



---

**דוח מפעילות צוות המומחים בנושא**

---

**שימוש בנתונים הנמדדים לאורך זמן**

**כמקור מידע**

**למחקר, למדיניות ולתכניות חינוכיות בישראל**

**ריכז וערך: עודד בושריאן**

**תשע"ז, 2017**

**היזמה למחקר יישומי בחינוך**

**האקדמיה הלאומית הישראלית למדעים**

ירושלים, תשע"ז, 2017  
היזמה למחקר יישומי בחינוך  
האקדמיה הלאומית הישראלית למדעים

עריכת לשון: עדה פלדור  
הפקה והגהות: אילה ולודבסקי-יובל  
עיצוב: אסתי ביהם  
עיצוב כריכה: עדי מנדלר

מסת"ב: 978-965-7601-10-5

הדוח רואה אור בהוצאת האקדמיה הלאומית הישראלית למדעים, ומונגש לציבור בכתובת  
<http://education.academy.ac.il>  
בכפוף לרשיון ייחוס – שימוש לא מסחרי – שיתוף זהה (CC-BY-NC-SA) לא כולל זכויות תרגום.

בכל שימוש בדוח או ציטוט ממנו יש לאזכר את המקור כדלקמן:  
בושריאן, ע' (עורך). (תשע"ז, 2017). שימוש בנתונים הנמדדים לאורך זמן כמקור מידע למחקר, למדיניות ולתכניות  
חינוכיות בישראל, דוח מפעילות לימודית. היזמה למחקר יישומי בחינוך, האקדמיה הלאומית הישראלית למדעים.

**האקדמיה הלאומית הישראלית למדעים** נוסדה בשנת 1959. חברים בה כמאה מדענים ומלומדים מן השורה הראשונה של ההשכלה והדעת במדינת ישראל. חוק האקדמיה הלאומית הישראלית למדעים, התשכ"א-1961, קובע כי מטרתיה ומשימותיה העיקריות הן לטפח ולקדם פעילות מדעית, לייעץ לממשלה בענייני מחקר ותכנון מדעי בעלי חשיבות לאומית, לקיים התקשרויות עם גופים מקבילים בחוץ לארץ, לייצג את המדע הישראלי בגופים ובאירועים מדעיים בינלאומיים ולהוציא לאור כתבים שיש בהם כדי לקדם את המדע.

**היזמה למחקר יישומי בחינוך** מעמידה לרשות מקבלי החלטות בחינוך ידע מדעי עדכני ומבוקר. ידע כזה חיוני לגיבוש מושכל של מדיניות ולתכנון מיטבי של התערבויות – לשיפור הישגי החינוך בישראל.

חזון היזמה: ידע מחקרי הוא מרכיב חיוני בתכנון מדיניות ציבורית או התערבות רחבת היקף. בשלב התכנון ידע מחקרי מבוקר תומך בגיבוש מדיניות שסיכויי הצלחתה טובים יותר, ולאחר מכן הוא זה שמאפשר דיון ציבורי רציונלי. היזמה מיישמת חזון זה בתחום החינוך.

דרכי הפעולה של היזמה: היזמה עוסקת בנושאים שמעלים מקבלי החלטות, ומתייעצת עם בכירים במשרד החינוך ועם בעלי עניין נוספים. ועדת ההיגוי של היזמה, שאותה ממנה נשיא האקדמיה הלאומית, אחראית על תכנית העבודה של היזמה ועל תהליכי השיפוט של התוצרים.

היזמה פועלת באמצעות ועדות מומחים ועל ידי זימון מפגשים לימודיים משותפים לחוקרים, לאנשי מקצוע ולמקבלי החלטות. היא מפרסמת את תוצרי עבודתה ומנגישה אותם לציבור הרחב. חברי ועדות המומחים מכהנים בהן בהתנדבות.

תולדות היזמה: היזמה הוקמה בשלהי שנת 2003 – מיזם משותף של האקדמיה הלאומית הישראלית למדעים, משרד החינוך ויד הנדיב. מתחילת שנת 2010 היזמה פועלת כיחידה של האקדמיה הלאומית הישראלית למדעים.

בקיץ תש"ע (2010) הוסיפה הכנסת תיקון לחוק האקדמיה הלאומית הישראלית למדעים, והסדירה את אופן ההתקשרות בין משרדי ממשלה המבקשים ייעוץ – לבין האקדמיה. היזמה היא המנהלת את פעולות הייעוץ שהאקדמיה נותנת בתחום החינוך לממשלה ולרשויות שונות.

## תודות<sup>1</sup>

הדוח שלפניכם הוא פרי עבודתו של צוות מומחים שהתכנס על מנת לקדם את הטוב הציבורי בישראל. מרכז הצוות, שגם ערך את הדוח, מבקש להודות בראש ובראשונה לחברי צוות המומחים על עבודתם המסורה, על זמינותם ועל מוכנותם לתרום מזמנם ללא תמורה. ברצוני להודות ליו"ר צוות המומחים פרופ' לאה אחדות, חברה בצוות התכנית לכלכלה וחברה במכון ון ליר בירושלים ומרצה בכירה בבית הספר למדעי החברה והניהול במרכז האקדמי רופין, שניהלה את עבודת צוות המומחים ברגישות וביעילות. אני מבקש להודות גם לכל אחד ואחת מחברי צוות המומחים: לפרופ' מיכל בלר, נשיאת מכללת לוינסקי לחינוך, לנעם זוסמן, כלכלן בחטיבת המחקר של בנק ישראל, לפרופ' טל ז'רסקי, פרופסור חבר בפקולטה למשפטים באוניברסיטת חיפה, לד"ר איריס טבק, מרצה בכירה במחלקה לחינוך באוניברסיטת בן-גוריון בנגב, ראש המסלול לתכניות לימודים והוראה ולפרופ' אורלי מנור חברת סגל בבית הספר לבריאות הציבור ורפואה חברתית של האוניברסיטה העברית והדסה.

