

Effects of Coaching for PET, in Hebrew & in Arabic

Ayelet Goldzweig , Ayeletg@nite.org.il

Carmel Oren, Carmel@nite.org.il

Yonatan Saar , Yonatan@nite.org.il

NATIONAL INSTITUTE FOR TESTING AND EVALUATION (NITE)

The goal of the study was to estimate the added value (in terms of scores) of preparatory courses to PET scores, over self-studying for the PET, in Hebrew and in Arabic, while statistically controlling for intervening background factors and particularly for basic ability. In the first part of the study, we evaluated the added value of PET scores using a linear regression model predicting PET score from preparation method and background variables, among first time examinees. Due to the limited success of this model in eliminating self-selection effects, in the next stage we used another method, based on repeat examinees. The latter provided better control for self-selection bias and improved estimation precision. Results show that preparatory courses have small positive effects that differ among language groups. In Hebrew, preparatory course contributed about 5 points to the first sitting and about 26 points to the second sitting, on average. However, in Arabic there was no difference in the contribution of a course to any of the sittings, and it was estimated at about 12 points, on average, in both cases. Despite the caution needed in inferring causation, the findings of this study support and extend the knowledge in this field.

Keywords: Higher education admission, Coaching, PET, estimation precision

Table of Contents

Introduction.....	2
Method	3
First part – first-time examinees.....	3
Second part – repeat examinees	3
Results.....	4
First part – first-time examinees.....	4
Second part – repeat examinees	4
Summary and Discussion.....	5
References.....	5

Introduction

Over the last few years, a majority of Psychometric Entrance Test (PET) examinees in Israel have participated in a commercial coaching course in preparation for the test. This raises the question of the efficacy of such preparation courses, compared to self-study using other means (e.g., preparation books, computer programs, etc.). An additional question is whether there is a difference between Hebrew- and Arabic-speaking examinees with regard to the preparation methods they use and the benefits they derive from them.

Research on special preparation for scholastic aptitude tests has been conducted mainly in the USA and has typically focused on the effects of coaching on the SAT scores. The general consensus is that, on average, coaching increases SAT scores by a total of about 30 points (out of the maximum total of 1600): 10-20 points on the Mathematics section and 5-10 points on the Critical Reading section (Briggs, 2001, 2009; Buchmann, Condrón, & Roscigno, 2010; College Board, 1999; Powers, 2012; Powers & Rock, 1998). Although slightly higher estimates have been reported recently (Montgomery & Lilly, 2012), it is clear that the benefit of coaching is well below the hundreds of additional points guaranteed by commercial coaching institutes.

Research in Israel likewise suggests that the efficacy of coaching courses as a means of preparation for the PET is much lower than claimed by commercial coaching institutes. The PET scores are scaled from 200 to 800 with a mean of 500 and a standard deviation of 100. Allalouf & Ben-Shakhar (1998) conducted an experimental research study, in which students from pre-academic preparatory programs were randomly assigned to either an experimental or control group. The experimental group participated in a coaching course for the Verbal Reasoning and Quantitative Reasoning domains of the PET, while the control group did not. Both groups were tested twice and a comparison of their mean improvement from test to test revealed that on average, the coached group had gained 24 points more than the un-coached group. In addition, gains in the Quantitative domain were greater than in the Verbal domain.

Other researchers (Oren, 1993; Schatzman & Oren, 1997) have conducted observational studies to assess the coaching courses' efficacy, by comparing repeat examinees who used different preparation methods for their tests (coaching course vs. self-study). The distinct efficacy of a preparatory course over self-study before re-testing, was estimated in one of these studies as a 7-point gain (Oren 1993). Further research estimated this efficacy as a 5-9 point gain for the first test and as a 7-11-point gain for the second test (Schatzman & Oren, 1997). Since examinees in these studies were not randomly assigned to different groups, statistical procedures were applied to control for intervening variables, but they had little effect on the outcomes.

