

The Initiative for Applied Education Research  
The Israel Academy of Sciences and Humanities

Symposium

**Using Longitudinal Data as a Source of Information  
for Education research, Policy and Programs in Israel**

---

**Symposium Background Materials**

---

December 19, 2016  
The Van Leer Jerusalem Institute,

Unrevised Translation from the Hebrew Original

Academic coordinator: Oded Busharian

**The Israel Academy of Sciences and Humanities** was established by law in 1961. The Israel Academy of Sciences and Humanities Law declares that the Academy's principal objectives and tasks are to: assemble, under its umbrella, the finest Israeli scientists; foster and promote scientific activity; advise the government on research activities and scientific planning of national importance; maintain ties with similar institutions abroad; represent Israeli science at international institutes and conferences; and, publish articles that can further science.

The Israel Academy of Sciences and Humanities views the development and advancement of basic scientific research as a goal of national importance. The Academy's aim is to cultivate and promote basic research and scientific activity in Israel. So as to foster development of an infrastructure for successful basic scientific research, the Academy sets high standards in the humanities, the natural sciences and technology, encourages excellence in the field of scientific creativity and thinking, and initiates activities that will articulate the achievements of science in Israel.

**The Initiative for Applied Education Research (the Initiative)** places up-to-date, scientific, critically-appraised knowledge and information at the disposal of decision-makers in the field of education. Such information is crucial for the intelligent formulation of policy and optimal planning of interventions to improve achievements in the sphere of education in Israel.

The Initiative's vision: Research knowledge is an essential component in planning public policy or large-scale interventions. In the planning stages, critically-appraised research knowledge supports the formulation of policy that has a better chance of success and later, it is more apt to promote rational public discourse. The Initiative puts this vision into practice in the field of education.

The Initiative's *modi operandi*: The Initiative addresses issues raised by decision-makers and consults with senior Ministry of Education officials and other stakeholders. The Initiative's steering committee, appointed by the president of the Israel Academy of Sciences and Humanities, is charged with creating a work plan and is responsible for implementing the peer-review processes of the reports it produces.

The Initiative operates through expert committees and by convening joint learning sessions held for researchers, education professionals and decision-makers. It publishes reports of its work and makes them readily available to the public. Those serving on its expert committees do so without remuneration.

History of the Initiative: The Initiative was established in late 2003 as a joint venture of the Israel Academy of Sciences and Humanities, the Ministry of Education, and Yad Hanadiv (the Rothschild Foundation). The Initiative has been operating as a unit of the Israel Academy since the beginning of 2010.

In the summer of 2010, the Israeli Knesset amended the Israel Academy of Sciences and Humanities Law, regulating the Academy's advisory role vis-à-vis government ministries seeking its consulting services. The Initiative directs the consulting activities on education-related issues which the Israel Academy provides to the government and various authorities.

Expert team members:

Prof. Leah Achdut (chair) – Van Leer institute

Prof. Michal Beller - Levinsky College of Education

Dr. Iris Tabak - Ben Gurion University of the Negev

Prof. Orly Manor - The Hebrew University of Jerusalem

Prof. Tal Zarsky - Haifa University

Mr. Noam Zussman – Bank of Israel

## Contents

Agenda .....	4
Introduction .....	<b>שגיאה! הסימניה אינה מוגדרת.</b>
Lectures abstracts .....	11
Bibliography .....	21
Short bios.....	22

## Agenda

<b>08:30-09:00</b>	Gathering
<b>09:00-09:15</b>	<b>Opening: Professor Leah Achdut</b> , Chairperson, Expert Team
<b>09:15-11:30</b>	<b>First session: The Potential and Importance of Longitudinal Research in the Field of Education</b> <b>Chair: Professor Leah Achdut</b> , Chairperson, Expert Team
<b>09:15-09:55</b>	Opportunities and Challenges in Gathering and Analyzing Longitudinal Data in Education (lecture in English; simultaneous translation will be possible) <b>Professor David Kaplan</b> , University of Wisconsin-Madison
<b>09:55-10:35</b>	The Potential of Longitudinal Studies in the Field of Education in Israel <b>Professor Moshe Justman</b> , Ben-Gurion University of the Negev
<b>10:35-11:15</b>	Long-term Effects of Educational Interventions in Primary and Secondary Schools in Israel <b>Professor Victor Lavy</b> , Hebrew University of Jerusalem
<b>11:15-11:40</b>	Recess – Coffee and Cake
<b>11:40-13:15</b>	<b>Second session: Use of Administrative Data for Research Purposes</b> <b>Chair: Dr. Iris Tabak</b> , Member, Expert Team
<b>11:40-12:10</b>	Administrative Data in the Field of Education in Israel: The Review Process and its Outcome <b>Dr. Amalia Ran</b> , MOFET Institute
<b>12:10-12:40</b>	Actions of the Ministry of Health as an Example of Making Administrative Data Available to Researchers in Israel <b>Ms. Aline Attias</b> , Ministry of Health
<b>12:40-12:50</b>	Summing up: The Gap between Reality and the Ideal, and the Possible <b>Mr. Noam Zussman</b> , Member, Expert Team
<b>12:50-13:20</b>	Audience reactions and questions <b>Mr. Yoel Finkel</b> , Central Bureau of Statistics
<b>13:20-14:00</b>	Lunch
<b>14:00-15:45</b>	<b>Third Session: The Balance between Protection of Privacy and the Public Good that May Come from Longitudinal Studies</b> <b>Chair: Professor Tal Zarsky</b> , Member, Expert Team
<b>14:00-14:30</b>	The Public Good versus the Rights of the Individual: An Ethical-Moral View <b>Dr. Yuval Eylon</b> , the Open university
<b>14:30-15:00</b>	The Benefits of Education Research versus the Risk of Privacy Infringement as Reflected in the Concept of Proportionality <b>Professor Raanan Sulitzeanu-Kenan</b> , Israel Democracy Institute
<b>15:00-15:30</b>	The Israeli Regulator: Working between Probability and Risk <b>Adv. Limor Schmerling</b> , Information and Technology Authority
<b>15:30-15:45</b>	Privacy Protection as a Question of Risk Management – Response to Limor Schmerling’s Remarks <b>Professor Tal Zarsky</b> , Member, Expert Team
<b>16:00-15:45</b>	Summary and Insights <b>Professor Michal Beller</b> , Member, Expert Team

## Introduction

The education system in Israel, as in other countries, is a key provider of public services, a pillar in the process of building the population's human capital, and has a decisive influence on the wellbeing of individuals and on economic growth. Education systems face complex challenges – including advancing equality of opportunity and providing high quality services on an increasingly growing scale – all this with persistently limited resources.

