

# **The Predictive Validity of the Components of the Process of Selection for Higher Education in Israel: A Correction for Sample-Selection Bias Using Heckman's Method**

Tamar Kennet-Cohen and Shmuel Bronner  
The National Institute of Evaluation (NITE)

## **Abstract**

Studies of predictive validity are usually conducted on selected samples. In such circumstances, the question arises as to whether the sample statistics (correlation coefficients and regression coefficients) estimate correctly the population parameters prior to selection. The assumption prevalent in the professional psychometric literature is that regression coefficients do not change as a result of the selection process. Correction of the validity coefficients for the effect of the selection ("correction for restriction of range") is based on this assumption. This report explores the assumption that the regression coefficients remain unchanged by the selection process.

First, evidence is presented which demonstrates that this assumption does in fact hold in the context of selection processes for institutions of higher education. Studies carried out in the United States and in Israel provide empirical proof that correlation coefficients which were obtained in selected samples and corrected for restriction of range under the assumption that the regression equations do not change as a result of the selection, accurately reflect correlations calculated for an unselected population.

Second, we describe the set of circumstances in which the selection process may lead to obtaining biased estimates of the regression parameters. Analysis of a model describing the effect of selection (Heckman, 1979) supports the claim that the process characterizing selection for institutions of higher education tends, if anything, to lead to a downward bias in the estimation of the regression coefficients. This assumption is supported by the work of researchers familiar with the variables typically involved in selection processes for higher education and the relationships among these variables.

Given the a priori assumption that the selection process leads to a downward bias in the regression coefficients (and, therefore, also to an underestimate of the correlation coefficients, obtained by the restriction of range correction formulas), it is desirable to have a statistical model which permits empirical investigation of the existence and direction of bias resulting from selection, and suggests a method for correcting it. Such a model, proposed by Heckman (1979), was applied by Ayalon and Yogev (1997) in an investigation of the predictive validity of the components of the selection process for Tel Aviv University.

The main argument of the current report is that Ayalon and Yogev have applied Heckman's model erroneously, which led to their obtaining incorrect estimates for the parameters which they wished to estimate. This study presents and supports this argument through a theoretical explication of Heckman's model. The argument is further supported by an analysis of simulated data, the results of which are reported herein. These results testify to a systematic departure of the estimates produced by Ayalon and Yogev's application of Heckman's model from the true population parameters. The study shows that if Heckman's model is applied correctly, then analyses of the simulated data provide accurate estimates of the population parameters.

Both applications of Heckman's model were used in a study of the predictive validity of the major components of the selection process (the total score on the Psychometric Entrance Test - PET, and the average of grades on the high school matriculation certificate - HSM) for applicants to Tel Aviv University for the academic year 1994-95. When the correct application was used, both predictors were found (on average, across all departments) to have positive predictive validity and a positive marginal contribution in predicting first-year grade average.