The Initiative for Applied Education **Research**

Expert Team Project Report

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

Coordinated and edited by Oded Busharian

Project Report — Summary

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

Excerpts from the Expert Team Project Report

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

Coordinated and edited by

Oded Busharian

Translated from the Hebrew Original

Jerusalem, 2017

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

Jerusalem, 2017
The Initiative for Applied Education Research
The Israel Academy of Sciences and Humanities

Language editor (Hebrew original): Ada Paldor
English translation: Judyth Eichenholz
Proofing and production: Ayala Vlodavsky
Design: Esti Boehm
Cover design: Adi Mendler

This document is published by the Israel Academy of Sciences and Humanities
and is available to the public at: http://education.academy.ac.il

Any use of, or quote from, this document should cite the source as follows:
Busharian, O. (ed.) (2017). *Using Longitudinal Data as a Source of Information for Education Policy and Programs in Israel, Project Report*. Jerusalem: The Initiative for Applied Education Research, Israel Academy of Sciences and Humanities.

The Israel Academy of Sciences and Humanities was founded in 1959. Its membership currently comprises close to 100 top Israeli scientists and scholars. The Israel Academy of Sciences and Humanities Law, 1961, declares that its principal objectives and tasks are to foster and promote scientific activity; to advise the Government on research activities and scientific planning of national importance; to maintain ties with foreign academies of science; to represent Israeli science at international institutes and conferences; and, to publish articles that can further science.

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. This kind of information is crucial for the intelligent formulation of policy and for optimal planning of interventions to improve educational achievements in Israel.

The Initiative's vision: Research knowledge is an essential component for planning public policy or comprehensive interventions. In the planning phase, critically-appraised research knowledge supports the formulation of policy whose chance of success is greater, and at a later point, enables rational public discourse to take place. The Initiative implements this vision in the field of education.

The Initiative's method of operation: The issues the Initiative addresses are those raised by decision-makers and it consults with senior Ministry of Education officials and other stakeholders. The Initiative's steering committee, appointed by the president of the Israel Academy, is responsible for the Initiative's work program and the peer-review processes of documents it creates. The Initiative operates by means of expert committees and by convening joint symposia for researchers, professionals in the field and decision-makers. It publishes a variety of reports and makes them available to the public. Members of expert committees carry out their work on a voluntary basis.

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 the Rothschild Foundation (Yad Hanadiv). Since the beginning of 2010, the Initiative has been operating as a unit of the Israel Academy. In the summer of 2010, the Israeli Knesset amended the Israel Academy of Sciences and Humanities Law, regulating the Israel Academy's advisory role vis-a-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 – Ruppin Academic Center and the Van Leer Institute in Jerusalem (chairperson);

Prof. Michal Beller – Levinsky College of Education;

Prof. Orly Manor – Hebrew University of Jerusalem and Hadassah Hospital;

Dr. Iris Tabak – Ben-Gurion University of the Negev;

Prof. Tal Zarsky – University of Haifa;

Mr. Noam Zussman – Bank of Israel.

Activity coordinator: Mr. Oded Busharian

Acknowledgements[1]

The report before you is the fruit of the expert team's labor, which had come together to promote public good in Israel. The team coordinator, who also edited this report, would like to, first and foremost, thank the expert team members for their dedicated work, their availability and their willingness to contribute of their time with no remuneration involved. I would like to thank the expert team chairperson, Professor Leah Achdut, a staff member of the Van Leer Institute's Economics and Society Program and head of the Department of Economics and Management at the Ruppin Academic Center, who directed the work of the expert team with sensitivity and efficiency. I would also like to thank each and every one of the expert team members: Professor Michal Beller, president of the Levinsky College of Education, Professor Orly Manor, faculty member in the School of Public Health and Social Medicine at the Hebrew University of Jerusalem and Hadassah Hospital, Dr. Iris Tabak, senior lecturer in the Department of Education at Ben-Gurion University of the Negev and head of the Curricula and Teaching track, Professor Tal Zarsky, associate professor in the Law Faculty at the University of Haifa, and Mr. Noam Zussman, economist at the Bank of Israel's Research Division.

The path of the activity and its key method of study took place through learning sessions in which researchers involved in areas related to the activity participated, as did policymakers and officials at public institutions who have knowledge and experience in those areas. The expert team wishes to thank the speakers and lecturers at the various learning sessions for their readiness to give of their time, knowledge and experience towards the success of this activity. I would, therefore, like to thank the professionals at the Ministry of Education (MOE), the Central Bureau of Statistics (CBS), and The National Authority for Measurement and Evaluation in Education (RAMA), the three main public institutions that possess and utilize educational data: Dr. Itay Asher, in charge of Education Experiments and Research and acting chief scientist at the Ministry of Education, who presented some of the challenges Ministry professionals face when they wish to use the data the Ministry collects or to share these data with researchers; Ms. Nurit Dovrin, director of social surveys at the CBS, Ms. Oriya Levy, senior professional coordinator for the longitudinal survey conducted by the CBS, Ms. Nardit Stein-Kapach, longitudinal survey coordinator at the CBS – all three of whom together presented the CBS's longitudinal survey, the opportunities it provides, and the difficulties the CBS is encountering in conducting it. Mr. Yoel Finkel, Associate National Statistician at the CBS told us of his activities at the CBS and the partnerships it creates in order to realize its vision of unified national statistics that are accessible for research purposes. Thank you to Dr. Chaim Gatt, director of the Ministry of Education's information systems and responsible for the Computing Administration's research and statistics, who presented the MOE's research room and the possibilities it opens up for researchers; Dr. Hagit Glickman, director general of RAMA (The National Authority for Measurement and Evaluation in Education), who talked about the use RAMA professionals make of longitudinal data in order to promote education in Israel. Mr.

[1] The names appear in alphabetical order; the professional affiliations listed are those that were current at the time the activity took place.

Yossi Gidanian, director of the Instruction, Higher Education and Teaching Personnel sector at the CBS who presented the long road the CBS traveled since the 1980s on the issue of collection and utilization of education and higher learning data in Israel. Mr. Pinhas Klein, economist and senior coordinator (statistical processing and analysis) at the MOE's Economics and Statistics Division, part of the Economics and Budgets Administration, shared information with us about the Administration's budgetary transparency system which it makes available to researchers and the public at large. Ms. Aliza Krovi, head of the ISOPED System and the Education Registrar in the Education and Higher Learning sector at the CBS, who presented the ISOPED system being developed at CBS. Mr. David Maagan, head of the Higher Education and Teaching Personnel Statistics sector at the CBS, who spoke with us about the use the CBS makes of administrative data for longitudinal research purposes. Mr. Adnan Mansour, director of the Learning Transitions and Follow-up sector at the CBS, brought the team up to date with respect to the process of receiving data files from the CBS. Adv. Dorit Morag, legal counsel at the MOE, answered the learning session participant's questions concerning the MOE's position on transmitting data and its cooperation with researchers and other public institutions. Adv. Brian Negin, legal counsel at the CBS and responsible for freedom of information, presented his practical experience with reference to information, privacy and confidentiality. Dr. Michal Tabibian Mizrahi, Director for Strategy and Planning at the MOE, who spoke about the data required for strategic planning at the MOE.

Next in line for thanks (and not in order of importance) – the speakers and presenters at the learning sessions who arrived from other government ministries or from private organizations: Thanks to Mr. Ziv Ofek, founder and executive director of CDI-Negev, who presented innovative ways to make data available while protecting the privacy of the people behind them. Ms. Talal Dolev, founder and former director of 360°, the national program for children and youth at risk, who spoke of the way in which the program confers a culture of data use to education and therapy practitioners in Israel. The team expresses special thanks to Professor Adam Gamoran, president of the William T. Grant Foundation, who shared with us the results of the American National Academy of Education's expert committee which he chaired, as well as his attempts to promote use of administrative statistics for research purposes in the United States. Dr. Ran Wolff, of the Yahoo Research Laboratory in Haifa, who presented a sociological analysis of the concept of privacy and spoke about the way large private technology companies, which possess much data about their users, relate to the issue of privacy protection.

The expert team sincerely thanks the speakers at the seminar which concluded the team's year of activity: Professor David Kaplan, the Patricia Busk Professor of Quantitative Methods in the Department of Educational Psychology at the University of Wisconsin-Madison. Professor Moshe Justman professor emeritus in the Economics Department, Ben-Gurion University in the Negev, Professor Victor Lavy, the William Haber Professor of Economics at the Hebrew University of Jerusalem and professor in the Department of Economics at Warwick University in the U.K, Dr. Amalia Ran, researcher for the MOFET Institute and lecturer and researcher in Latin American Studies at various institutions. Ms. Aline Attias, director of the Big Data Platform and Anonymization at the Ministry of Health's Digital Health and Computing Division.

Professor Raanan Sulitzeanu-Kenan, senior lecturer in political science and public policy at the Hebrew University of Jerusalem and a research fellow at the Israel Democracy Institute. Adv. Limor Shmerling Magazanik, director of the Department of Public Affairs and Government at the Ministry of Justice's Israeli Law and Information Technology Authority (ILITA).