To evaluate the education system's activities and continuously improve them, comprehensive, high-quality databases are required. In Israel, there has been a growing awareness in recent years of the advantages inherent in longitudinal data as a basis for policymaking, evaluation of intervention programs, and high quality research concerning various aspects of the education system. In parallel, the demand for longitudinal data has increased and their use has broadened alongside the traditional use of cross-sectional data. All this with the understanding that acquiring an education is a multi-dimensional and ongoing process and that as a process, it is affected by decisions made at different points in time by educators, parents, students, and others. Tracking these details over time may deepen our understanding of the processes related to acquiring an education and to additional opportunities for education.

The Initiative for Applied Education Research assembled an expert committee with the support and encouragement of Yad Hanadiv, which will lead a learning process on the topic of “The Use of Data Measured over Time as a Source of Information for Education Policy and Programs.” The members of the committee are: Professor Lea Achdut – Ruppin Academic Center and the Van Leer Institute in Jerusalem (chairperson); Professor Michal Beller – Levinsky College of Education; Professor Orly Manor – Hebrew University of Jerusalem and Hadassah Hospital; Professor Iris Tabak – Ben-Gurion University of the Negev; Professor Tal Zarsky – University of Haifa; Mr. Noam Zussman – Bank of Israel. Mr. Oded Busharian of the Initiative is the team coordinator.

The team had three main objectives: a) Examining methods to create longitudinal data infrastructures in the field of education and related areas, making the data available and expanding their use for research and policymaking purposes; b) Promoting cooperation between entities that possess longitudinal data with the goal of creating integrative databases; c) Examining technological and legal-ethical aspects involved in creating longitudinal infrastructures and making them available to broad audiences while placing emphasis on privacy protection.

The team held three meetings with government ministry representatives (mainly the Ministry of Education and the Central Bureau of Statistics), academicians and other experts who shared their experience and knowledge with the team members (the agenda of the meetings appears in Appendix B, the participants' brief bio sketches appear in Appendix C). Two meetings were devoted to the data infrastructures belonging to the Ministry of Education (MOE) and the Central Bureau of Statistics (CBS), during which the existing databases were reviewed, the manner in which they are organized and maintained, methods by which the public, and particularly researchers, can access them, etc. Another meeting was mainly devoted to the legal aspects involved in creating longitudinal data, in making them available to the public, and future

plans – while stressing privacy protection. The expert team learned about the fields of anonymization and information security. The team issued two calls for literature reviews: “Mapping Information Bases in the Field of Education,” and selected Dr. Amalia Ran of the Information Center at the MOFET Institute for the task, and “Information Security and Privacy.” In the expert team’s estimation, the proposals received regarding the second topic were not of a high enough quality and it was therefore decided to obtain the required information through holding several meetings and telephone calls with experts, and through a non-comprehensive review prepared by the team coordinator.

As mentioned, the team addressed the issue of longitudinal data and as such, the differences between cross-sectional and longitudinal data will now be discussed, including the different types of longitudinal data and the research possibilities they provide.

Longitudinal and cross-sectional data: In general, there are two types of databases – cross-sectional and longitudinal. Cross-sectional data describe a population at a given point in time. In the social sciences, cross-sectional data is frequently utilized to find a relationship between various phenomena or behaviors of groups in the population. Cross-sectional data make it possible to simultaneously examine many variables although, in general, it is difficult to draw conclusions with respect to processes that are ongoing over time through data obtained in this manner.<sup>1</sup> In contrast, longitudinal data, broadly defined, are data collected for each variable during two or more periods, or the data are collected over time regarding those same variables being studied. According to this broad definition, every sequence of observations conducted one after the other can also be referred to as longitudinal data.

Types of longitudinal data: The research literature typically talks about three types of methods for generating longitudinal data. The first type is “repeated cross-sectional studies,” each one of which describes the same population at a different point in time. In this case, the samples – sub-groups of the total research population – can change from one period of measurement to another (independent samples) as long as they are part of the same population. Another type of longitudinal data is obtained through “retrospective surveys” – surveys conducted at one point in time but include questions about the subject’s past and thus collect information about his history. The third type of longitudinal data, which is the type the expert committee’s work dealt with and to which we will relate in this report, is data obtained from “panel studies” – which measure the same subjects<sup>2</sup> at two or more points in time.

It is customary to refer to two types of panel data: A representative panel and a cohort study. True to its name, a representative panel sample is a collection of data about a representative sample of the population being studied at two or more points in time. The difference between a representative panel sample and a repeated cross-sectional study is that the former tracks the same sample – the same sub-group of the population – over the entire period of time. Cohort studies track one group of people that share similar attributes or who have experienced

---

<sup>1</sup> However, cross-sectional data might include information on the memories of past data objects.

<sup>2</sup> Generally, this refers to following the same group of people over time but it can also track any research population, for example, schools, classes or families.

significant life events during a specific period of time. The most frequent example of a cohort study is tracking a group of individuals who were born during a certain period of time. This type of follow-up study is called a birth cohort. Other examples of cohort studies are research that tracks a group of people married during the same year, those exposed to a particular drug, and so on.

