

BAR-ILAN UNIVERSITY

**Measuring educational effectiveness in Israel's
elementary schools**

Reut Tsidkany

Submitted in partial fulfillment of the requirements for the Master's
Degree in the School of Education, Bar-Ilan University

Ramat-Gan, Israel

2021

Abstract

This research is comprised of two studies both examine educational effectiveness, as expressed in improvement in academic achievement of students enrolled in elementary schools in Israel. While there are many studies that measure academic achievement, this work is unique in that it examines educational effectiveness by measuring the improvement gained in academic achievement along time.

The first study analyzes school effectiveness, focusing on school structure (comparing eight-year or six-year elementary schools). The decision regarding the structure of each school is exogenous to the school, being determined at the local authority level and/or the Ministry of Education, where school principals cannot influence that decision.

The second study addresses intra-school processes, with a focus on internal intervention programs. Intervention programs can be implemented at the managerial prerogative of each individual school, where the required resources are available to the school board. This study examines the effectiveness of a specific program that incentivizes students to improve their achievement in mathematics, using “awards”.

Two educational structures are common in Israeli elementary schools. In the traditional structure, students attend elementary school until eighth grade, then move on to high school (hereinafter - eight-year schools). About 20% of Israeli students currently study in schools with this structure (Dvir, 2018a). In the prevailing structure, students attend elementary school until sixth grade, after which they move on to middle school and from there to high school (hereinafter – split-structure schools).

This study (Part 1) had two objectives: The first was to examine the extent of educational effectiveness, as expressed in improvements in the student academic achievements, of eight-year schools compared with split-structure schools. The

hypothesis was that eight-year schools are more effective than split-structure schools, given the former's students experience fewer transitions than do their peers who attend the latter (Opletka and Tobin, 2008; Yinon and Rudnicki, 2003). The second objective was to identify factors (previous achievements and background variables; societal sector, the Ministry of Education's needs-based funding formula¹, the school's location/ peripherality index, class size, and annual governmental financial investment per student) that contribute to explaining the educational effectiveness of schools. The hypothesis was that schools in the Hebrew-speaking sector, where students from established backgrounds, located in the center of the country, in small classes, and who are entitled to a high average allocation of funds per student, will be more effective.

We sampled 120 schools for this study, of them 40 eight-year elementary schools, 40 six-year elementary schools, and 40 middle schools. To represent each school's level of achievements, we used the average scores on the Meitzav standardized tests (school efficiency indices, hereinafter - Meitzav) in mathematics, English, and science over the years 2012-2016 (extracted from the Ministry of Education's website 'Almost Everything About Educational Institutions'). In addition, for each school, we ascertained their characteristics: ethnicity, needs-based funding formula, class size, and annual governmental financial investment per student, using the Ministry of Education's "Education Transparency" website, and the peripherality index through Israel's Central Bureau of Statistics website.

The research method is based on "value-added" model regressions, in line with previous studies that examined educational effectiveness (for details see Gilboa, 2010; Canaan, et al., 2019; Chetty, et al., 2011; Darling-Hammond, et al., 2011; Jackson, et

¹ This formula considers the parental education; family income; school location; migration status.

al., 2020). The basic “value-added” model requires a measurement of at least two scores over time, for the same subject; to calculate the difference between the two scores, previous achievements and the effects of the background variables are deducted, to allow comparison of subjects with different achievements and backgrounds. Over time, the basic model evolved into a Contextual Value Added (CVA) model (Leckie & Goldstein, 2017; Ray, 2009) in which the characteristics and background variables of the subjects were added to the model. The CVA model contributes to the accuracy of identification of the impact of the independent variable (e.g., the school) over the study period, by adding its background variables into the analysis. Thus, in this study, we measured the improvement in achievement using the schools’ grades in the Meitzav tests that were conducted in the fifth and eighth grades respectively, over the years 2012-2016, and added the abovementioned characteristics of the schools as background variables.