One of the main problems of non-experimental study designs is the self-selection of examinees into groups, which may bias the sample. In the context of test preparation, several distinct factors may influence self-selection: First, the test takers' basic inherent ability may affect their choice of preparation method, so that, for example,

stronger examinees may choose not to attend a coaching course. Secondly, a coaching course may not be an affordable option for examinees of lower economic status. Additionally, motivational as well as other factors may lead to the selection of one method of preparation over another. The current study consists of two parts: The first part addresses the two specific factors of self-selection, basic ability and economic status, and discusses how we attempt to statistically control for them. The second part then focuses more closely on the question of basic ability and on the careful endeavor to exclude the impact it might have. The motivational factor and additional factors pertaining to self-selection are not addressed.

The principal goal of the current study is to estimate the added value (in terms of PET scores) of coaching courses over self-study for the test, in the Hebrew and Arabic versions, respectively, while statistically controlling for intervening background factors, and particularly for basic ability.

Method

The background factors used in this study are: gender, economic status, parental education and weighted average matriculation score as an indicator for basic ability. The sampling is based on the feedback questionnaire ("MASHOV") given at the end of the PET exam and the data is analyzed separately for each language.

First part – first-time examinees

Initially, a linear regression model was constructed that predicted PET scores on the basis of preparation method and background variables among 52,327 first-time examinees. This model allowed us to estimate the exclusive contribution of a coaching course to PET scores, apart from background variables.

Given the limited success of this model in eliminating self-selection effects, and assuming that the basic ability of test takers plays a major role in their self-selection with regard to different preparation methods, in the next stage of the study we applied a different paradigm, one which allows a better statistical control of the examinees' ability.

Second part – repeat examinees

We identified 15,828 examinees who sat the test two or more times during the study period (2000-2010). The research only took into account the first two test-sittings. Test takers were then divided into four groups according to the method of preparation they undertook prior to each of the two test-sittings (self-self, self-course, course-self, course-course).

First, a regression model predicting the mean improvement between sittings was built to control for the effects of background variables, much as in the first part of the study. In keeping with the model, adjusted average test scores and the mean difference between them were calculated for each group.

Subsequently, a series of paired comparisons between groups was carried out, which made it possible, under certain assumptions, to estimate the specific contribution of the coaching course to the PET score as distinct from other factors.

Each of these comparisons included two groups, who had prepared in a similar way for one of their PET sittings, but in a different way for the second. Since grouping was identical in both test-sittings, we assumed that group differences, comprised of differences in basic ability and in factors not accounted for by the model, also remained constant in the two sittings. In this constellation, a difference between group scores in the first sitting should be followed by an identical, corresponding, score difference in the second sitting, and vice versa. Therefore, a deviation from the expected score difference from one sitting to the other could be attributed to the difference in preparation method before the relevant sitting.

For example, to estimate the specific efficacy of the coaching course prior to the second sitting, we compared the "self-self" group to the "self-course" group. The two groups had prepared in a similar way for their first sitting, but differed in the preparation method they undertook before their second one. Therefore, the difference between the group mean scores on the first sitting cannot be attributed to the preparation method prior to the first sitting. Since all other factors, including basic ability, remained constant for both sittings, except for preparation method prior to the second sitting, when the mean score difference between groups on the second sitting differed from the mean score difference between them on the first sitting, we interpreted this deviation as an estimate for the specific effect of the coaching course used prior to the second sitting. Following the same logic, additional comparisons were conducted.

Results

First part – first-time examinees

The analysis revealed that among Hebrew-speaking examinees, course participants scored an average of 4.7 points less than those who had prepared on their own, whereas among Arabic-speaking examinees, course participants scored an average of 9.6 points more than those who had prepared on their own.

In both cases, partial and semi-partial correlations between the type of preparation and test scores were found to be negligible.

Another analysis revealed that the background variables explained only 1.5% of the variance in the type of preparation.

Second part – repeat examinees

Overall, we found that in the Hebrew PET, coaching courses added, on average, about 5 points to the first sitting and about 26 points to the second sitting. However, in the Arabic version there was no difference in the distinctive contribution of a preparatory course to scores in either of the sittings; and it was estimated at about 12 points, on average, for both test-sittings.