The main advantages of longitudinal data and longitudinal research: The first and perhaps most important advantage of longitudinal research over cross-sectional research is clearly the possibility of examining phenomena over time and to thus discover developments and identify patterns of change over time. Such research enables a series of events, circumstances, attributes and behaviors of the research population to be tracked.<sup>3</sup> By tracking over time, it is also possible to find relationships between events that took place at different times and not only events that occurred close in time to one another.

Longitudinal studies also enable better measurement of the changes in relationships between variables over time or alternatively, examination of the stability of the relationships over time. For example, if socioeconomic status influences IQ level at age three, but the relationship loses its strength by age 16, longitudinal research can much more easily discover this than cross-sectional research. Longitudinal research can also discover a delayed effect – a relationship between variables created only after time has passed – for instance, when a certain change occurs at a young age has an influence, but only after much time has passed. An example of this is the study by Chetty, Hendren & Katz (2015) which discovered that moving at a young age to a neighborhood that is of higher socioeconomic status affects the income of the children who moved at a young age, even though other studies did not find a relationship between such a move and academic achievement.

Not only do longitudinal data reinforce our insights with respect to change processes and relationships between variables, but when used within a suitable research design, they also enable the identification of causality. With cross-sectional research, there is difficulty in identifying causality, except in the case of randomized controlled trials. Determining causality allows the researcher to recommend interventions and is a necessary condition for reliable evaluation of intervention programs' success and for formulating policy.

Longitudinal research also gives researchers flexibility in the sense that it enables them to add focal points to the research during data collection (such as, for instance, adding more information while conducting the study). The researcher can choose to add additional questions to a questionnaire and in this way, examine hypotheses that arose after the start of the data collection process.<sup>4</sup>

The availability of longitudinal data has great importance especially for research in the field of education. The process of acquiring an education is an ongoing one, involves many players (students, teachers, parents, etc.), and during its course, many transitions occur between the different levels of education, and human capital is accrued. As a result, tracking the students and

---

<sup>3</sup> The professional term for this is “diachronic analysis.”

<sup>4</sup> Obviously, this option is limited to longitudinal studies based on panel surveys. As we will see below, not all longitudinal research is based on surveys.

their environment over time has great importance. Some argue that the complexity of the educational process, in practice, does not make its description possible without considering the past of the study participants.

Moreover, the influence of educational processes occurs in different time frames and its impact is not necessarily the same throughout the years. There are, for instance, events whose effect is immediately perceptible (for example, a student using drugs), and there are those whose outcome will be expressed after a certain amount of time (for example, a change in the quality of teachers is likely to affect test results only after some time). Only longitudinal data can teach us about the time it takes for an effect to occur.<sup>5</sup>

Longitudinal data from representative panel samples can be collected using two main methods: The first is the use of longitudinal surveys – a survey distributed among the same respondents at different periods of time. The second is the use of data collected in the past for a sample population, for instance, administrative data collected by the state. Longitudinal surveys have many advantages but they also come a few disadvantages that may hinder their use in Israel. Firstly, the cost of conducting such research is relatively high (both financially and time-wise) and requires funding over a long span of time. The scientific output of such surveys is attained after the passage of a relatively long time and therefore research based on their results can only begin after at least two “waves” of data have been collected. Furthermore, at times there is a need for a large number of interviews in order to enable analysis of change over time.

Scrupulous adherence is required to the operating procedures for sampling and data collection as well as in determining clear follow-up guidelines; commitment of the participants and the researchers over a long time period is required. Attrition, a situation in which those who were interviewed in previous waves cannot be interviewed in the present one, may impair the representativeness of the sample. This fact makes it very difficult to preserve representativeness and may even create a selection bias if more people with certain attributes (or whom certain circumstances affected) tend to leave the study and are not available to be surveyed. Finally, there is the problem of coverage of the target population, as a result of the sample being assembled in the first wave while in actual fact, the population changes over time. For example, longitudinal surveys conducted in the 1980s in the United States do not represent the current composition of the population of the U.S. from the standpoint of country of origin, and this is due to the immigration from Latin America and the decline in the birthrate of whites.

Compared to panel data, obtaining information through retrospective longitudinal data is relatively inexpensive and simple (since this is a single survey – as is the case for cross-sectional data). But these data also suffer from the problem of representativeness due to memory bias – mainly due to cognitive conditions, attitudes, emotions and motivation. There is also the problem of the interviewee’s tolerance owing to the large amount of information he is must provide. If the research seeks to measure change over a relatively short period of time (weeks or months), a retrospective study is preferable for investigating events and behaviors. If, however,

---

<sup>5</sup> For more information see Shavit & Feniger (2007)(Hebrew).

the study is meant to supply information about events that occurred years before it is reasonable to assume that a precise picture will not be attained by relying only on the participants' memory.

Considering these disadvantages and after having heard the challenges faced by the Central Bureau of Statistics in its work on the long-term survey whose transmission has recently begun, the expert committee members assessed that it would be difficult to locate an organization in Israel willing to run another long-term survey. This being the case, it was decided to place the emphasis on the second option for obtaining longitudinal data – linking up to administrative data measured by the state for the same individuals at different time periods.

Longitudinal data from administrative sources enjoys several advantages compared to those acquired from surveys: Administrative data generally cover the entire population of relevance and thus faithfully represent all of the population groups including the smallest among them. In contrast with longitudinal surveys in which there is attrition, in longitudinal administrative data this phenomenon is a drop in the ocean. In the normal course of events, administrative data are generated in any case and consequently, making them available to the public involves negligible budgetary cost. Administrative data are generally available on an annual basis and within a short amount of time from their collection, as opposed to surveys which are conducted once every few years and entail a relatively long process until they are available.

Despite our decision to focus on administrative data, it is important to note that there may be quite a number of disadvantages in their regard: The data tend to be less rich than those included in surveys, they do not include responses to subjective questions, and in certain areas their quality may be inferior. Beyond all these, creating longitudinal data from administrative data involves a legal and ethical problem of privacy protection and as a result, access to this data may also be reduced.