The study findings indicate that the eight-year schools and split-structure schools are equally as effective. That is, in terms of improvement in achievement, the eight-year school structure is not preferred over the split structure. Among the variables examined, it was found that each school’s students’ previous achievements and socio-economic indicators contribute to explaining the overall effectiveness of schools. The most influential factor was previous achievements: the higher the school scored in the first period, the higher it scored in the second period. In addition, the lower the socio-economic status, the lower the eighth-grade achievement in English and mathematics. That is, the lower the status of the school’s students (the higher the school's needs-based funding formula), the lower the eighth grade Meitzav score. The other variables (ethnicity, peripherality, class size, and annual governmental financial investment per student) were not found to have a statistically significant effect on achievement.

Part 2 of the work focuses on a school-based intervention program; a unique initiative, implemented by the school, that gave “awards” to students for improving their achievement in mathematics. This program was run during the 2017-2018 academic year, in an elementary school located in the social periphery (Cluster 4 according to the socio-economic index of the Central Bureau of Statistics (CBS, 2021)), whose students’ achievements are relatively low (e.g., a score of 55 out of 100 in the Fifth Grade Meitzav Mathematics test in 2016 (2016), decile three according to the Ministry of Education, National Authority for Measurement and Evaluation in Education, 2016). The purpose of the “awards” program, like other incentive programs, was to improve students’ achievement and raise their motivation for learning. The program gave “awards” with token value (e.g., stationary items with real value approximately \$1) to students who improved their academic achievement in mathematics.

The objectives of this study were twofold. The first was to examine the educational effectiveness of the “awards” program. The hypothesis was that, in line with previous studies that have examined the educational effectiveness of student incentives (Angrist & Lavy, 2009; Bash & Fischer, 2020; Bettinger, 2012; Dulleck, et al., 2016; Fryer, 2011; Jalava, et al., 2015; Le, 2020; Levitt, et al., 2016a; Levitt, et al., 2016b), the incentives would be effective, meaning that the degree of improvement in students’ achievement in mathematics would be higher for those taking part in the “awards” program, than for those in the control group, who studied in a traditional educational setting.

The second objective was to examine, for students who participated in the “awards” program, whether those who “won” more “awards” improved their achievements to a greater degree, and to identify the background variables of students (e.g., gender, grade, number of siblings, marital status, and migration status) that predict improved

achievement in the “awards” program. It was hypothesized that among students who participated in the “awards” program, those who won a larger number of “awards” would improve their test scores in the second period (at the end of the program) to a greater degree than their counterparts who did not win an “award”. In addition, it was expected that boys in fifth grade, with few siblings, whose parents are in a relationship, and whose family are immigrants, would improve their achievements to a greater degree than would their peers.

The study involved 426 students – 94 students in the treatment group (participated in the “awards” program) and 332 students in the control group (studied in the traditional framework). For each student, two grades were taken from math exams: a grade from the beginning of the school year (the first period) and a grade from the end of the year (the second period). The raw exam scores from the second period, for both groups (treatment and control), were converted to a uniform scale and multiplied by a factor of 1.3 to allow comparison of tests with different levels of difficulty (for more information, see Böhlmark & Lindahl, 2015; Kyriakides, et al., 2019; Mok, et al., 2015). Each student in the treatment group was assigned a personal target score, calculated relative to their starting point – the initial test score (first period). The target was defined so that it would not be easy for the student to achieve, but rather would require them to invest in their learning (Chiang, et. Al., 2017). Over the subject year, six short quizzes were held. Students who achieved their target scores received a token “award” (a game or stationery valued at approximately \$1). At the end of the program, after the second period exam, students who achieved their personal target were given an additional “award”; a certificate of appreciation from the school principal.

The quantitative research methodology for this study was based on regressions using the “Difference in Differences” (DID) technique (Angrist & Pischke, 2008; Bach &

Fischer, 2020; Battistin & Meroni, 2016; Dulleck, et al., 2016; Gertler, et al. 2016; Strello, et al., 2021; Torrats-Espinosa, 2020). This method is common in studies involving “natural experiments” and deals with examining the causal relationships created after events such as a sharp economic change, the introduction of new legislation, or occurrence of a unique phenomenon (the “treatment”). It requires identification of two groups with similar characteristics, where the “treatment” being examined affected only one of the two. Calculating the differences between the two groups, across two time periods, produces the “Difference in Differences” – the net effect of the “treatment” (Angrist & Pischke, 2008; Gertler, et al. 2016). The DID method is required in this study because the implementation of the “awards” program included all the students in the school selected for the study (including both treatment and control groups), and it is not possible to predict the degree of improvement that would have taken place in the absence of the “rewards” program.