In addition to the speakers at the learning sessions, participants included invitees who did not speak but were respondents to lectures and took part in the discussion that developed in connection with the lecture. We also thank these participants for their contribution: Professor Michael Birnhack, lecturer in the Law Faculty at Tel-Aviv University, Adv. Sima Adiv, of the MOE's legal counsel; Ms. Renee Osisohn of the MOE's Chief Scientist's Office; Ms. Nurit Lipshtat, director of the Longitudinal Research area at RAMA; Dr. Gilmor Keshet-Maor, director of the Sciences Division at the MOE's Pedagogic Secretariat; Adv. Zivit Mozer, responsible for Public Affairs and Government at the Ministry of Justice's Israeli Law and Information Technology Authority; Mr. Avner Kantor, research fellow at the Internet Research Center at the University of Haifa, who leads the "Open PISA" project to make PISA test data accessible and researchable (openpisa.org); Dr. Tammy Harel Ben Shahar, lecturer in the Faculty of Law at the University of Haifa and director of the faculty's Law and Social Change Clinics; Adv. Ruti Horn, former legal counsel to the National Insurance Institute; Ms. Miri Sabag Endweld, head of the National Insurance Institute's Economics Research Division, part of the Research and Planning Administration. Our thanks are conveyed to Professor Audrey Addi-Raccah of the School of Education at Tel-Aviv University; Professor Ron Astor of the School of Social Work and School of Education at the University of Southern California; Professor Rami Benbenishti of the School of Social Work at Bar-Ilan University; Professor Yoav Benjamini of the School of Mathematical Sciences at Tel-Aviv University; Ms. Efrat Degani-Toperoff of Yad Hanadiv; Mr. Yuval Vurgan, head of the Education Area at the Knesset's Research and Information Administration; Ms. Eti Weissblei of the Knesset's Research and Information Center; Mr. Assaf Weininger, also from the Knesset's Research and Information Center; Mr. Ido Levita of the Senior Division for Planning and Strategy at the MOE; Ms. Hila Lancry Miost of the Senior Division for Planning and Strategy at the MOE; Ms. Sofia Mintz, former head of the Computing and Information Systems Administration at the MOE; Mr. Gal Fisher, head of the Education Sector at Yad Hanadiv; Ms. Aliza Krovi, head of the ISOPED System and the Education Registrar at the CBS; Dr. Analia Schlosser of the School of Economics at Tel-Aviv University; Ms. Mazie Sherter of the Senior Division for Planning and Strategy at the MOE.

The team extends its thanks to the authors of the scientific literature review on the topic of mapping databases in Israel in the field of education – Dr. Amalia Ron and Dr. Liat Josefsberg Ben-Yehoshua of the Information Center at the MOFET Institute, who dedicated many hours to gathering the information, writing it up and editing it, and submitted a comprehensive and high quality report.

The expert team thanks Dr. Avital Darmon, the director of the Initiative for Applied Education Research as well as Mr. Uri Harosh, Ms. Ayala Vlodavsky, and Ms. Ada Paldor, whose efforts, time and talents were critical to creating this report.

This report underwent the customary process of independent peer review. The report editor and the expert team are thankful for the review which helped ensure the report's clarity, its quality and its independence. Responsibility for the report contents rests with the editor of this document.

Table of Contents[2]

[2] The Table of Contents reflects the contents of the full Hebrew-language report. Grey text signifies parts of the report that were not translated.

Introduction and Summary

Motivation for the Expert Team's Creation and its Activity

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, with the support and encouragement of Yad Hanadiv, assembled an expert team 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 team are: Professor Leah 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; Dr. 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 (a list of the meetings participants' appear in the Appendix of this report). 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 also studied 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.

In the winter of 2016, the expert team held a seminar to mark the conclusion of its activity and present the findings it amassed to the public at large. The seminar revolved around three topics which correspond to the topics that arose during the learning process: The first session dealt with longitudinal data and their use for research purposes and in formulating an intelligent education policy. The second session concerned existing administrative data in Israel in the education field and the methods (current and desired) for making them accessible to researchers. The third session included a discussion about the need to protect administrative data subjects' privacy and the methods for doing so.

Summary: Insights from the Expert Team's Activity

During the course of the activity, the expert team formulated a number of insights. Some were explicit outgrowths from lectures given by speakers at the learning sessions, some came from reading the research literature on the topic, and some from the team members' familiarity with the topic and their expertise in the area. In the summary which follows, we will present these insights within the topic areas the present report addresses. It is important to mention that these insights are not based on wide-ranging research or on a comprehensive review of the scientific literature on the topic. Thus, these insights should not be viewed as recommendations unanimously endorsed by the expert team. Nonetheless, the expert team hopes that the insights will enable a more intelligent articulation of policy on these topics.

Longitudinal Data, Longitudinal Research and their Use in Education Research and Policy Formulation and Evaluation

1. The experience the Central Bureau of Statistics (CBS) accumulated from a decision making process that went on for about 10 years, concerning whether to conduct a longitudinal survey ("long-term survey") is instructive regarding the many difficulties involved in enlisting sources of public funding for conducting longitudinal surveys –owing to their high cost and the long-term commitment the funding bodies must take on already at the outset. In addition, the experience amassed in Europe and Israel around the SHARE project shows that with the years, these surveys become more complex and require great effort in locating sources of funding for each wave of interviews. Likewise, difficulties deriving from respondents

dropping out emerge over time. We believe that it is desirable to discuss the question of whether to conduct additional longitudinal surveys, for example in the field of education, in the context of deliberations over creation of a national data infrastructure in whose framework a decision about embarking on a longitudinal survey would also depend on the assurance of long-term funding for its execution.

2. It is important to note that although the expert team decided to focus on the option of combining administrative data measured over time in order to generate longitudinal data, there are many areas in which longitudinal research based on proactively-gathered data has an advantage. So, for example, research based on gathered data enables the explanations for a certain phenomenon to be more easily examined than research based on administrative data. Another advantage of the former is that since longitudinal survey respondents generally consent to the active collection of data (in contrast to administrative data collection in which the state does not seek the citizen's consent), the risk of violating privacy is much lower, as least from the legal standpoint.

3. The use of administrative sources of information to create the infrastructure for longitudinal data in various fields such as education, health, labor and wages is constantly growing in the research community and among decision makers in many countries. Israel, too, has come quite a way in this direction. The team had the impression that in recent years, professionals at the CBS as well as at the Ministry of Education have internalized the importance of making education data available for research purposes and have done much in order to promote convenient and efficient access to these data by researchers. Still, the team has heard from several sources that the waiting time for data and its price are still relatively high particularly because the state does not allocate sufficient funds and personnel positions for making them available. Decreasing the waiting time for data and reducing its cost would likely greatly advance researchers' use of the data.

4. There is place to continue and to consider further expanding this process and increasing cooperation between public entities based on an overall strategy and a systemic view of the needs and benefits, and in line with research and policy making priorities. In the absence of the government's comprehensive strategy it is important for each government ministry, and in our case the Ministry of Education, to set policy and targets for the creation of longitudinal data and to make them accessible to the public. We have the impression that this is the path the Ministry of Health is currently adopting.

5. An interesting insight emerging from Professor David Kaplan's lecture at the seminar is that longitudinal data need not be exclusively at the individual level. Administrative data collected over time about schools or even about countries (such as TIMSS and PISA) can be considered longitudinal data from which we can learn about these very schools or countries.

6. Great importance is attached to the use of administrative data not only as stand-alone longitudinal data, but also within a comprehensive strategic framework of linking longitudinal data with cross-sectional data (from surveys or other administrative sources). Linking of this kind can also help in evaluating the results of intervention programs, including those of

controlled experiments. This strategy has many advantages and has led to its greater use both in Israel and worldwide.[3]

Making Education Administrative Data Accessible in Israel

1. The expert team stressed the role of the CBS both in integrating information and in formulating the priorities for its collection. This also applies to information gathered by other public entities – as indicated in the Statistics Ordinance. There is room to expand the cooperation agreements between the CBS and government ministries, similar to the agreements it signed with the Bank of Israel and the National Insurance Institute. Naturally, it would necessitate modifying the scope of the resources available to the CBS - including resources for preparation of data files for the research community so that they can carry out these steps.
2. The team was of the impression that the government ministries would be able to derive much benefit from systematic collection of data in the areas of their endeavor and from storage of the data in central integrated information systems. These data would undoubtedly serve ministries in making their activity more efficient, and in evaluating and setting policy goals. These same data could also serve the research community and the public at large.
3. Due to the extensive use of outsourcing, government ministries would be hard pressed to gain a complete picture of their own activities unless channels are created through which service providers would regularly transmit full information about their activity to the ministries.
4. The importance of determining policy and of making information available for research purposes requires us to be on a constant search for statistical, technological and legal tools to protect the information and to, as much as possible, reduce the risk of violations of privacy. Included among these is, naturally, anonymizing the information as well as legal and technological-contractual tools, a few of which are presented in this report.
5. The main insight reached by the team was that even if these tools are utilized, there is a risk to the privacy of at least one of the individuals "behind" the data. Nevertheless, by using known methods, it is possible to create a situation where the cost of subsequent identification would be very high in terms of time and resources. Thus, managing known risks is an option – examining the inherent benefit of making information available to the public versus the risk of privacy violation.
6. The team is unanimous in its view regarding the need to build the infrastructure for remote access systems ("virtual research rooms") that will allow researchers access to data from their own offices (as distinct from dedicated research rooms such as those at the CBS and the National Insurance Institute), similar to the solution for researchers proposed by the Ministry of Education.
7. Publication of database catalogs (including the metadata) found in government ministries and the CBS can assist both the organizations themselves as well as the population of

[3] For example, see Dynarski (2014).

potential users. The CBS can also consider preparing a database catalog especially created for researchers for work in the research room and this can also serve other researchers.