For many years already, prospective and retrospective longitudinal surveys have been conducted around the world. Prominent examples of such are the British National Child Development Sample (NCDS), the American National Longitudinal Study of Youth (NLSY), the German Socio-Economic Panel (SOEP), the American Panel Study of Income Dynamics (PSID), the British Household Panel Study (BHPS), and the German life history research study. It is important to mention the trend of combining prospective and retrospective panel data and to thus benefit from both these methods. The professional literature also recommends this approach or the combination of longitudinal with cross-sectional data.

In the past, few longitudinal surveys were conducted in Israel and they were mainly independently initiated by researchers studying specific populations. Currently, there are two important longitudinal surveys being conducted: The first, being conducted by the CBS, is the "Survey of Households in Israel – Long-Term Survey" begun in 2012 and to date, four waves of interviews have taken place. The survey focuses on processes Israeli households undergo over time, throughout the life cycle, with the objective of examining transitions and changes in different areas as a function of background characteristics of households and people in Israeli society, and to assist in the documentation effort of social and economic processes for policy planning and research. Some of the topics researched are "core questions" and others will

change every few waves and will constitute the “alternating segments.” Owing to the nature of the survey, it can serve as a platform for creating longitudinal data in the field of education. The second survey is the European Survey of Health, Ageing and Retirement in Europe (SHARE) which Israel joined in 2004.<sup>6</sup> This is an example of an international survey being currently conducted in 27 European countries and in Israel with the aim of creating a multidisciplinary database of longitudinal data on the older population (age 50 and above) in order to track the process of aging and its implications for a range of areas in life. To date, four waves of data collection have been conducted in Israel and the results of the first three waves have been published. Finally, retrospective data were combined with prospective data within the survey framework. One of the interview waves was dedicated to life history.

Along with the process of globalization, in recent decades international comparisons of education system inputs and outputs has enjoyed a revival and as a result, the number of surveys has increased as has the number of participating countries, including Israel. The surveys examined teachers’ perceptions and teaching methods (TALIS), student literacy in different subject areas (PIRLS, PISA and TIMMS) and adults’ basic skills (PIAAC). These are cross-sectional surveys; there are no international longitudinal surveys in the field of education.

---

<sup>6</sup> For more information, see: <http://igdc.huji.ac.il/home/share/introduction.aspx>.

## Lectures abstracts

### **Opportunities and Challenges in Collecting and Analyzing Longitudinal Educational Data**

David Kaplan

The purpose of this talk is to highlight the opportunities and challenges associated with the collection of longitudinal data in education. The talk is divided into three parts. Part 1 will outline the range of insights into educational processes that can be obtained from a careful collection of longitudinal data. These insights can be gleaned from such analyses as growth curve modeling and latent transition analysis. Examples of these possibilities will be provided. Part 2 will cover the challenges associated with the collection of educational longitudinal data, including issues surrounding the implementation and validation of file matching methods. The challenges surrounding data privacy will be briefly discussed. Part 3 will focus on extended analytical and policy-relevant opportunities that can be derived from the collection of educational longitudinal data, including the use of educational longitudinal data for building policy indicator/forecasting systems.

### **Four examples of the importance of longitudinal studies for policymaking and interventions**

Victor Lavy

Professor Victor Lavi will present four different studies which exemplify the use of administrative data as a source of information for longitudinal studies. Together, those four studies show the advantages of longitudinal studies for understanding the educational process.

1. **Teachers' Pay for Performance in the Long-Run: The Dynamic Pattern of Treatment Effects on Students' Educational and Labor Market Outcomes in Adulthood**: This paper examines the dynamic effects of a teachers' pay for performance (PFP) experiment on long-term outcomes at adulthood. I use a teachers' PFP experiment which I conducted a decade and half ago in Israel. In Lavy (*AER 2009*) I analyzed the short-term effects of this experiment on students' cognitive high school outcomes. Now, this earlier research presents an unusual opportunity to evaluate whether an intervention that offered teachers performance based bonuses for student test achievements has had a lasting impact on adult well-being. This paper provides the first evidence of links between teachers' PFP during high school and students' schooling and labor market outcomes in their late 20s and early 30s. Using longitudinal administrative data I show that the program led to a gradual increase in university education of the high school treated students, reaching a gain of 0.25 years of schooling at age 28-30. The effects on employment and earnings were initially negative, coinciding with a higher enrollment rate in university, but became positive and significant with time. These gains are largely mediated by the positive effect of the program on several high school outcomes, including quantity and quality gains in the high stake matriculation exams.
2. **The Long Run Economic Consequences of High-Stakes Examinations: Evidence from Transitory Variation in Pollution (joined with Avi Ebenstein and Sefi Roth)**: Cognitive performance during high-stakes exams can be affected by random disturbances that, even if

transitory, may have permanent consequences. We evaluate this hypothesis among Israeli students who took a series of matriculation exams between 2000 and 2002. Exploiting variation across the same student taking multiple exams, we find that transitory PM2.5 exposure is associated with a significant decline in student performance. We then examine these students in 2010 and find that PM2.5 exposure during exams is negatively associated with post-secondary educational attainment and earnings. The results highlight how reliance on noisy signals of student quality can lead to allocative inefficiency.