For the first research question, which addressed the educational effectiveness of the “awards” program, the “difference in differences” was calculated as the difference between two differences, as follows: The first difference that was calculated assessed the effect of time and was defined as the difference between the average test scores from the first period and the average test scores from the second period, for each of the two groups (before / after). Comparing grades over time allows assessment of the overall improvement or regression in student grades. The second difference that was calculated assessed the effect of the "treatment" and was defined as the difference between the average of the achievements of each of the two groups (treatment / control). Calculating the difference between the two differences produced the "difference in differences", representing the net effect of the "awards" program.

For the second research question, which examines the effect of the number of “awards” on the improvement in achievement, a logistical regression was performed to test the strengths of the relationships between the number of “awards” the students “won” during the program and their improvement in achievement, while controlling for each student's background variables (gender, grade level, number of siblings, marital status, immigrants).

The findings of DID analysis indicate that in the treatment group (which participated in the “awards” program) the improvement in achievement in mathematics along time was greater than 80% of a full score. That is, an improvement of eight points (out of 100) compared to the improvement in the control group. Findings of the regression analysis according to the DID model teach that there is a significant positive relationship between students’ participation in the “awards” program and their improved achievement. The findings of the logistic regression analysis indicate that the chance that a student who “won” an “award” would improve their achievement was 1.5 times higher than the changes of their fellow students who did not “win” “awards” throughout the year to improve.

Regarding the students' background variables, it was found that the grade level (fifth grade / sixth grade) and the mother tongue (representing migration) are statistically significant predictors of their improvement in achievement. A sixth-grader has a 70 percent lower chance of improving their performance than does a fifth-grader. The main difference between the two grade levels relates to student assessment tools. In the fifth grade, the assessment tool is a Meitzav test from previous years, while in the sixth grade, the tests were conducted by the teaching staff. That is, the reliability of the data increases as the measurement tool becomes more professional. In addition, it was found that the chance of students whose mother tongue is not Hebrew, and whose family has

a background of migration, to improve their achievements, was 3.5 times higher than that of their Hebrew-speaking counterparts. The other background variables were not found to be statistically significant (gender, number of siblings, parents' relationship status).

This work contributes to the search for the factors that will lead to the improvement of educational effectiveness of primary schools, as expressed in the improvement of the achievements of their students. Regarding the educational structure of the schools, it was found that the six-year and split-structure schools were equally as effective. Of the other systemic factors, past achievements of a school's student cohort and the socio-economic status of its students were the only predictors of achievement. From the perspective of internal programs in individual schools, the "awards" program was found to be effective. Programs of this nature are under the control of the school's teaching staff and their significant advantage is that with low investment (in terms of financial and human resources), schools can see significant improvements in achievement. In addition, unlike incentive programs that provide financial rewards that are controversial from an educational perspective, this program provided token rewards that are given to school students as a matter of course. Thus, the research results suggest that other schools should use similar "awards" programs, adapting them to their respective needs.

There is an opportunity for future research that builds upon the methodology used in the study on the systemic aspects of educational effectiveness by further considering the educational structure of schools, but with the student as the subject – examining individual students' progress between two points in time, in schools with different structures. In addition, there is a need for research that examines schools' educational effectiveness through emotional and social prisms, utilizing Meitzav tests that evaluate the schools' social climates, as well as to examine educational effectiveness according

to other systemic characteristics (e.g., school size). Finally, from the systemic perspective, the differences between the characteristics of the schools could be further elucidated, and the sample size increased, in order to identify additional characteristics that affect educational effectiveness. Further studies on intra-school programs are needed in order to understand the effectiveness of “awards” programs in additional subjects (beyond mathematics). In addition, it would be beneficial to examine the effectiveness of such programs on students of a variety of ages, as well as changes in motivation among participating students, during and after “awards” programs.