8. The team points to a significant shortage of information about several education sub-fields. The lack stems from the fact that part of the information was never collected and another part was collected but is not available. The need for information is called for in light of additional education indicators that should be developed (for more details, see Justman & Bukobza, 2010). Below are some areas in which information gaps exist:
   - Budget allocation to the education system by local authorities, non-profits and parents
   - Activity of local authorities in the education system
   - Activity of non-profit organizations within the education system and outside it
   - Detailed information about non-public preschools (students and teachers)
   - Information about the ultra-Orthodox (Jewish) sector (for example, academic achievement) and in the East Jerusalem (Arab) education system
   - Participation of students and teachers in intervention programs
   - Information about infrastructure in the education system (buildings, computers, etc.)
   - Link between teachers and students at the subject level (key conditions for evaluating instructional quality, inequality in education, etc.)
   - Non-formal education (youth groups, community centers, etc.)
   - "Soft" skills, learning and teaching processes
   - Attitudes, values, etc. required to round out the information on achievements on tests
   - Data collected by the Israel Defense Forces at the time of enlistment and during service

Protecting Privacy and Database Anonymization

1. A dichotomous hierarchy between anonymized information, on the one hand, and information containing identifiable details, on the other hand, does not exist. This insight was presented in Professor Zarsky's lecture, as well as in Adv. Brian Negin's and in Adv. Limor Shmerling's lectures (all given at the learning session held in August). This insight increasingly appears in the literature addressing anonymization (for example: Garfinkel, 2015; HITRUST, 2015; De-Identification Working Group, 2015). There is certainly information which definitely does not allow identification of a specific individual (for example, a file that presents only a summary or statistical attributes of the data), and there is information that definitely allows such identification (for example, a data file that includes ID numbers). Most data files, however, are located in the range between these two extremes.

2. In order to clarify where on this continuum a data file is located, the risk of identification must be weighed. Garfinkel (2015, p. 6) suggests the following range (from lowest risk of privacy violation to highest):

Data that does not concern individuals → data about individuals although with no possibility of identifying any particular individual included in the file → data that can be identified as belonging to one of X individuals (K-anonymity) → data in which a specific individual can be identified → data with an identifying variable.

In his lecture, Prof. Zarsky added that the risk of privacy violation is not influenced only by the nature of the data but also by questions such as:

- What level of expertise is required in order to identify individuals included in a data file?
- Can individuals in the data file be identified if the hacker is in possession of additional information?
- Can this additional information be easily located?
- Can this additional information be easily found, but would involve committing a crime?
- How many people in the data file are likely to be vulnerable to a violation of their privacy?

3. In her lecture at the learning session in August, Adv. Limor Shmerling of the Israeli Law, Technology and Information Authority (ILITA) presented information that can help determine the degree of risk: A table that represents the Israeli regulator's conception of managing the risk in exposing information. It should be mentioned that in addition to questions about the information and about the risk in its exposure, this table also includes a question about the damage that may be caused to individuals as a result of their personal information being exposed:

High Risk to Privacy		Low Risk to Privacy
Many	**Number of people about whom there is information?**	Few
A lot	**How much information is stored about each person?**	A little
High	**What is the sensitivity of the stored information?**	Low
Unlimited	**For how long is the information stored?**	Short time
A lot	**What amount of information is transmitted?**	A little
Many	**Number of external interfaces with the information system?**	No interfaces
Great	**What is the individual risk of harm due to discovery of information?**	Non-existent or low
Strong link to identity	**Is the information strongly linked to identity?**	Weak or no link to identity
Possible	**Can the information be used for other purposes?**	Not possible
Unaware	**Are the people aware of use of the information?**	Aware
No	**Has the information been collected with the peoples' agreement?**	Yes
No	**Is the information being used with the agreement of the people?**	Yes

4. According to Adv. Brian Negin, legal counsel to the CBS, even at the regulatory level, there is no guidance that unequivocally declares that a certain data file is absolutely anonymized. In practice, all the guidelines provided include some degree of judgment granted to the person who determines whether or not the file is sufficiently anonymized. Thus, for example, the European Union standards require entities to engage in "all the relevant measures likely to be used by a third party in order to identify the statistical unit." This guidance does not determine what the relevant measures are, or what constitutes "likely." The guidance in Australia requires removal of information that "can, with high probability, enable identification of this person." Here, too, it is not precisely defined what is meant by "high probability." Similar phrasing appears in the rest of the guidelines.
5. Even if statistical, legal, and technological tools were used to protect the information, there will almost always be more risk of a violation of privacy of at least one of the data subjects. In practice, it is impossible to guarantee that there is no possibility of identifying any one of the data subjects.
6. Nonetheless, by using the tools known in the field, it is possible to create conditions that make the cost of re-identification, in terms of time and resources, very high. Thus, this field should be understood not as one in which there are declarations of what is absolutely permitted or prohibited but rather, as a field in which risks are managed, and assessment is made of whether the risk to privacy (after engaging in all the possible protection steps) is greater than the chance that this data-based research will contribute to the public good.
7. It is important to mention that risk management must be conducted together with assessment of opportunity; the benefit must be taken into account. The possible benefit in general must be estimated, and in our case, the potential benefit to be derived from the research in general, and the specific benefit to be derived from each study, according to the aims and importance of the research proposed. This point is significant since at present, true cooperation does not exist between education researchers and education policymakers with respect to the decision making process of whether and how to allow access to education data in Israel.

Report Structure

In the pages that follow, the main essence of the information collected by the expert team during the activity will be presented. The report will focus on three topics and a separate chapter will be devoted to each. The first chapter will address longitudinal data, longitudinal research, and their importance for research in the field of education and for the design of efficient and effective education policy. The second chapter will examine existing administrative data in Israel in the field of education and access to them for research purposes and education policy management needs, as enabled by the various entities that possess them. The third chapter will examine the legal and regulatory situation in Israel vis à vis safeguarding the privacy of data subjects while making administrative data accessible for research purposes, and will also present different methods and approaches for assessing risk to privacy and to reducing this risk.

Chapter One – Longitudinal Data and Longitudinal Research

As mentioned, the team's activity was concerned with longitudinal data in education and the research options enabled by these data. The first chapter will thus be devoted to this kind of data and to research based on them. In the first part of the chapter, a general introduction to the field will be presented and will include the definitions of longitudinal data, longitudinal research, the differences between cross-sectional data and longitudinal data, the different types of longitudinal data and the research possibilities they provide. Following this, we will present summaries of the lectures given at the seminar on the topic of longitudinal research in which the advantages and possibilities for use of longitudinal data for research needs and policy analysis purposes are illustrated. Each one of the lectures will present the topic from a somewhat different perspective.

Longitudinal Data and Longitudinal Research: A Brief Introduction

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.[4] 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 team's work dealt with and to which we will relate in this report, is data obtained from *"panel studies"* – which measure the same subjects[5] at two or more points in time.

[4] However, cross-sectional data might include information on the memories of past data objects.
[5] 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.

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 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.[6] 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.[7]

[6] The professional term for this is "diachronic analysis."

[7] 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.

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 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.[8]

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

[8] For more information see Shavit & Feniger, 2007 (Hebrew).

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, 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 team 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.[9] 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.

Summary of Lectures Presented at the Seminar on Longitudinal Data and Longitudinal Research[10]

Opportunities and Challenges in Collecting and Analyzing Longitudinal Educational Data

David Kaplan

The aim of this lecture is to illustrate the possibilities and challenges of using longitudinal information for policy analysis. I will talk about longitudinal studies based on the data collected in the Early Childhood Longitudinal Survey (ECLS-K-1998). This is a large scale survey studying a representative sample of the population of American children in the kindergarten cohort of 1998-99 during those years. This survey is, arguably, the most comprehensive longitudinal study to examine the development of academic abilities among elementary to secondary US school children.

The children surveyed came from private schools as well as public schools and included children attending kindergartens for a full-day and for part of a day. They came from diverse socioeconomic and ethnic backgrounds. The parents, the teachers, and the school and kindergarten principals also participated in the survey and provided information about the children's cognitive, social, physical and emotional development. Information was also gathered with respect to the children's home

[9] For more information, see: the European Survey of Health, Ageing and Retirement in Europe (SHARE).