3. **Long Run Effects of Free [School Choice: College Attainment, Employment, Earnings, and Social Outcomes](#) at Adulthood:** In this paper I study the long term consequences of free school choice by taking advantage of an experiment conducted two decades ago in the city of Tel Aviv, Israel. This school choice program was very effective in improving high school attainment and cognitive achievements six years later (Lavy, *Review of Economic Studies*, 2010) and now I examine whether these effects persist beyond high school. Using administrative data, the results indicate that treated students, who on average are from low socio-economic background, experience significant gains in post-secondary enrollment and in completed years of education and also have higher earnings at age 30. These significant positive treatment effects reflect mainly an increase in academic education, through increased enrollment in three-years academic colleges but not in research universities, and some shift away from vocational education at adulthood.
4. **On the Origins of Gender Gaps in Human Capital: Short and Long Term Consequences of Teachers' Biases:** We estimate the effect of primary school teachers' gender biases in Israel on boys' and girls' academic achievements during middle and high school and on the choice of advanced level courses in math and sciences during high school. For identification, we rely on the random assignment of teachers and students to classes in primary schools. Our results suggest that teachers' biases favoring girls have an asymmetric effect by gender — a positive effect on girls' achievements and negative effect on boys' and vice versa. Such gender biases also impact students' enrollment in advanced level math courses in high school – girls positively and boys negatively. These results suggest that teachers' biased behavior at early stages of schooling has long run implications for occupational choices and earnings at adulthood, because enrollment in advanced courses in math and science in high school is a prerequisite for post-secondary schooling in engineering, computer science and so on.

### **Use of Longitudinal Administrative Data in Empirical Education Research**

Moshe Justman

The broad use of standardized tests in primary and middle schools, which began with the “No Child Left Behind” legislation in the United States, has spread worldwide. Such use has opened the door to longitudinal studies that track data on students from primary school up until completion of their secondary school studies and also continue with follow up on matriculation examinations, higher education, and entrance into the labor market.

Longitudinal follow-up studies such as these have a number of structural advantages:

**First**, they allow the researcher to track an entire cohort (or to systematically sample from the entire cohort), this in contrast to studies that begin to track at a later age and therefore the population they choose suffers from selection bias since it includes only students who stayed in school until that age. For example, a study that examines matriculation grades will only include those who took the exam and will thus miss out on information that could have been collected on students who dropped out earlier.

**Second**, the test results at earlier stages of study provide an indication of the student's learning ability at that stage and as a result, they allow isolation of the education system's influence on the student, her achievements and choices, an effect that is ongoing until the later stages of study. Knowledge gleaned in this manner can serve as evaluation of educational interventions and reforms at the level of the student, the teacher, the school and the education system, and can also serve to support such interventions.

In his lecture, Professor Justman will illustrate these advantages through two studies he recently conducted. One, in collaboration with Brendan Hough, utilizes Australian data and showcases the high degree of precision with which matriculation exam results can be predicted (in their Australian version) based on scores on standardized tests taken in ninth grade and the student's socioeconomic background. Predictions obtained in this manner can serve as a tool for advising the student (at a relatively early stage) regarding a learning track and a career. These predictions can also be used to assess – based on a comparison between the predicted and actual results – the school's functioning at the secondary school level.

The second study, conducted in collaboration with Naomi Friedman-Sokoler, is based on data in Israel and connects between the *Meitzav* test scores (Growth and Effectiveness Measures for Schools) in eighth grade and these same students' scores on matriculation examinations five years later. The study focused on gender differences in choosing math and science as subjects to major in for the matriculation exams. From the longitudinal data, it can be learned that boys' preference for physics and computers and girls' preference for biology and chemistry is not affected by differences in their math abilities, as many have claimed, since these differences in preference exist even when boys and girls have identical grades in the eighth grade *Meitzav* math tests. These results can also serve in designing policy to increase girls' participation in the leading science professions, such as engineering and computer science, as a career choice.

This review focuses on contemporary and historic databases on the education system in Israel, including sources on pre-school programs, elementary and secondary schools, institutions of higher education, and the non-formal education system. The survey maps the information, which was gathered from administrative sources, studies, and databases from public institutions, private institutes, and non-profit organizations. It includes sources, such as administrative files and statistical surveys, particularly databases, which contain the following details: information about students and their achievements, information about teachers and the pedagogical staff, information on educational institutions, information on education budgets and funds, information on infrastructure, non-formal education, and candidates for military service.

The aim of this review is to map the existing databases on education in Israel under the Initiative for Practical Research on Education by the Israeli National Academy for Sciences, which investigates the use of measurable longitudinal data as a source of information for policy and educational programming in Israel: longitudinal surveys as a case study. In order to complete this mission, the review will present three types of databases: open databases, which are available for the public; semi-open databases, which require entry permissions; and classified databases, which are not open for the public. The information was gathered from administrative sources, surveys, and databases from public authorities, private institutes, and non-profit organizations, and it will be presented as follows:

- Principle institutions in Israel and the description of their mapped databases, including data on national level or based upon samples, which represent the research population.
- Additional databases held by major municipalities, educational networks, and non-profit organizations.
- Other databases created specifically for researches on education. These databases will be organized in a bibliographic list.

According to the team of experts on behalf of the Initiative for Practical Research on Education of the National Academy of Sciences, in recent years, there is a growing awareness in Israel toward the advantages of longitudinal data as a source for policy decision-making and qualified research on different aspects of the education system. Similarly, a growing demand for longitudinal data reshapes the use of this information, while understanding that schooling is a dynamic, multi-dimensional and an ongoing process, which is impacted by decisions taken over different periods of time by educators, parents, students and other agents. Therefore, the following pages will review the databases and files, which focus on education, in order to constitute a source of information for policy decision-making in this field.

All the databases in this framework were selected according to the following criteria:

- Databases on education that contain data on students, teachers, educational institutions, etc.
- Databases that contain identification details on the research unit (such as student's ID number, name of institution, institution code, etc.), in order to relate variables from different databases.
- Data that may be computerized.

Hence, this review consists of 120 databases, which are open or semi-open, as well as a list of 35 additional classified databases, which are held by the Ministry of Education, and are not available for the wide public. The review will present shortly the mapped databases in questionnaires, which were completed either by the team of analysts on behalf of the Information Center in MOFET Institute or by the liaison in each one of the interviewed institutions. Classified databases or databases, which were not available from different reasons, will be described shortly also. Available databases, which were not mapped due to the lack of collaboration on behalf of the relevant agencies, will not be included in this review, although they will be mentioned according to the depositor of the information and the relevant contact person. In order to facilitate the search in this guide, all the databases were organized in an

alphabetical order according to three categories mentioned previously. Additionally, the reader may use the index list at the end of this review, in order to search for a specific database.

