מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
المركز القطري للامتحانات والتقييم
מיסודן של האוניברסיטאות בישראל

# The Reliability of the Psychometric Entrance Test and Its Validity in Predicting Academic Success

Tamar Kennet-Cohen
Research Department

January 2016

In many university departments, the number of applicants exceeds the number of available places, necessitating a selection process to choose the most suitable candidates. The quality of tools used for selection is an important issue for academic institutions, candidates, and society as a whole. Two of the basic measures used to evaluate the quality of selection tools are reliability and validity. **Reliability** of a tool examines to what extent repeat measurements of the same examinee using the same tool will yield similar results. **Validity** of a tool examines to what extent it measures what it is supposed to. Academic institutions are interested in accepting candidates with the highest chances of succeeding in their studies, and thus, the scores yielded by the admissions tool are meant to reflect the candidates' chances of success in their studies. Therefore, a positive relationship between the scores yielded by the admissions tool and academic achievements are evidence of the validity of the admissions tool in predicting academic success. This paper provides concise data about the reliability of the Psychometric Entrance Test (PET) and its validity in predicting academic success.

## Reliability

One of the means of measuring reliability is to examine internal consistency by investigating to what extent the questions on the test measure the same variable (or variables). There are different ways of measuring the test's internal consistency. The data presented below were computed using the Kuder-Richardson method, which is identical to the coefficient alpha and is appropriate for tests with items with dichotomous choices such as "true" and "false". The value of the reliability coefficient ranges from 0 to 1.00. A coefficient of 1.00 indicates perfect reliability and a coefficient of 0 indicates an absence of reliability. The accepted rule for high-stakes tests such as the PET is that its reliability should be at least 0.80.

Table 1 presents reliability coefficients for PET scores. The reliability data is based on more than fifty tests administered in the new format, which includes a writing task. The table presents the median values of the reliability coefficients of these tests.

## Table 1: Reliability Coefficients of PET Scores

| Test Domain Scores | | | General Scores* | | |
|---|---|---|---|---|---|
| Verbal Reasoning | Quantitative Reasoning | English | General Multi-Domain | General Humanities-Oriented | General Sciences-Oriented |
| 0.88 | 0.87 | 0.92 | 0.95 | 0.94 | 0.94 |

*For the Multi-Domain Score, the weights of the scores on the Verbal Reasoning and Quantitative Reasoning domains are double that of the score on the English domain. For the Humanities-Oriented Score, the score on the Verbal Reasoning domain has three times the weight of each of the other two scores. For the Sciences-Oriented Score, the score on the Quantitative Reasoning domain has three times the weight of each of the other two scores.

The data shows that the PET is very reliable.

### Predictive Validity

The Israeli universities and the National Institute for Testing and Evaluation (NITE), which acts on their behalf, conduct ongoing studies to examine PET's efficacy in predicting academic success. These studies examine the relationship between PET scores, average score on the high-school matriculation certificate (HSM), and a composite score consisting of the equally weighted HSM and PET, with freshman grade point average (FGPA) and undergraduate grade point average (UGPA).

The most accepted means of measuring the predictive validity of an admissions tool is the Pearson correlation coefficient between achievements in the admissions tool with those in academic studies. It is also known as the validity coefficient of the admissions tool. The value of the correlation coefficient ranges from -1.00 to +1.00. A correlation coefficient of +1.00 indicates a perfect positive relationship, and a correlation coefficient of 0 indicates no relationship. The accepted rule for interpreting correlation coefficients (in absolute values) is that a correlation below 0.10 is considered negligible, a correlation in the 0.10-0.29 range is considered low, the 0.30-0.49 range is considered medium, and 0.50 and above is considered high (Cohen, 1988). It is important to note that high reliability is a condition for high predictive validity.

Estimating the validity coefficients of admissions tools necessitates dealing with the problem of range restriction. We are interested in the predictive validity of the admissions tool for all candidates, but it can only be computed among those candidates who were accepted and began their studies, since data about academic achievement is only available for them. The problem is that the range of scores – both in the admissions tool and in academic achievement – in this group of students is more limited, or restricted, which reduces the observed validity coefficient. To obtain an estimation of the true validity coefficient, one that reflects the validity for all candidates, statistical corrections must be made which rely on observed correlations between the variables in the student sample, and on data on the distribution of those variables that are known for all candidates. The presented data was computed using the correction formula for range restriction, which is appropriate for data of this nature (Gulliksen, 1987, Chap. 12).

Table 2 presents the validity coefficients[1] of the admissions tools in predicting two criteria: FGPA (Oren, Kennet-Cohen, Turvall, & Allalouf, 2014) and UGPA (Kleper, Turvall & Oren, 2014). The findings with respect to FGPA were based on data from 100,863 students from all the universities in Israel from the 2005/6 to 2009/10 academic years. The findings with respect to UGPA were based on data from 100,180 students from all the universities in Israel from the 2001/2 to 2006/7 academic years.