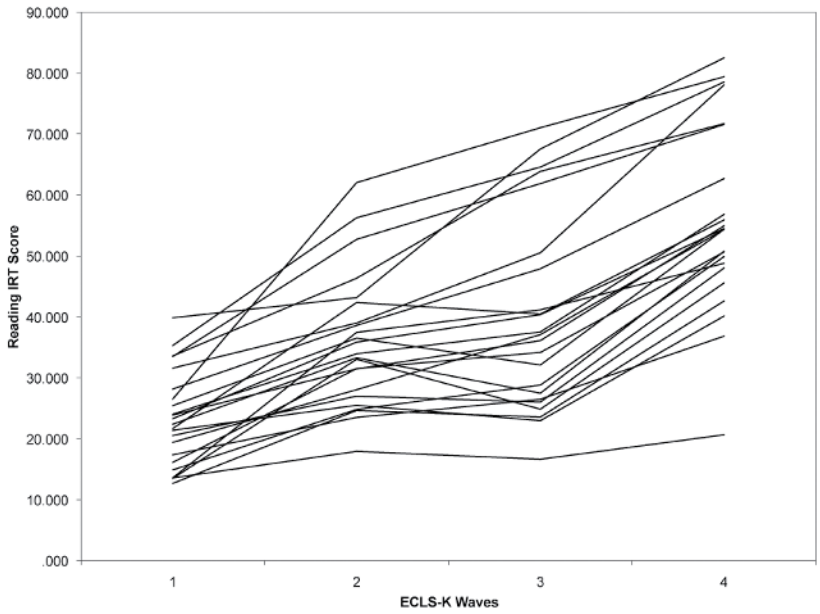
[10] Some of the lecture abstracts do not appear here due to technical reasons. The scientific literature review abstract that appears in the next chapter of this report is essentially a summary of Dr. Amalia Ran's lecture.

environment, educational activities in their home, their school and classroom environments, their curriculum and their teachers' level of training.

As mentioned, the survey tracked the children from kindergarten age to the end of eighth grade. Its objective was to provide reliable and comprehensive information that would enable a better description and understanding of children's development and the experiences they have at those ages. The data are likewise intended to facilitate a better understanding of the relationship between early experiences of kindergarten children and their later development, and their subsequent studies and experiences in school. The data collected through the years enabled researchers and decision-makers to learn how the experiences children undergo at home, in school, in the classroom, and in the community shape their psychological, social and emotional development.

In order to begin to describe the advantages of collecting and using longitudinal data and using suitable statistical research methods and models, we will start with an example: A follow-up study of the data collected on 20 different children who participated in ECLS-K 1998. The children were randomly selected.

The growth curves can be examined and described in two ways: Examining change in outcomes for the same child over time, and examining differences between children in their rates of change over time.



Change in Continuous Variables over Discrete Time

In examining the differences between the children we see that they started off at different points and their abilities developed along different paths over the follow-up period. The children's abilities also reached different levels (even greater than the point they started from) at the end of the follow-up. We have models that describe the average growth rate of children and we can

compare their growth rate and abilities at each one of the points at which abilities were measured. We can create non-linear models of development and take each child's growth rate at each stage of his life into consideration.

After we examined the children's ability at the starting point, the growth in their ability, and their ability at the "end point" (the end of the follow-up), we can go on to the next level and try to understand the differences in each child's growth rate and find variables that explain these differences. Clearly, what especially interests us are the variables that policy can impact and how these variables are related to the children's background data and starting point. We can also look at interventions and examine whether they influence the children's growth curve.

On the third level – all the above can be performed while taking into account the children's learning environment (e.g. classroom and/or school effects). The statistical models available today enable examination of the school's contribution to the children's growth curve as a function of these children's attributes and background.

When we are in possession of sufficiently reliable longitudinal data, statistical models allow us to examine the children's developmental rate and to create complex models to explain it. It is possible to represent grades at the "endpoint" as a function of the status at the beginning and of growth rate. A more complex model can add variables such as gender and examine whether a correlation still exists between the starting point and the rate of growth. Even more can be done: Events that affect measures at each time point (time varying covariate) can be taken into consideration.

But even more can be done: It is possible to analyze two developmental processes occurring simultaneously and examine each one of them separately *as well as jointly in terms of how their growth processes are related*. For example, a study I conducted together with Nancy Jordan (Jordan & Kaplan, 2008) using longitudinal data we collected ourselves looks at whether how children's use of their fingers for counting predicts the accuracy of their count. We studied whether the use of fingers changes over time and whether how they used their fingers affects the accuracy of the count (and therefore, if it was advisable to wean them from using their fingers by a certain age). We found two simultaneously occurring processes: Use of fingers lessened somewhat over time and then at a certain point, suddenly ceased altogether and in contrast, count accuracy increased linearly over time. The use of longitudinal data enables a connection to be made between these two processes and for them to be examined together. Furthermore, it allows them to be combined with other variables we are interested in learning about: Gender differences, differences between children from different socioeconomic levels, and so on. We found that children of lower socioeconomic status (SES) do indeed use their fingers for a longer period of time.

A numerical example: We looked at data from the ECLS-K on 3,575 children. We examined the indicators of their reading level at different points in time, looking at two SES levels, whether they attended full-day or part-day kindergarten, and the age at which they started to attend kindergarten. The research questions were, what is the rate of growth in reading competency among children participating in the study; is there a difference between children attending kindergarten for a full day versus a partial day; and, whether these differences persist when the children's SES and age of entry into kindergarten are taken into account.

First, we found that there is no significant difference between the rates of development of reading skills between children attending full-day or part-day kindergarten. However, when we introduced the variable of SES background into the equation, the picture started to look different. We see a bigger difference, one that reflects a difference between registering for a full or a part day as a function of socioeconomic status and age of entry into kindergarten.

After we looked at the average growth curve, we also found unique patterns of growth trajectories. In the research at hand, division into three distinct groups of reading trajectories could be discerned. In this case, a "one size fits all" curve would miss interesting findings with respect to each one of these three trajectory patterns. It was important for us to know if there are sub-populations with unique growth curves. Interventions must also be adapted to the particular growth curve of each sub-population. Thus, we estimated mixture models of growth (so-called growth mixture modeling) – those that would be appropriate for each sub-population in the full population.

Models such as these are part of a group of general mixture models based on the idea that each research population might be comprised of sub-populations (whose number and size obviously varies), and that models must relate to these sub-populations. This general idea can also be applied to other statistical models. Sub-populations divided according to the relevant theoretical models for each and every research study (corresponding to the variables that theory tells us are likely to influence the model), and according to empirical findings (according to groups we actually found within the population). When applying the idea to models describing growth or development (growth mixture modeling), we divide the population according to the unique shapes of the growth curves.

As mentioned, in the study in question, we found three trajectory patterns: The first and largest group (72% of the children) was comprised of children who started out with relatively low scores in reading but developed nicely over the years. The second group was made up of children who started at a relatively high level and after a period of growth they reached a high, relatively fixed level. The third group was composed of children who started at a relatively low level and relative to the other groups, did not succeed in advancing.

After we derived the three different growth curves it was possible to check whether there were differences between the groups and within each one of the groups. Each one of the groups could be divided into children in part-day kindergarten and those in full-day kindergarten. They can also be divided by SES. When we did so, the research yielded an interesting finding: In the weak growth group (the third group) there was a difference between children attending kindergarten for a full day versus a partial day. This was not the case for the other groups (when age of entry into kindergarten and SES are controlled for, these differences pretty much disappear). In other words, it was found that attending full-day kindergarten can have a positive effect on children who are on a relatively slow growth path. Obviously, what is important is not this specific example but what can be learned from existing methodologies when there are suitable data.

In the example we presented, we saw development focused on "continuous" scores measured over discrete time. Focusing on "categorical" results in which change or development leads to achieving (or not achieving) a certain status is also an option, for example, if a certain ability

was acquired or not, or moving through Piaget's stages of development (Piaget, 1964). These are transitions from one status to another. The ECLS-K allows us to assess such transitions as well. The ECLS-K provides "transformations" of reading scores into probabilities that the child will be able to master specific tasks, as well as dichotomous proficiency scores (can/cannot). This is achieved by creating clusters of questions for the area of reading that are all of similar difficulty level. When the child succeeds in solving three out of four items at a particular difficulty level, he is categorized as possessing the ability relevant to that level.

In the example I will now present (Kaplan & Walpole, 2005), we tried to use these tools in order to track the progress of children. This model is divided into five stages of reading comprehension: letter recognition, beginning sounds, ending sounds, sight words, and words in context. The results allowed us to examine not only the attainment of different levels of expertise by children from different SES groups but also how attaining a certain level of expertise influences development of the following level and how various factors in the child's environment influence acquisition of these skills.[11]

We will now examine one of the central challenges in gathering and utilizing longitudinal data in the field of education for policy analysis purposes – file matching.[12] The question is how can we connect between files from different sources in order to obtain a richer data set for research purposes, and assuming it is possible, how can the quality, validity and reliability of the data be preserved in the new file. For example, we can add information about the OECD member-countries to the information obtained from the PISA tests, in order to study growth in the different countries. Such combinations are referred to as creating synthetic longitudinal data.

An example of synthetic longitudinal data is the combination of ECLS-K data with ECLS-Birth data – a longitudinal survey that tracked children from birth. Longitudinal data can be generated in these cases even if the respondents in each one of the data files are not the same respondents, which is the case in this fusion. In such cases a connection must be made between individuals in the different files based on their background data – this means that we must be in possession of abundant and accurate background information about the respondents. In addition, statistical methods must be used in order to test the validity of the information in the new data file. If there is information from two files – from two different groups of people – a serious problem of missing data is produced – the respondents in one file did not respond to the questions to which the respondents in the second file did. Furthermore, it is important to pay attention to every change taking place in the context (or contexts) – in the environment in which the data were collected especially in cases where the data were gathered at different times. Policy change, economic crises or immigration for example, can greatly affect the data in one of the files thus yielding a less credible fused data file. Finally, the issue of which population the new synthetic file relates

[11] The research results were presented in detail at the seminar however in this summary, we will present them only briefly due to considerations of space.