In this review, the team of analysts employed various methodological tools for collecting the information on the databases and describing them. First, the team mapped the relevant agencies and institutions, which relate to the field of education (see Appendix 1). Official request letters on behalf the Initiative for Practical Research on Education of the National Academy of Sciences and on behalf of the team of analysts of the Information Center in Mofet Institute were sent to liaisons in 70 selected institutions. Second, the team conducted follow-ups by phone conversations and electronic mails with the different agencies, as well as interviews and meetings with the relevant representatives, in order to complete the data collection process. At the same time, the team identified and classified additional databases by searching the Web for different Internet sites, which facilitate this information, as well as by examining online files and other sources, such as annual reports, summaries and scholarly researches.

Upon completing the data collection, the team evaluated the existing sources, in order to estimate its relevance for the specified requirements mentioned previously. Databases, which contained all the required information, constitute part of this guide and the general database on education. Lastly, upon completing the data evaluation, the relevant databases were mapped according to identical criteria, and they constitute the core of this review. These criteria contain variables, such as name of the database, depositor of database, initial date of collecting information, last date of collecting information, frequency of information update, population, research unit, number of units, variables, source of information, etc.

We should emphasize that this survey focuses on a very general field, which relates to many spheres of life and to immense number of agencies. Most of the information is not accessible for private researchers. Hence, an official appeal on behalf of the National Academy of Sciences and the Initiative for Practical Research on Education was required, in order to address the relevant liaisons in charge of these databases. Consequently, the results in this review rely on the cooperation of these agencies in accordance with the Law for Freedom of Information. Additionally, databases from institutions and agencies, which did not respond to our repetitive requests, were not included in this framework. They are classified under non-responsive organizations.

It is important to note that the present guide does not analyze the databases or investigates the type of variables gathered by these sources. The analytical task in this review is based upon a search of databases on education by addressing directly the relevant liaisons in each agency, and conducting an ongoing follow-up on the acceptance to share this information or the refusal to grant access to it. Furthermore, a virtual search according to different indexes was conducted by the team of analysts, who also scanned the official websites of the relevant agencies according to different key words, such as: databases, education, online databases, educational programs, etc. Upon collecting the information and evaluating it, the databases were mapped according to our detailed proposal (see Appendix 1) in an alphabetical order.

Moreover, this guide does not propose any specific research orientation; it aims to offer the researcher, who reviews the catalogue of databases mapped here, a variety of databases on

education, which were collected from different institutions and agencies. Therefore, its aim is to assist in mapping the existing information in Israel on the field of education. For this purpose, this review scanned databases and files on preschoolers, pupils in schools, institutions of higher education, non-formal education, yeshivas, professional training programs, etc., as well as researches in education that contain additional files and databases.

Lastly, we wish to thank all our liaisons in the different organizations and agencies, who assisted us in completing this mission, for their cooperation and willingness to share their databases. Our gratitude goes also to the team of analysts in the Information Center of Mofet Institute for their contribution in completing this catalogue.

### **The Ministry of Health's Timna Project**

Aline Attias

**Background:** The health community in Israel generates a wide range of clinical, administrative and other data about the treatment provided to people living in the State of Israel, public health, and the state of the health system. The range and scope of the data alongside the challenges and complexity of providing medical services are fertile ground for research-based statistical study.

Advanced technologies from the field of Big Data make it possible to analyze data in ways that until recently could not be processed by a computer: This refers not only to improved performance or to analyzing large databases but to the use of cognitive computing, that is, “machine learning” and processing of unstructured information such as free text, imaging, video, voice, data from sensors and all information available on the Internet. New processing methods enable rules to be identified even in data sets where statistical significance is not attained.

**National Big Data system for health:** In the framework of the Timna project (Hebrew acronym for “Big Data Research Infrastructure”) the Ministry of Health is promoting a process of data collection, storage and analysis using the described technologies in the aim to help the health system and various research bodies to generate benefit from the data for the public.

The platform the Ministry is building is part of a worldwide trend in the health field to realize the enormous potential inherent in the data to improve health in order to, among other things, enable:

- Create innovative treatment processes that correspond to the patterns and templates in the information that could not be identified until today.
- Individually customized treatment and customized treatment plans that have the highest probability of leading to recovery, a plan that would be adapted to the patient's profile rather than providing homogeneous treatment to a heterogeneous population.
- Promoting preventive medicine through creation of pre-disease profiles which would enable provision of behavioral recommendations and preventive treatment; today, most of the health system's energies are directed towards coping with disease after it appears.
- Proper use of resources to establish an existing health system in a world of limited resources, through meticulous examination of bottlenecks and analysis of the causes of proper or deficient performance.

- Setting evidence-based policy at the strategic, operational and tactical levels while monitoring and supervising the performance of health organizations and regulatory tools.

The research platform being established by the Ministry of Health will be based on a central database that will allow data from different sources within, and outside of, the health system to be cross-referenced. All of the data entering the database will undergo anonymization of all identifiers and in all of the fields that can lead to identification – quasi-identifiers.<sup>7</sup>

Following a process of receipt of ethics and regulatory authorizations, a dedicated research environment will be created for conducting specific research. “Research files” (that will contain a cross-referencing of all the data and information authorized for the specific research from the central database) will undergo deep anonymization including meeting k-anonymity conditions. The research environment will include tools for data processing and will enable the research team to collaborate and to receive the help of data scientists, as needed. The research environment will be autonomous and will not allow data to be transmitted to the world outside. Any additional information in the research group’s possession will be entered into the research environment only under conditions of supplementary deep anonymization. There will also be a separation between research environments so that transfer of data between research environments will not be possible.

In order to facilitate attainment of the functionalities required by the type of research, the research instruments in the research environment will be determined by the researchers’ experience and preferences. In the intent of encouraging innovation, it will be possible to make use of new developments and technologies, including installation of unique tools, according to the research team’s preference, using a “Bring Your Own License” model.

