	0	**0**
	11	9	8	12	8	**48**
	9	10	14	10	11	**54**
Social Sciences – Verbal	0	0	0	0	0	**0**
	10	5	8	8	6	**37**
	8	13	11	12	15	**59**
Law	0	0	0	0	0	**0**
	4	1	1	0	1	**7**
	0	2	1	2	2	**7**
Social Sciences – Quantitative	0	0	0	0	0	**0**
	3	3	3	5	2	**16**
	2	2	3	1	4	**12**
Natural Sciences	0	0	0	0	0	**0**
	3	5	5	6	4	**23**
	3	4	5	5	5	**22**
Engineering and Architecture	0	0	0	0	0	**0**
	4	4	4	4	3	**19**
	2	1	0	2	1	**6**
Medicine	0	0	0	0	0	**0**
	3	3	4	5	5	**20**
	1	1	1	0	2	**5**
Nursing	0	0	0	0	0	**0**
	3	1	5	2	3	**14**
	2	5	2	6	10	**25**
Total	**0**	**0**	**0**	**0**	**0**	**0**
	41	**31**	**38**	**42**	**32**	**184**
	27	**38**	**37**	**38**	**50**	**190**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic

Table 9: Distribution of academic departments with under- and over-prediction in the criterion, with the PET as predictor, as well as the academic departments without bias, according to faculty and academic year, regarding the examinees in Arabic

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	3*	2	0	0	0	**5**
	17	17	22	22	19	**97**
	0	0	0	0	0	**0**
Social Sciences – Verbal	0	0	0	0	0	**0**
	18	18	19	19	21	**95**
	0	0	0	1	0	**1**
Law	0	0	0	0	0	**0**
	4	3	2	2	3	**14**
	0	0	0	0	0	**0**
Social Sciences – Quantitative	0	0	1	0	0	**1**
	5	5	5	6	6	**27**
	0	0	0	0	0	**0**
Natural Sciences	1	0	2	2	0	**5**
	5	9	8	9	9	**40**
	0	0	0	0	0	**0**
Engineering and Architecture	1	0	0	0	0	**1**
	5	5	4	6	4	**24**
	0	0	0	0	0	**0**
Medicine	1	0	0	0	1	**2**
	3	4	5	5	6	**23**
	0	0	0	0	0	**0**
Nursing	0	0	1	0	0	**1**
	5	6	6	8	13	**38**
	0	0	0	0	0	**0**
Total	**6**	**2**	**4**	**2**	**1**	**15**
	62	**67**	**71**	**77**	**81**	**358**
	0	**0**	**0**	**1**	**0**	**1**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic

Table 10: Distribution of academic departments with under- and over-prediction in the criterion, with the verbal subsection of the PET as predictor, as well as the academic departments without bias, according to faculty and academic year, regarding the examinees in Arabic

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	0*	0	0	0	0	**0**
	20	19	22	22	19	**102**
	0	0	0	0	0	**0**
Social Sciences – Verbal	0	0	0	0	0	**0**
	18	18	19	20	20	**95**
	0	0	0	0	1	**1**
Law	0	0	0	0	0	**0**
	4	3	2	2	3	**14**
	0	0	0	0	0	**0**
Social Sciences – Quantitative	0	0	0	0	0	**0**
	5	5	6	6	6	**28**
	0	0	0	0	0	**0**
Natural Sciences	0	0	1	0	0	**1**
	6	9	9	11	8	**43**
	0	0	0	0	1	**1**
Engineering and Architecture	1	0	0	0	0	**1**
	5	5	4	6	4	**24**
	0	0	0	0	0	**0**
Medicine	0	0	0	1	0	**1**
	4	4	5	4	7	**24**
	0	0	0	0	0	**0**
Nursing	0	0	1	0	0	**1**
	5	6	6	8	13	**38**
	0	0	0	0	0	**0**
Total	**1**	**0**	**2**	**1**	**0**	**4**
	67	**69**	**73**	**79**	**80**	**368**
	0	**0**	**0**	**0**	**2**	**2**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic

Table 11: Distribution of academic departments with under- and over-prediction in the criterion, with the quantitative subsection of the PET as predictor, as well as the academic departments without bias, according to faculty and academic year, regarding the examinees in Arabic

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	0*	1	0	0	0	**1**
	18	15	20	20	19	**92**
	2	3	2	2	0	**9**
Social Sciences – Verbal	0	0	0	1	0	**1**
	17	17	18	17	19	**88**
	1	1	1	2	2	**7**
Law	0	0	0	0	0	**0**
	4	3	2	2	3	**14**
	0	0	0	0	0	**0**
Social Sciences – Quantitative	0	0	0	0	0	**0**
	5	5	6	5	6	**27**
	0	0	0	1	0	**1**
Natural Sciences	0	0	1	1	0	**2**
	6	9	9	10	9	**43**
	0	0	0	0	0	**0**
Engineering and Architecture	0	0	0	0	0	**0**
	6	5	4	6	4	**25**
	0	0	0	0	0	**0**
Medicine	0	0	0	0	0	**0**
	4	4	5	5	7	**25**
	0	0	0	0	0	**0**
Nursing	0	0	0	0	0	**0**
	4	1	7	5	10	**27**
	1	5	0	3	3	**12**
Total	**0**	**1**	**1**	**2**	**0**	**4**
	64	**59**	**71**	**70**	**77**	**341**
	4	**9**	**3**	**8**	**5**	**29**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic

Table 12: Distribution of academic departments with under- and over-prediction in the criterion, with the English subsection of the PET as predictor, as well as the academic departments without bias, according to faculty and academic year, regarding the examinees in Arabic

Faculty	1995/1996	1996/1997	1997/1998	1998/1999	2002/2003	Total
Arts and Humanities	0*	1	0	0	1	**2**
	20	17	22	22	18	**99**
	0	1	0	0	0	**1**
Social Sciences – Verbal	0	0	0	0	0	**0**
	17	18	19	20	21	**95**
	1	0	0	0	0	**1**
Law	0	0	0	0	0	**0**
	4	3	2	2	3	**14**
	0	0	0	0	0	**0**
Social Sciences – Quantitative	0	0	0	0	0	**0**
	5	5	6	6	6	**28**
	0	0	0	0	0	**0**
Natural Sciences	0	0	1	1	1	**3**
	6	9	9	10	8	**42**
	0	0	0	0	0	**0**
Engineering and Architecture	0	1	0	0	1	**2**
	6	4	4	6	3	**23**
	0	0	0	0	0	**0**
Medicine	0	1	1	0	2	**4**
	4	3	4	5	5	**21**
	0	0	0	0	0	**0**
Nursing	1	0	1	1	0	**3**
	4	6	6	7	13	**36**
	0	0	0	0	0	**0**
Total	**1**	**3**	**3**	**2**	**5**	**14**
	66	**65**	**72**	**78**	**77**	**358**
	1	**1**	**0**	**0**	**0**	**2**

*First row: Academic departments with under-prediction for examinees in Arabic
Second row: Academic departments without bias for examinees in Arabic
Third row: Academic departments with over-prediction for examinees in Arabic