[12] The second main challenge Professor Kaplan intended to address was privacy protection but since at the seminar (as well as in this report) a separate section was devoted to this challenge, Professor Kaplan chose to skip that portion of his lecture.

to must be rethought. It cannot be taken for granted that the new file reflects one or both of the original populations.

It is important to mention that there are quite a number of statistical methods to handle missing data.[13] Each one of the methods will yield slightly different results and therefore, special attention must be given to examining which method is appropriate to which data file. Next, the criterion for the level of the file's validity must be determined, the degree to which it indeed preserves the data from both the data files of which it is comprised. Rässler (2002) talks about four levels of validity.

- Preserving individual values – the highest level in which we succeed in identifying the answer a particular respondent appearing in the first data file would have given had he responded to the questions in the second file. It is safe to assume that it is not possible to achieve this level of validity
- Preserving joint distributions
- Preserving correlation/covariance structure – very important for drawing statistical conclusions
- Preserving marginal distributions – this is the minimal level needed to draw any research conclusion at all

To examine the existing methods of combining data, we conducted a "proof-of-concept" study fusing data from the TALIS and PISA tests in Finland. The advantage of using data from Finland is that the schools participated in both assessments. What we did was to remove a portion of the information from each of the surveys and combine them as if the schools did not respond to both surveys. Following this, we checked whether the results we obtain through missing data imputation corresponded to the results actually obtained from both surveys. Without going into technical details, we found that at least one of the methods we applied in order to handle missing data worked well in preserving the validity of the data in the context of these TALIS and PISA.

What are the challenges that we have yet to meet with respect to data fusion?

- We have still not addressed multilevel data fusion, for example, data about children on the one hand and data about their schools, on the other
- Sampling weights – there are still methodological questions about the weight of each part of the sample and how to adapt it to the general population so that the sample will remain representative of the entire research population
- And again, temporal concerns related to contextual changes over time, especially when using data for policy evaluation

On the other side of the coin, there are more than a few advantages in such data fusions, as presented earlier in the report.

[13] A partial list of these methods can be found in Professor Kaplan's presentation (in Hebrew) on the Initiative for Applied Education Research's website. The list does not appear here so as to avoid making this lecture summary too technical.

Longitudinal studies are very expensive and it is very difficult to use them to measure all the indices relevant to the research (even if there is rotation among the research questions in different years). Maybe it is best to regard longitudinal data as a part of a system of indicators that attest to the health of a country's education system. These indicators must be relevant to the scientific theories related to the aims of a country's education system, they must facilitate tracking of changes (or examination of planned changes) in education policy, and must be followed in order to confirm that they are indeed changing together with changes in other measurements being collected for the country's education system (to track their validity over time). Indicator systems can be used for policy purposes, including forecasting.

On the other hand, if forecasting and prediction are the goals of the indicators from longitudinal surveys it is important that they not be derived from a single theoretical model but rather from diverse theoretical models that are related to what is being measured. One of the things that can be done is to combine statistical models in order to optimal predictive results and the most accurate forecasts. Over time, these models can be calibrated in line with the accuracy of the predicted outcome.

The Potential of Longitudinal Research in the Field of Education in Israel

Moshe Justman

In his lecture, Professor Moshe Justman demonstrated the value of longitudinal data in two recently conducted research studies – in Israel and Australia. These studies made use of administrative longitudinal data measured over time which, as mentioned, is an alternative to longitudinal surveys. In Israel, much education data are collected though in contrast, there are very few longitudinal surveys and even fewer longitudinal surveys in the education field.

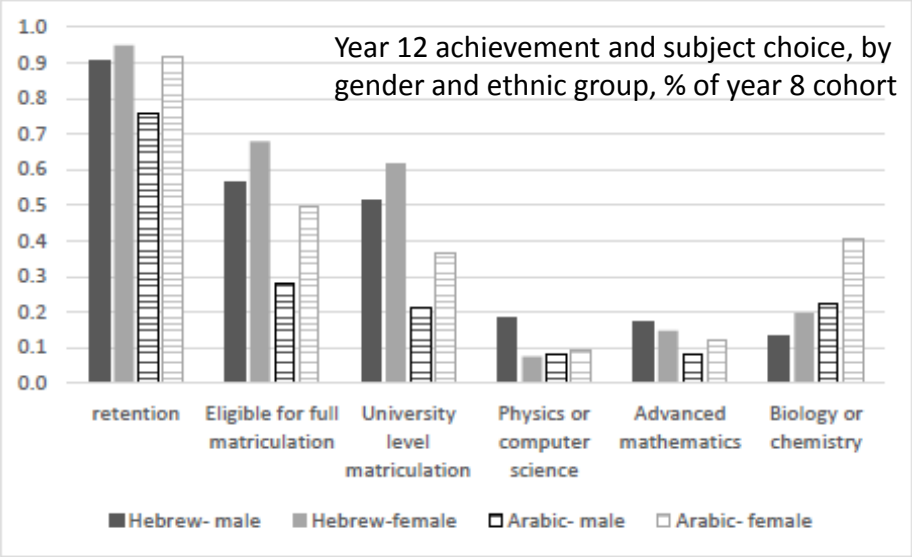
The first example presented by Prof. Justman is a study conducted together with Dr. Naomi Friedman-Sokuler which compared gender gaps in the choice of science subjects, both in the Jewish and Arab sectors. They discovered that there are sizable differences between Jews and Arabs in this area. Longitudinal data enable us to control for ability data (as expressed in the scores on the Meitzav test – "Growth and Effectiveness Measures for Schools"), to compare data for male students and female students possessing the same ability levels and to see the choices they later make in choosing subjects for matriculation.

There are gender differences in the professions men and women choose: Women are less represented in the hi-tech professions and this phenomenon contributes to wage disparities in Israel and around the world. These differences reflect differences in choosing subjects in university and even earlier – on the matriculation examinations. In the past, it was claimed that these differences were the result of gaps in the mathematical abilities between boys and girls however, research studies (including this one) show that this is not the case and that these gaps are maintained even when the mathematical ability of both the genders is controlled for (Riegle-Crumb et al., 2012; Friedman-Sokuler & Justman, 2016; Justman & Mendez, 2016). Other explanations for this phenomenon relate to gaps that stem from the differences in teachers' attitudes towards boys and towards girls, and to the differences in the labor market and in – non-egalitarian – incentives

employers grant women and men for choosing the matriculation subjects they did, at the time. In the current study, the researchers controlled for the respondents' socioeconomic status since it is known that deprivations that develop from these gaps influence boys more than girls. In this study, there was an attempt to discover if there are differences that arise from cultural background. The researchers looked at differences between Arab education and Jewish education with reference to the following variables:

1. Dropout rate
2. Matriculation certificate eligibility
3. Choice of major subject for matriculation

The information: A data file prepared in the CBS research room with longitudinal data obtained by combining Meitzav data (Meitzav 8) with the matriculation data for those same students who were in eighth grade during academic year 2002-3. Obviously, data were missing and statistical methods were used in order to provide full data for students who did not participate in both the tests. Students for whom there was no information about at least one of their parent's education level were omitted. The big advantage of using these data was that there was a very high number of participants. Sampling error and the bias due to dropping out of the study are much smaller than what typically occurs when working with surveys.



Ethnic Difference in Raw Gender Gaps

Initial descriptive findings with reference to persistence and choosing subjects, *mathematical ability not controlled for:*

- Rate of persistence at school – girls are more persistent than boys on all the measures: More girls than boys finish 12 years of study, more girls than boys obtain full matriculation certificates, and more girls than boys attain high-quality matriculation certificates (i.e., the

level required for university admission). An interesting piece of data is that this disparity between boys and girls on all three of the measures is much wider in the Arab sector than in the Jewish sector.

- Choosing subjects for study – in looking at the measure of choosing to study physics and computer science, we see interesting results. In the Jewish sector, there is a large gap in favor of boys – 20% of the boys in this sector chose these subjects, in contrast to 7% of the girls. The surprising part is that in Arab education, the situation is reversed – more Arab girls choose to study physics and computer science than Arab boys. A similar picture is revealed concerning the choice to study five units of mathematics – in the Jewish sector, more boys choose to study math at this advanced level and in the Arab sector, more girls choose math at this level. The surprise here is even greater considering that the Arab sector is conceived of as more traditional and more conservative and for this reason, we would have expected the gender attributes to be more differentiated.
- In choosing to study biology, the gap favors the girls in both the Jewish and Arab sectors though the gap is greater in the Arab sector.

The longitudinal data show that some of the gaps in the Arab sector in choosing subject majors reflect the achievement gap which favors girls, a gap that is already apparent on the Meitzav eighth grade test – big differences in language and not as large, but still significant, differences in mathematics.

At the next stage, the gender gap between Jews and Arabs and male and female students who attained equal results on the Meitzav was compared. Even after controlling for SES and Meitzav results, there was still a difference between Arab girls and Arab boys of over 7% more than the difference seen between boys and girls in Hebrew education. In contrast, with respect to obtaining a higher quality matriculation certificate, there is no difference in the size of the gaps. Here is the added value of longitudinal data – they make it possible to detect the added value of each one of the stages of study.