Summary and Discussion

In the first part of the study we tried to estimate the contribution of coaching courses using a simple linear regression model, but the success of this method in controlling for the self-selection bias was limited. According to the originally observed mean scores and the findings of the first regression model for the Hebrew test, coaching course participants scored lower than those who studied on their own. The most plausible explanation for this finding is non-random self-selection into groups.

It seems that the repeat examinees model we applied in the second stage of the study was helpful in addressing this problem, at least to some extent. Once we applied the repeat examinees model, the efficacy of coaching courses was estimated as having a rather small but nonetheless positive effect. In other words, we managed to obtain an estimate that gives a reasonable indication of effect with regard to its direction, and is probably more accurate. Apparently, the repeat examinees model enabled us to reduce in part the problematic impact of self-selection on the estimation, and achieve error reduction and improved estimation of the distinct benefit of coaching courses.

The positive, albeit small, impact of coaching courses on PET scores identified by this study is in keeping with past research findings. Interestingly, a difference between Hebrew and Arabic- test examinees has been observed in this context. This might imply that the self-selection itself differs between the two groups, especially in light of the known differences between these two populations.

It should be noted that the current research has certain limitations. The repeat examinee model assumes that differences between groups remain constant for both sittings, that the method of preparation is independent from the examinees' ability to gain from it, and that there are no complex interactions between benefits from different combinations of preparation techniques. Moreover, the repeat examinees sample itself, as such, is also presumably non-randomly self-selected. Furthermore, since the entire study is observational, it is difficult to draw conclusions about causation.

However, despite the methodological limitations which dictate caution in inferring causation, the substantial findings of this study support and extend the knowledge in this field.

References

- Allalouf, A. & Ben-Shakhar, G. (1998) The Effect of Coaching on the Predictive Validity of Scholastic Aptitude Test. *Journal of Educational Measurement*, 35(1), 31-47.
DOI: 10.1111/j.1745-3984.1998.tb00526.x
- Briggs, D. C. (2001). The Effect of Admissions Test Preparation: Evidence from NELS: 88. *Chance* 14 (1), 10-22.
- Briggs, D.C. (2009). *Preparation for College Admission Exams*. NACAC Discussion Paper. Arlington, VA: National Association for College Admission Counseling.

- Buchmann, C., Condron, D. J. & Roscigno, V. J. (2010). Shadow Education, American Style: Test Preparation, the SAT and College Enrollment. *Social Forces* 89 (2), 435-461. doi:10.1353/sof.2010.0105
- Byun, S. & Park, H. (2012). The Academic Success of East Asian American Youth: The Role of Shadow Education. *Sociology of Education* 85: 40-60.
doi:10.1177/0038040711417009
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- College Board. (1999). *Coaching and the SAT I*. Research Note 06, Educational Testing Service. New York, NY: College Board.
- Domingue, B. W. & Briggs, D. C. (2009) Using linear regression and propensity score matching to estimate the effect of coaching on the SAT. *Multiple Linear Regression Viewpoints*, 35 (1), 12-29.
- Hansen, B. B. (2004). Full Matching in an Observational Study of Coaching for the SAT. *Journal of the American Statistical Association*, 99 (467): 609-618.
doi: 10.1198/016214504000000647
- Montgomery, P. & Lilly, J. (2012). Systematic reviews of the effects of preparatory courses on university entrance examinations in high school-age students. *International Journal of Social Welfare*, 21, 3–12. doi: 10.1111/j.1468-2397.2011.00812.x
- Oren, C. (1993). *On the utility obtained from different methods of preparation for the PET* (In Hebrew). Report No. 170, National Institute for Testing and Evaluation, Jerusalem.
- Powers, D.E. (2012). Understanding the Impact of Special Preparation for Admissions Tests. ETS Research Report No. RR-12-05, ETS R&D Scientific and Policy Contributions Series No. SPC-12-02.
- Powers, D.E. & Rock D. A. (1998). *Effects of Coaching on SAT I: Reasoning Scores*. College Board Research Report 98-6. New York, NY: College Board.
- Schatzman, K. & Oren, C. (1997). *The usefulness of psychometric exam preparation courses Compared to self-preparation* (In Hebrew). Report No. 232, National Institute for Testing and Evaluation, Jerusalem.