**Anonymization:** The anonymization process is garnering special attention within the Timna project due to the high sensitivity inherent in exposing medical information. At the same time, the more anonymous the information, containing a more limited range of details – the lower its research value. The balance must be found between individual protections and actions taken for the sake of public health, in general.

The deep anonymization process performed on the research file includes data encryption corresponding to different data types, while retaining their research utility:

- One-way tokenization and omission of sensitive data
- Corruption via noise or circling, for example
- Grouping through presenting a range of values, for example
- Exchange through, for example, synthetic data or in a response based on a changing random sampling from the entire population

**Regulation:** In parallel to establishing the technology platform and operational preparedness, the Ministry of Health is also acting to regulate the secondary use of medical information, that is, for research purposes.

---

<sup>7</sup> For a survey of the professional terms that appear here, see Chapter Three in these materials.

In the aim of formulating rules and guidelines, the Ministry of Health placed the topic on the agenda and for discussion in various public forums while also studying the situation around the world. The Ministry of Health is adopting the OECD draft guidelines published in April 2016 both for the process of regulation formulation and on the issue of software.

The draft guidelines, in part, requires that its formulation be conducted in consultation with a wide range of stakeholders and built through intra-governmental coordination, and will strive to promote cooperation among (public and private) organizations in processing the information, while acknowledging the role of technology in advancing innovation. The draft guidelines also dictate essential principles regarding the method of authorizing research requests, securing information and privacy protection, skills development and training employees in the field, control and enforcement.

A public committee (to include several work groups that will examine various facets) is deliberating these issues and its intention is to shortly submit its recommendations for regulation on the issue.

### **Achieving Balance between the Public Benefit from Establishment of Databases based on Personal Information and the Individual's Right to Privacy**

Limor Shmerling Magazanik

**The Israeli Law, Technology and Information Authority (ILITA)** is a regulatory and enforcement authority and a professional entity, expert in protecting individual privacy in Israel. Its powers are based on the Privacy Protection Law of 1981 and its work relates to all the databases held by private parties and public bodies.

ILITA's aim is to promote protection of each person's basic right to privacy and to enforce the provisions of the Privacy Protection Law. The Authority's activities are performed using diverse regulatory tools including administrative and criminal supervision and enforcement, publication of market guidelines, regulation and registration of information databases, strengthening personal information security and raising public awareness of the right to privacy. ILITA is active on an international scale, in the field of education, and the fields of publicity and training.

**The right to privacy** is a basic right in Israeli law. It has legal status and its protection is meticulously and explicitly anchored in the Privacy Protection Law. Its main principle declares that use of an individual's personal information constitutes a violation of his privacy, if his prior consent has not been obtained. Alternatively, an individual's personal information can be utilized without harming his privacy if given legal authorization that meets the legislative requirements of Basic Law: Human Dignity and Liberty.

In the context of databases that contain personal information, the discussion centers on the limits on its use that apply to the database owner, and on the right of the individual to control the management of personal information that concerns him, to edit or delete information and to decide with whom and how to share it.

Our desire is to ensure that despite this, if violation of an individual's privacy did occur, none of his other rights will be harmed – namely, the right to equality and avoidance of discrimination,

the right to life, the right to property, the right to innocence, the right to freedom of thought and expression, employee rights, etc.

When the government collects personal information about its citizens, it is the obligation of public organizations and policymakers to define the purposes for which the data was collected and to guarantee – through implementation of appropriate means to protect information and an effective enforcement system – that there will be no deviation from the stated aims. In addition, the range of effective control available to the individual and the means at his disposal in order to control information about himself, must be defined.

Since this is a constitutional right, the project planners must guarantee that the use of personal information will be for a worthy cause and to no greater extent than required.

**Principles in establishing a database:** Creating a database for research in the field of education means establishing a new super-database in Israel, a database which does not currently exist, that includes sensitive personal information about most of the population, and whose data was not collected from the individuals themselves. Establishing a database requires collection of information from databases, some of which are held by the state authorities and some in private hands – and unifying them into a common database. On the matter of transferring information between them, public organizations are subject to the provisions in Section D of the Privacy Protection Law and are not permitted to share with anyone the information they have collected for the purpose defined in the law or for other purposes, including those that are public in nature – only under the conditions set by the law.

Moreover: Even if at the end of the day, the information will be made available after undergoing anonymization processes (which do not entirely guarantee the impossibility of identification), it must, throughout, follow the rules of proper management outlined in the privacy protection laws since at its source, it is identified personal information.

We are thus referring to creation of a new database which poses new challenges and risks to privacy that are not identical to those present in each one of the database parts. On the assumption that there is no intention to request consent for a new use of the information from each one of the individuals whose information is found in the database, it may be that the decision to establish it must be backed by primary legislation. To do so, there would be a need to explain and justify before the authorized parties – according to the applicable legislative tests – the necessity of establishing the database and its aim, and to prove that it is required in order to realize a worthy public purpose.

A responsible party must be designated “database owner”: A governmental authority that will manage the project’s establishment as well as its maintenance throughout its lifecycle, in line with the law’s requirements. Legislation will also list the requirements for securing the information in the database.

In order to characterize the project, a survey should be conducted to assess the risks to privacy and to choose to execute the project according to the principles of “privacy by design.” The

database's creation may be realized in different ways, in line with the operational needs of the system and corresponding to the risks identified and which require a solution.

In her remarks at the seminar, Attorney Shmerling will detail and illustrate the considerations and dilemmas in the process of creating solutions of this type, processes in which ILITA was involved in the context of government projects – the “credit information database,” the “pension clearing house,” and the “research on medical information database.”