From the perspective of choosing science subjects – physics and computer science – the difference between the gender gap in the Arab sector and the gender gap in the Hebrew sector is ten percentage points – a very significant difference. In addition, there is also a four percentage point difference in the gender gaps with reference to choosing advanced level math and a nine percentage point difference in choosing to study biology.[14] As mentioned, these facts hold up after taking SES and Meitzav scores into account.

An important point to mention and which is important for the study's implications – the advantage of the Arab girls over the boys is not maintained at the university level. Though this point was not examined in the present study, it is still important to note.

The second example of a longitudinal research study is based on administrative data Professor Justman collected in Australia together with Brendan Houng (Houng & Justman, 2016). The

[14] It is important to emphasize that what is meant by the above is not that the gap between Arab boys and Arab girls is 10% greater than the difference between girls and boys in Hebrew education (for example, 10% vs. 11%) but rather that in absolute terms, it amounts to a 10% difference – i.e., 20% vs. 10%.

researchers created a data file from the results on the NAPLAN (National Assessment Program – Literacy and Numeracy) test, administered to third, fifth, seventh and ninth grade students, and from the twelfth grade matriculation exam results. The researchers also created an SES classification system based on the parents' education level and employment. The data came from the 2009 and 2010 matriculation exams and created a model that, based on the students' outcomes on the NAPLAN and SES, predicts the student's probability of being in the upper half of the median on the distribution of the matriculation exam achievements. This model was tested on the 2010 cohort.

The model was later tested and compared the model's forecasts relative to the students' actual matriculation results. For example, all the children who, according to the model, had a 20% chance of making it into the top 50% were examined to check whether they actually did. It was found that for almost every group (99% .., 60%,...2%, 1%) the forecast was indeed correct. In other words, for each student, if their SES status and MEITZAV results are known, the chances of making it into the higher half of the median of the matriculation exam achievement distribution is also known.

A pessimistic implication seen from the results is that the student's socioeconomic status has an effect on his chances of success. An optimistic implication is that children with high MEITZAV results have a good chance at the matriculation exam regardless of their SES background. The inferences of this model for policy is that with it, the child can receive better counseling and be told what his chances for matriculation achievement are. Obviously, each child has a chance to succeed but a child with low MEITZAV scores in eighth grade has rather low chances of succeeding on the matriculation exams. It may be wise (together with the child and parents) to think about whether the child should invest in the matriculation track or if there is another track that would lead him to success.

Another, no less interesting, option is to examine how the student advances from one test to the next (in third, fifth, seventh and ninth grades) as compared to what is forecast for him. Another option – to examine how students in different schools succeed and whether there is a difference between different schools in the percentage of their students who succeed (as compared to the model's forecasts).

Long-Term Effects of Educational Interventions in Primary and Secondary Schools

Victor Lavy

The aim of the lecture is to demonstrate, using current research studies, the importance of longitudinal data to evaluate the success of interventions, the way such interventions can shape the development of individuals, and the importance of longitudinal studies in assessing policy and its influence. These studies are based on follow-up studies conducted over the shorter-term (three years, five years) and just recently, during the past year, we checked the longer-term results.

In recent years, economists are increasingly attempting to examine interventions and policy not (only) in light of the success of the interventions' and the policies' initial goals, but with respect

to their influence on the long-term wellbeing of individuals – into their adulthood. Is there an improvement in the individual's lifestyle and his contribution to society and his family? There is an attempt to understand the causal (not correlational) relationships between interventions and policy change and the individual's long-term wellbeing. The research examines not only the aims of the intervention but its overall effect on the individual. The range of outcome variables is very broad. There are many studies that focus on the effect of peer quality, class size, school size, and so on, on individuals in the long term.

Together with research colleagues, we used administrative data in order to add a layer of in-depth theoretical study of long-term results of interventions. The data came from the CBS, the Ministry of Education and the National Insurance Institute. The research uses the same databases:

- Health tax file – access to information about every individual for every year of his life (since 1995) and whether he attended any type of educational institution, including a post-secondary institution. Every educational institution is required to transmit information to the National Insurance Institute about its students since they are entitled to a discount in the health tax.
- Income tax file – here we can obtain information about every individual's income, whether or not he worked, and the number of months worked during the year. From these data, conclusions can be drawn about the spouse's income as well as the children's income.
- Census data enables access to information about every individual, his address, and family status as well as the ability to connect his income to his parents' income.

The advantages of the data: They are available and immediate and the cost of acquiring them is relatively low. The disadvantage is that details such as the specific institution the student attended and the degree he studied for cannot be obtained.

The first study – The effect of being able to freely choose the high school on students in the city of Tel Aviv. This program went into effect in 1993-94 and in its framework, students from the ninth district in Tel Aviv (a very weak population) were given the option of choosing their middle school and high school when they graduated from primary school. An earlier study published six years ago checked the short- and medium-term effects of this free choice on the students until they completed their studies. The results showed that there was a very significant positive effect on the rate of their persistence in their studies, on the matriculation eligibility rate, on the number of study units undertaken, and on lesser exposure to violence.

The study we conducted in the past year, and which I would like to now present, checked into whether those short and medium-term changes had (caused) a long term effect on the outcome variables. We obtained a number of interesting results:

- Between 2000 and 2011, there was a steady increase in the percentage of those attending university although there was no difference between those who participated in the program and the control group. Thus, in the long term, the program had no effect. A similar picture is seen with respect to the total number of years studied.

- In contrast, the rate of those attending colleges[15] shows that the program led to an increase of five percentage points or 25% in those studying in colleges, as compared to the control group. This effect was not seen in the initial years but was evident after a few years.
- When examining the average wage of program participants, we see that in the first years the effect was negligible. In the long term, however, there is a positive influence of almost 3,000 NIS more (about 3%-4%) annually.

The second research study examines the long-term effects of a controlled experiment (conducted in cooperation with the Ministry of Education in 2000-01), a program which remunerated teachers based on their students' achievements. The teachers' achievements were examined through their students' achievements on the matriculation tests. The experiment began in January and ended with the matriculation exams in June. Taking part were schools whose level was similar to the average level of all the schools in Israel although the students' average achievements in math and English were relatively weaker.

Participating in the experiment were all the math and English teachers whose students sat for the matriculation exams during that year and a control group of teachers from another school whose students also sat for the exams. In the short term, positive results were found in the percentage of students taking the English and math matriculation exams, in the rate of those passing as opposed to failing, and in the average grade on the matriculation exams.

The research examined the long-term effects on diverse measures: Attending an academic institution, employment, wages, age at marriage and at birth of children, and spouse's educational level. Using National Insurance Institute data, we measured its influence on the receipt of allowances from the National Insurance Institute: Unemployment payments, income supplements, etc. Below are the findings:

- Rate of university attendance: The number of students rose from year to year. In the long term, a positive influence of close to five percentage points (or about 25%) was found in the rate of students attending university, as compared to the control group.
- Number of years at school: It was found that the program raised the number of years at school by 17 or 18 percentage points.
- Wages: Here too, the effect of the program was negative at first since the rate of those attending university was higher. However, in the long run, a significant positive effect was found of close to an additional 6,000 NIS above the average annual wage. It is important to understand that this result can be obtained only through longitudinal data and almost only through administrative data – there is no longitudinal study in Israel that has been in effect for this many years.
- Additional findings: It was found that the program led to delaying the age at marriage by half a year and as a result, this led to a delay in the birth of the first child, both for males and females.

[15] Colleges, as distinct from universities, which generally have somewhat lower admission requirements.

The study points to the importance of using administrative data and of combining administrative data with surveys to engage in longitudinal research. These studies can play a very significant role in policy and in greatly improving our capacity to understand the outcomes of educational policy and intervention programs.

The third research study conducted, this time not as an evaluation study of an intervention program, examined the long-term effects of the results on high-stakes tests – matriculation examinations.

As is well-known, high school students in Israel, within a relatively short period of time, are tested on a relatively high number of matriculation subjects. These tests are of great importance to the students' future, as they influence the study majors that will be open to them when they apply to institutions of higher education and thus, also influence the rest of their lives. The question was whether the use of tests causes random events (an argument with parents, sickness, an inadequate breakfast) that occur to the individual on the day of the test to influence his future development well up to an older age. In this case, a positive answer would perhaps reveal something about the fairness and effectiveness in the use of such tests.

In order to check the influence of a random event and to distinguish between it and the student's ability, we checked episodes in which the individual was exposed to air pollution on the day of the test. Epidemiological studies show that air pollution negatively affects cognitive functioning in those exposed to it (here, we will not go into the reason why – clearly, the article discusses the reasons for this). Using administrative data, we are able to know when each individual took the matriculation exam. The second part of the puzzle was the data from the Environment Ministry which showed the air pollution measures for the different days of the tests. Thus, the way exposure to air pollution affects test results was examined.

Compared to the attainments on other matriculation exams, we were able to see that there is a significant negative causal effect between exposure to air pollution and the matriculation exam results of those exposed to pollution. These results stem mainly from checking the days on which there was a high level of air pollution, for example, the day following the *Lag b'Omer* holiday (celebrated with many bonfires throughout the country). The effect is small with respect to low and medium levels of air pollution but grows significantly where air pollution levels are high.