## Table 2: Validity Coefficients of Admissions Tools in Predicting FGPA and UGPA

| Criterion | Multi-Domain Composite | HSM Average | PET Scores | | | |
|---|---|---|---|---|---|---|
| | | | General Multi-Domain | Test Domain Scores | | |
| | | | | Verbal Reasoning | Quantitative Reasoning | English |
| FGPA | 0.46 | 0.36 | 0.43 | 0.36 | 0.37 | 0.32 |
| UGPA | 0.47 | 0.38 | 0.41 | 0.34 | 0.33 | 0.31 |

This data provides evidence of the predictive validity of the PET general score; its advantage, in terms of predictive validity, over the HSM average; and its contribution to the prediction of the criterion over and above what the HSM average predicts. This contribution is reflected in the gap between the predictive validity of the HSM average and that of the composite score. The validity coefficients in predicting FGPA that are routinely computed by NITE are similar to those obtained when predicting UGPA.

1 The correlations presented in this paper were computed within units of analysis (a department in a certain academic institution in a certain academic year). The results presented are weighted averages of these correlations (weighted by the number of students in the unit of analysis).

Table 3 presents the validity coefficients of the admissions tools in predicting FGPA (Oren et al., 2014) by area of study. The table presents the validity coefficients of the three PET general scores (multi-domain general score, humanities-oriented general score, and sciences-oriented general score) and the validity coefficients of the three composite scores, which are computed using the three PET general scores (multi-domain composite score, humanities-oriented composite score, and sciences-oriented composite score).

TABLE 3: VALIDITY COEFFICIENTS OF ADMISSIONS TOOLS IN PREDICTING FGPA BY AREA OF STUDY*

| Areas of Study | Composite Scores | | | HSM Ave. | PET Scores | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | General Scores | | | Test Domain Scores | | |
| | Multi-Domain Comp. | Human.-Oriented Comp. | Sciences-Oriented Comp. | | General Multi-Domain | General Human.-Oriented | General Sciences-Oriented | Verbal Reasoning | Quant. Reasoning | English |
| Law | 0.44 | 0.45 | 0.43 | 0.37 | 0.45 | 0.46 | 0.43 | 0.39 | 0.34 | 0.38 |
| Humanities | 0.48 | 0.49 | 0.47 | 0.38 | 0.43 | 0.43 | 0.41 | 0.39 | 0.34 | 0.34 |
| Social Sciences - Verbal | 0.41 | 0.41 | 0.40 | 0.27 | 0.42 | 0.42 | 0.40 | 0.37 | 0.33 | 0.33 |
| Paramedical Professions | 0.19 | 0.20 | 0.18 | 0.12 | 0.24 | 0.24 | 0.22 | 0.18 | 0.13 | 0.24 |
| Medicine | 0.35 | 0.34 | 0.36 | 0.30 | 0.33 | 0.30 | 0.35 | 0.24 | 0.31 | 0.24 |
| Social Sciences - Quantitative | 0.53 | 0.52 | 0.54 | 0.47 | 0.43 | 0.39 | 0.44 | 0.31 | 0.39 | 0.29 |
| Natural Sciences | 0.57 | 0.55 | 0.58 | 0.45 | 0.52 | 0.48 | 0.54 | 0.42 | 0.50 | 0.38 |
| Engineering | 0.53 | 0.52 | 0.54 | 0.46 | 0.45 | 0.43 | 0.47 | 0.37 | 0.42 | 0.31 |
| All Departments | 0.46 | 0.46 | 0.46 | 0.36 | 0.43 | 0.41 | 0.43 | 0.36 | 0.37 | 0.32 |

*The shaded areas: The highest validity among the three composite scores and the three general scores on the PET.

The table shows consistent evidence regarding the contribution of the PET general score to the prediction of the criterion over and above what the HSM average predicts. In every area of study, the validity of the composite score is higher than that of the HSM average. In addition, the figures show that there is a slight advantage for the humanities-oriented score in predicting the criterion in verbal areas of study, whereas there is an advantage for the sciences-oriented score in predicting the criterion in quantitative areas of study.

To conclude, the data presented in this study is evidence of high reliability and validity (relative to other admissions systems in Israel and elsewhere) and support the use of the PET in the selection process for undergraduate studies in universities in Israel. More detailed information about the findings reported here may be found in the references listed below.

**SOURCES:**


קלפר, ד', טורוול, א' ואורן, כ' (2014). **תוקף הניבוי של כלי המיון לאוניברסיטאות בישראל מול ממוצע תואר בוגר** (דוח מס' 403). ירושלים: מרכז ארצי לבחינות ולהערכה.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Erlbaum. (Original work published 1950.)

Oren, C., Kennet-Cohen, T., Turvall, E., & Allalouf, A. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. *Psicothema*, 26, 117-126.