#### The credit information database currently being established by the Bank of Israel

The aim of this project is to introduce competition and to lower the interest rates on credit for households from bank and non-bank lenders. Analysis of this issue revealed, in part, the problem of the gap in information between the banks and the non-bank lenders, gaps which, in the opinion of the project initiators, are the reason that competition does not exist. The solution recommended is establishment of a database of the citizenry's positive and negative credit, a database which will allow a report to be immediately ordered about each person applying for credit from a lender.

The project raised questions regarding individual consent to the collection of information and its transmission, the need for legislation, identity of the database owner, identity of authorized database users, location of the database, possibility of using outsourcing, and structure and characterization of the system. These are all factors that will minimize the risk of violation of the privacy of the information subjects and will enable research on statistical information databases.

#### The pension clearing house established by the Ministry of Finance's Commissioner of Capital Markets, Insurance and Savings

The aim of the project was to make it easier for the public to manage its own pension funds in an optimal manner. Analysis of the problems revealed, in part, the difficulty of people, financial advisors, and insurance agents to collect information on the various pension products an individual possesses that are held by diverse insurance companies. The solution was designed to enable quick and efficient collection and centralization of such information about the individual so that relevant pension advice could be provided on its basis.

The project raised questions with respect to individual consent to the collection of information and its transmission, the need for legislation, identity of the database owner and its location, identity of authorized database users, possibility of using outsourcing, characterization and structure of the system so as to minimize the risk of harm to the people, digital identification, and duration of data retention.

#### About the medical information database for research being planned by the Ministry of Health

The goal of the project is to advance medical research and health services through the use of the large databases containing long histories amassed by the sick funds, hospitals and medical institutions in Israel.

The project raised questions related to individual consent, need for legislation, location of the database, and use of cloud computing services.

## Bibliography

- יוסטמן, מ' ובקובזה ג' (עורכים) (2010), קווים מנחים לרענון מערך האינדיקטורים לחינוך בישראל: סיכום, מסקנות והמלצות, היזמה למחקר יישומי בחינוך, האקדמיה הלאומית הישראלית למדעים, ירושלים.
- רן, ע. (2016) מיפוי מאגרי מידע בישראל בתחום החינוך. ל' יוספסברג בן-יהושע (עורכת), סקירה מוזמנת כחומר רקע לעבודת צוות המומחים בנושא "שימוש בנתונים הנמדדים לאורך זמן כמקור מידע למדיניות ולתכניות חינוכיות בישראל", היזמה למחקר יישומי בחינוך.
- שביט, יופניגר י. (2007), *מחקרי אורך במערכת החינוך הישראלית*, נייר עבודה שהוגש למדען הראשי במשרד החינוך.
- קאנט, ע. (1984) *הנחת יסוד למטפיזיקה של המדות*, תרגום: מ. שפי, הוצאת מאגנס, תשמ"ד 1984.
- Chetty, R., Hendren, N., & Katz, L. F. (2015). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.
- Dynarski, s. (2014), *Building Better Longitudinal Surveys (on the Cheap) Through Links to Administrative Data*, National Academy of Education.
- El Emam, K., Jabbouri, S., Sams, S., Drouet, Y., & Power, M. (2006). Evaluating common de-identification heuristics for personal health information. *Journal of Medical Internet Research*, 8(4), e28.
- HITRUST De-Identification Working Group, (2015), De-Identification Framework, Health alliance Trust web-site, on-line at:  
[https://hitrustalliance.net/documents/de\\_id/HiTrustDeIdentificationPresentation.pdf](https://hitrustalliance.net/documents/de_id/HiTrustDeIdentificationPresentation.pdf)
- Garfinkel, S (2015), De-Identification of Personal Information, U.S. Department of Commerce, National Institute of Standards and Technology Internal Report 8053, on-line at:  
<http://dx.doi.org/10.6028/NIST.IR.8053>
- McFarland, S. J (2012) Why we care about privacy, on-line at: <https://www.scu.edu/ethics/focus-areas/internet-ethics/resources/why-we-care-about-privacy/>
- Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). I-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 3.
- Gavison. R, "Privacy and the limits of law," *Yale Law Journal* 89 (3) (Jan 1980).
- Dynarski, s. (2014), *Building Better Longitudinal Surveys (on the Cheap) Through Links to Administrative Data*, National Academy of Education.

## Short bios

### Expert team members:

**Leah Achdut** (Chair), Member of the Economics and Society Program at the Van Leer Institute and a senior lecturer at the School of Social Sciences and Management at the Ruppin Academic Center.

**Michal Beller**, President of Levinsky College of Education.

**Orly Manor** has, since 1990, been a member of the academic staff of the School of Public Health and Community Medicine at the Hebrew University and Hadassah Medical Center.

**Iris Tabak**, A senior lecturer in the Department of Education at Ben-Gurion University of the Negev

**Tal Zarsky**, An Associate Professor in the Law Faculty at the University of Haifa

**Noam Zussman**, Economist, Research Department, Bank of Israel

### Speakers at the Conference (By order of appearance)

**David Kaplan**, the Patricia Busk Professor of Quantitative Methods in the Department of Educational Psychology at the University of Wisconsin – Madison.

**Moshe Justman**, Professor in the Department of Economics at Ben-Gurion University and former dean of its Faculty of Humanities and Social Sciences.

**Victor Lavy**, Professor in the Department of Economics at the Hebrew university in Jerusalem, and a Professor in the Department of Economics in the university of Warwick, England.

**Amalia Ran**, lecturer and researcher in the field of Latin American Studies.

**Aline Attias**, director of the Big Data platform and anonymization at the Ministry of Health's Digital Health and Computing Division.

**Yoel Finkel**, deputy government statistician at the Central Bureau of Statistics (CBS)

**Yuval Eylon**, Staff member in the Department of History, Philosophy, and Judaic Studies in The Open University of Israel

**Raanan Sulitzeanu-Kenan**, senior lecturer in political science and public policy at the Hebrew University of Jerusalem.

**Limor Shmerling-Magazanik**, director of Department of Public Affairs and Government at the Ministry of Justice's Israeli Law and Information Technology Authority (ILITA)