The more significant part of the study showed that such exposure to air pollution on the day of the test, taken at age 18, has a significant negative influence on various measures in the long term: When those who were tested at age 18 reached age 32, a significant negative correlation was found between attending university and the number of years of schooling, as well as a negative causal influence on wages at age 32.

Countries such as Denmark, Sweden, Norway and the United States are making administrative data available to researchers in a rapid and effective manner while, naturally, protecting the privacy of the data subjects. As a result, we see that most of the research published in leading economic journals come from these countries. The importance of such longitudinal research from the standpoint of the possibilities for assessing policy and producing useful policy is much greater.

We thank the Ministry of Education for use of their virtual laboratory, and to the other organizations that make Israeli education data available. All of the Ministry's administrative files are located in a database with raw data and it is possible to make connections using a fictitious identity for each subject included. The data can be remotely accessed without submitting a request for files. This is a model worthwhile replicating in other ministries as well since it greatly reduces the path to the start of the study. Despite more than a few difficulties, the National Insurance Institute makes administrative files available in its physical research room while protecting privacy and preventing information leakage.

Chapter Two – Longitudinal Administrative Education Data in Israel

Scientific Literature Review Summary: Administrative Databases on Education in Israel[16]

Amalia Ran, Liat Josefsberg Ben-Yehoshua

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. In order to complete this mission, the review focused on 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:

- Main institutions in Israel and the description of their mapped databases, including data on the 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 Applied Education Research by the Israel Academy for 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.

[16] This summary was translated by the authors.

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

- Databases on education that contain variables 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 Applied Education Research by the Israel Academy for Sciences and on behalf of the team of analysts of the Information Center at the 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 Israel Academy of Sciences and the Initiative for Applied Education Research 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, that did not respond to our repeated 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 willingness 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 carrying out this project, for their cooperation and willingness to share their databses. Our gratitude goes also to the team of analysts in the Information Center at the Mofet Institute for their contribution in completing this catalogue.

Bibliography

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.

Friedman-Sokuler, N., & Justman, M. (2016). Gender streaming and prior achievement in high school science and mathematics. *Economics of Education Review*.

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

Gavison. R, "Privacy and the limits of law," *Yale Law Journal 89* (3) (Jan 1980).

HITRUST De-Identification Working Group, (2015), De-Identification Framework, Health alliance Trust web-site, on-line at: https://hitrustalliance.net/documents/de_id/HiTrustDeIdentification Presentation.pdf

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: When do fingers help? *Developmental Science*, 11, 662-668.

Justman, M. and Bukobza, G. (Eds.). Guidelines to Revise the System of Education Indicators in Israel: Summary, Conclusions and Recommendations, By the Committee to Revise the System of Education Indicators in Israel. Israel Academy of Sciences and Humanities. Jerusalem: 2010.

Justman, M., & Méndez, S. J. (2016). Gendered Selection of STEM Subjects for Matriculation.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance to education policy. *Peabody Journal of Education*, 77, 189-215.

Kaplan, D. & Walpole, S. (2005). A stage-sequential model of reading transitions: Evidence from the Early Childhood Longitudinal Study. *Journal of Educational Psychology*, 97, 551-563.

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, 44, 457-467.

Kaplan, D., Kim J-S., & Kim, S-Y. (2009). Multilevel Latent Variable Modeling: Current Research and Recent Developments. In R. E. Millsap and A. Maydeu-Olivares (eds.), *The SAGE Handbook of Quantitative Methods in Psychology*. Newbury Park: SAGE Publications.

Kaplan, D. & McCarty, A. T. (2013). Data fusion with international large-scale assessments: A case study using the OECD PISA and TALIS surveys. *Large-scale Assessments in Education*, 1:6, doi: 10.1186/2196-0739-1-6.

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). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *1*(1), 3.

Piaget, J. (1964). Part I: Cognitive development in children: Piaget development and learning. *Journal of research in science teaching*, *2*(3), 176-186.

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The more things change, the more they stay the same? Prior achievement fails to explain gender inequality in entry into STEM college majors over time. *American Educational Research Journal*, *49*(6), 1048-1073.

קאנט, ע. (1984). *הנחת יסוד למטפיזיקה של המידות*, תרגום: מ. שפי, הוצאת מאגנס, תשמ"ד 1984.

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, י. ופניגר י. (2007), *מחקרי אורך במערכת החינוך הישראלית*, נייר עבודה שהוגש למדען הראשי במשרד החינוך.

Appendices

Appendix B: Brief Bios of the Main Partners to the Learning Process

Expert Team Members

Leah Achdut (Chair), Member of the Economics and Society Program at the Van Leer Institute and head of the department of economics and Management at the Ruppin Academic Center school of economy and business administration. In 2010, she was appointed Associate Professor of Economics and Management at the Ruppin Academic Center. During her years in civil service, she filled management roles in research units at the National Insurance Institute, directed the Histadrut's (Israel Trade Union Organization) Institute for Economic and Social Research, served as the deputy director for research and planning at the National Insurance Institute and was a member of the Institute's executive board. During that period, Professor Achdut headed committees concerned with developing an information and research infrastructure and developing community social projects, was responsible for preparation of the Poverty Report and the National Insurance Institute's Annual Survey, participated in governmental and public committees as a member or observer, and represented Israel at international forums. She was a member of the board on the Luxembourg Income Study and served as the National Insurance Institute's representative on the research committee of the International Social Security Association (ISSA). Since 2004, she has also been a partner in the Survey of Health, Ageing, and Retirement.
Professor Achdut holds a Bachelor's degree in economics and statistics and a Master's degree in economics, both received from the Hebrew University of Jerusalem.

Michal Beller, President of Levinsky College of Education. Her main area of expertise is educational measurement and assessment. Professor Beller was the founding director of the National Authority for Measurement and Evaluation in Education (known as RAMA, its Hebrew acronym) and its director since its establishment in 2005. Prior to founding RAMA, she was senior research director in the Research and Development Department of the Educational Testing Service (ETS) in Princeton, New Jersey, the largest testing institute in the world. Before that, she was an associate professor in the Department of Education and Psychology at the Open University, where she also founded and directed the Shoham Center (integrating technology in learning). For a period of approximately eight years, Beller headed the National Center for Testing and Evaluation (established by the Associated Heads of the Universities in Israel). Professor Beller has published many articles in the area of educational measurement and assessment. She has participated in many international conferences where she has represented Israel, and she serves as a member of professional steering committees in Israel and internationally.
Professor Beller holds a PhD degree in psychology from the Hebrew University of Jerusalem, received in 1983.
Served on the "Educational Measurement and Evaluation" Committee; upon her appointment as executive director of RAMA, she withdrew from the committee.

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. From 2009 to 2012, she served as the Director of the School, and in 2012, she was promoted to Full Professor. Since 2014, Professor Manor has been the Chairwoman of the Board of Directors of the National Institute for Health Policy Research. She has authored more than 150 scientific articles, has received many research grants, and was awarded university prizes for teaching and research. She was a visiting professor at the University of Pennsylvania, the Karolinska Institute in Stockholm and at the University of Geneva. Professor Manor heads the National Program for Quality Indicators in Community Healthcare.
Professor Manor holds a Bachelor's degree in economics and statistics and a Master's degree in statistics, both received from the Hebrew University of Jerusalem, and a PhD degree in statistics, from Stockholm University, received in 1985

Iris Tabak, A senior lecturer in the Department of Education at Ben-Gurion University of the Negev, where she heads the specialization in Curriculum and Instruction. Dr. Tabak is, together with Joshua Radinsky, the editor-in-chief of the Journal of the Learning Sciences, and is a former president of the International Society of the Learning Sciences (ISLS). Her research concerns conceptual knowledge, reasoning skills and professional identity. Specifically, she constructs models of professional vision on whose basis she develops teaching methods and learning technologies. This is the framework within which Dr. Tabak developed the "teacher as partner" approach and the concept of "synergistic scaffolding." With the objective of making science accessible, Dr. Tabak examines how it is possible to enhance the use of online information in order to make evidence-based health-related decisions. These studies take place within a broad methodological framework of design-based research and laboratory research.
Dr. Tabak holds a PhD degree from Northwestern University (US), received in 1999.
Served on the "Language and Literacy" Committee.

Tal Zarsky, An Associate Professor in the Law Faculty at the University of Haifa. His areas of interest are property law, law and technology, telecommunication law, privacy law and e-commerce. His research focuses on the Internet, Information privacy, communications and other topics related to law and technology. He is a member of Israel Internet Association's Infrastructure Steering Committee. Professor Zarsky holds a JSD (2004) from Columbia University (NY, US).

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

Speakers at the Conference (in order of appearance)

David Kaplan, the Patricia Busk Professor of Quantitative Methods in the Department of Educational Psychology at the University of Wisconsin – Madison. Dr. Kaplan holds affiliate appointments in the University of Wisconsin's Department of Population Health Sciences and the Center for Demography and Ecology, and is also an Honorary Research Fellow in the Department of Education at the University of Oxford. He is an elected member of the National Academy of Education, a recipient of the Humboldt Research Award, a Fellow of the American

Psychological Association (Division 5) and was a Jeanne Griffith Fellow at the National Center for Education Statistics.
Dr. Kaplan received his Ph.D. in education from UCLA in 1987.

Moshe Justman, Professor emeritus in the Department of Economics at Ben-Gurion University and former dean of its Faculty of Humanities and Social Sciences. His main areas of research are the political economy of education, equality of opportunity of and access to education, structure of the education system and its sources of funding and, measurement in the field of education. He has also extensively researched topics in industrial-technology policy and issues of regional development. Senior Fellow in the Van Leer Jerusalem Institute since 2007 and coordinator of activities on Civil Society at the Institute since 2012.
Chaired the 'Guidelines for Revising the System of Education Indicators in Israel' committee. From 2011 to 2013, he was a member of the Initiative steering committee.

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. . His main areas of research are Labor Economics, Economics of Education and development Economics. Prof. Lavy is a long time member of the leading economic research intitutes, mainly in England and the US. He has also served as a guest lecturer in leading universities and colleges across the globe. He is currently a member of the American Economic Association, the Econometric Society, the Royal Economic Society and the Society of Labor Economists.
Prof. Lavy holds a Ph.D., Economics from the University of Chicago, 1979.

Amalia Ran, lecturer and researcher in the field of Latin American Studies. Dr. Ran specializes in Latin American culture and literature and in Judaism on that continent, and has published widely on these subjects. She is a member of several academic organization, among them, the Latino-American Studies Association and the Latin American Jewish Studies Association, in which she serves as a member of the board of directors. Dr. Ran was a faculty member at the University of Nebraska and served as a Jewish Agency emissary for Israeli advocacy at the University of Maryland. In 2009, she returned to Israel as part of the Ministry of Absorption's Brain Gain program for returning Israeli scientists and researchers. In 2014, she established a business providing literary and academic translation and editing services in Spanish, English and Hebrew. Dr. Ran is also a lecturer and a workshop moderator on topics that include business culture, literature, and cinema, Judaism in Latin America, Israel-Diaspora relations, women and culture.
Dr. Ran received her BA and MA degrees from the Hebrew University of Jerusalem and in 2007, her PhD from the University of Maryland (US).

Aline Attias, director of the Big Data platform and anonymization at the Ministry of Health's Digital Health and Computing Division. From 1995 to 2014, she was involved in consulting, establishing and managing large-scale data storage centers and business intelligence infrastructures, mainly in the private sector. In her last position, she directed the Business Intelligence and Data Infrastructure Department for the Phoenix Insurance Co.
Ms. Attias holds a BS degree in information systems engineering, received from the Technion and an MA in management from Boston University (US), received in 2000.

Yoel Finkel, Associate National Statistician at the Central Bureau of Statistics (CBS). At the start of his career at the CBS, he was an economist in the Consumer Price Index Unit, and has since advanced from one position to the next on the CBS ladder: deputy director of the unit, unit director, division director, deputy director of the CBS, and currently, Associate National Statistician. As Associate National Statistician, Finkel serves as a senior partner in leading the CBS to its esteemed professional status in Israel and abroad. In enhancing the senior ranks of civil service, in joint learning processes and in creating work relationships between people at this rank, Finkel sees a necessary condition for implementing real reform in civil service and improving the service to benefit the public in the State of Israel.
Mr. Finkel holds a BA degree in economics and business administration and an MBA, both received from the Hebrew University of Jerusalem, and an MPA in public administration from Harvard University, received as a Wexner Fellow in 1994.

Yuval Eylon, Staff member in the Department of History, Philosophy, and Judaic Studies in The Open University of Israel. Dr. Eylon has published articles in the fields of Ethics, Ethics and law, And Meta ethics. His research intretst include Political philosophy, Philosophy of the law, And Philosophy of sports.
Dr. Eylon holds a PhD from the Hebrew university in Jerusalem, 2003.

Raanan Sulitzeanu-Kenan, senior lecturer in political science and public policy at the Hebrew University of Jerusalem. His areas of research are cognitive aspects of political and public decisions made by citizens and public officials (elected representatives, officials, and judges). His current research concerns understanding the effect of proportionality on public policy, responsibility judgments, and the effect of reputation on institutional cooperation. Professor Sulitzeanu-Kenan has been a member of the American Political Science Association since 2006 and a member of the Israeli Bar Association since 1999.
Professor Sulitzeanu-Kenan holds an LLB degree in law from the University of Haifa and a DPhil degree in political science from Oxford University (UK), received in 2006.

Limor Shmerling-Magazanik, expert in regulation and enforcement of privacy protection of personal information stored in computerized databases. Attorney Shmerling Magazanik is the director of Department of Public Affairs and Government at the Ministry of Justice's Israeli Law and Information Technology Authority (ILITA) and was previously the director of the Ministry's Licensing and Supervision Department. Since 2006, when she joined ILITA, she has been involved in directing the administrative enforcement and regulation under the Privacy Protection Law, the Credit Information Services Law, and the Electronic Signature Law. She has been involved in computer investigations, in integrating principles of information security and cyber protection, in proper management of databases and in processes of imposing administrative sanctions on topics such as proximity of purpose of use of information, consent, and information security. Representing ILITA, Attorney Shmerling Magazanik participates in public committees in areas related to privacy protection such as the Advisory Committee Supervising the Biometric Database Test Period, committees on making databases available to the public and transmission of information between public bodies, information infrastructures, cloud computing, government

networks, and a credit information database. In 2008, she served as the secretary of the Council for Cable and Satellite Broadcasting at the Ministry of Communications, and as advisor to the chairperson. Prior to entering public service, Attorney Shmerling Magazanik was involved in counseling and legal representation in the areas of corporate law, contracts, property and banking, and held product management and project management positions in the high-tech industry.
Attorney Shmerling-Magazanik holds an LLB degree in law, an MA degree in public law, and an MA degree in literature, all received from Tel-Aviv University.

Names and Positions of the Participants in the Learning Meetings (in alphabetical order)

Adv. Sima Adiv, lawyer at Legal Department, MOE.

Prof. Audrey Addi-Raccah, Head, Executive MA Program in Educational Management and Leadership Dep. of Educational Policy and Administration Constantiner School of Education Tel-Aviv University.

Dr. Itay Asher, Educational experiments and research officer and acting Chief Scientist at the MOE.

Prof. Rami Benbenishty, Louis and Gaby Weisfeld School of Social Work Bar Ilan University, Israel.

Efrat Degani-Toperoff, content development coordinator for education at Yad Hanadiv.

Nurit Dobrin, manager of the CBS Social Survey.

Talal Dolev, Director, Government Program for Integration of Citizens of Ethiopian Origins, former Director, National Program for Children and Youth at Risk.

Yoel Finkel, deputy government statistician at the CBS.

Gal Fisher, Education program Director, Yad Hanadiv.

Prof. Adam Gamoran, President, The William T. Grant Foundation.

Dr. Chaim Gatt, director of research and statistics in the Ministry of Education's IT division.

Yosef Gidanian, Head of the Education Section at the CBS.

Dr. Hagit Glickman, Director General the National Authority for Measurement and Evaluation in Education (RAMA).

Dr. Tammy Harel Ben Shahar, lecturer in the Faculty of Law at the University of Haifa and director of the faculty's Law and Social Change Clinics.

Adv. Ruti Horn, former legal counsel to the National Insurance Institute.

Avner Kantor, research fellow at the Internet Research Center at the University of Haifa, leads the "Open PISA" project to make PISA test data accessible and researchable (openpisa.org).

Dr. Gilmor Keshet, director of the Sciences Division at the MOE's Pedagogic Secretariat.

Pinhas Klein, economist and senior coordinator (statistical processing and analysis) at the MOE's Economics and Statistics Division, part of the Economics and Budgets Administration.

Aliza Krovi, head of the ISOPED System and the Education Registrar at the CBS.

Hila Lancry Miost of the Senior Division for Planning and Strategy at the MOE.

Ido Levita of the Senior Division for Planning and Strategy at the MOE.

Oriya Levy, Senior coordinator in the field of the long term survey, CBS.

Nurit Lipshtat, director of the Longitudinal Research area at the National Authority for Measurement and Evaluation in Education.

David Maagan, heads the Statistics of Higher Education and Teaching Personnel sector at the CBS.

Adnan Mansour, director of the Learning Transitions and Follow-up sector at the CBS.

Dr. Sofia Mintz, former head of the Computing and Information Systems Administration at the MOE.

Adv. Dorit Morag, Legal Advisor, MOE.

Adv. Zivit Mozer, responsible for Public Affairs and Government at the Ministry of Justice's Israeli Law and Information Technology Authority.

Adv. Brian Negin, legal advisor, CBS.

Ziv Ofek, founder and executive director of CDI-Negev.

Rene Osizon, The Israeli Ministry of Education, Department of Research.

Miri Sabag Endweld, head of the National Insurance Institute's Economics Research Division, part of the Research and Planning Administration.

Dr. Analia Schlosser, The Eitan Berglas School of Economics Tel Aviv University.

Nardit Stein-Kapach, longitudinal survey coordinator at the CBS.

Michal Tabibian-Mizrahi, Director for Strategy and Planning at the MOE.

Yuval Vurgan, Head of researchers' team, one of three multi-disciplinary teams, at the Knesset (Israeli Parliament) Research and Information Center.

Eti Weissblei, researcher, the Knesset's Research and Information Center.

Assaf Weininger, researcher, the Knesset's Research and Information Center.

Dr. Ran Wolff, researcher at the Yahoo Research Laboratory in Haifa.