מתווה הפעילות ודרך הלמידה המרכזית בה היו מפגשים לימודיים בהשתתפות חוקרים העוסקים בנושאים שהפעילות נסבה עליהם, ובהשתתפות קובעי מדיניות ובעלי תפקידים בגופים ציבוריים שקיים אצלם ידע וניסיון בנושאים הללו. צוות המומחים מבקש להודות לדוברים ולמרצים במפגשים הלימודיים השונים על מוכנותם לתרום מזמנם, מהידע שלהם ומניסיונם להצלחת הפעילות. אני מודה לכן לאנשי משרד החינוך, הלשכה המרכזית לסטטיסטיקה וראמ"ה – שלושת הגופים הציבוריים המרכזיים שמחזיקים ומשתמשים בנתוני חינוך: לד"ר איתי אשר, ממונה ניסויים ומחקרים חינוכיים וממלא מקום המדען הראשי במשרד החינוך, שהציג בפנינו חלק מהאתגרים העומדים בפני אנשי משרד החינוך בבואם להשתמש בנתונים הנאספים על ידי המשרד, או לחלוק נתונים אלו עם חוקרים. ליוסי גידניאן, מנהל אגף חינוך, השכלה וכוחות הוראה בלשכה המרכזית לסטטיסטיקה, שהציג בפנינו את הדרך הארוכה שעברה הלשכה המרכזית לסטטיסטיקה מאז שנות ה-80 בנושא איסוף הנתונים על חינוך והשכלה בישראל והשימוש בהם. לד"ר חגית גליקמן, מנכ"לית הרשות הארצית למדידה והערכה בחינוך (ראמ"ה), שדיברה על השימושים שעושים אנשי ראמ"ה בנתוני אורך כדי לקדם את החינוך בישראל. תודה לד"ר חיים גת, מנהל מערכות מידע והאחראי על המחקר והסטטיסטיקה במינהל התקשוב של משרד החינוך, שהציג את חדר המחקר במשרד החינוך ואת האפשרויות שהוא פותח בפני חוקרים. לנורית דוברין, מנהלת סקרים חברתיים בלשכה המרכזית לסטטיסטיקה, לאוריה לוי מרכזת מקצועית בכירה בתחום הסקר ארוך הטווח הנערך בלשכה המרכזית לסטטיסטיקה, ולנרדיט שטיין-קאפח, מרכזת סקר ארוך טווח בלשכה המרכזית לסטטיסטיקה – שלושתן הציגו יחד את הסקר ארוך הטווח של הלמ"ס, את האפשרויות למחקר שהוא מספק ואת הקשיים העומדים בפני הלמ"ס בעריכתו. לד"ר מיכל טביביאן-מזרחי, סמנכ"לית אסטרטגיה ותכנון במשרד החינוך, שדיברה על הנתונים הדרושים לצורך תכנון אסטרטגי במשרד החינוך. עו"ד דורית מורג, היועצת המשפטית של משרד החינוך ענתה על שאלות המשתתפים בנוגע לעמדת הייעוץ המשפטי במשרד החינוך בנושא העברת נתונים ושיתופם עם חוקרים וגופים ציבוריים אחרים. עדנאן מנסור, ראש תחום מעקבים ומהלכי לימודים בלשכה המרכזית לסטטיסטיקה עדכן את הצוות בתהליך של קבלת קובץ נתונים מהלמ"ס. דוד מעגן, מנהל תחום סטטיסטיקה של חינוך גבוה וכוחות הוראה בלשכה המרכזית לסטטיסטיקה, דיבר איתנו על השימוש שעושה הלמ"ס בנתונים מינהליים לצורך מחקרי אורך. בריאן ניגן, היועץ המשפטי וממונה חופש המידע של הלשכה המרכזית לסטטיסטיקה (למ"ס), הציג את ניסיונו המעשי בנושא מידע, פרטיות וסודיות. יואל פינקל, המשנה לסטטיסטיקן הלאומי בלשכה המרכזית לסטטיסטיקה (למ"ס), סיפר לנו על פעולותיה של

<sup>1</sup> השמות מופיעים לפי סדר האלפבית, והשייכים האקדמיים והמוסדיים הם אלה שהיו תקפים בזמן הפעילות.

הלמ"ס ועל שיתופי הפעולה שהיא מייצרת על מנת לממש חזון של סטטיסטיקה לאומית מאוחדת ומונגשת לצורכי מחקר, וכן הגיב ביום העיון במושב שהוקדש לנתונים מינהליים והנגשתם לצורכי מחקר. פנחס קליין, כלכלן ומרצה בכיר (עיבוד וניתוח סטטיסטי) באגף כלכלה וסטטיסטיקה, מינהל כלכלה ותקציבים במשרד החינוך, חלק איתנו מידע על מערכת השקיפות התקציבית שמנגיש המינהל לקהל הרחב ולחוקרים. עליזה קרובי, ראש תחום איזופד ומרשם השכלה באגף חינוך והשכלה בלשכה המרכזית לסטטיסטיקה הציגה את מערכת האיזופד שמפתחים בלשכה.

שניים בסדר התודות (ולא בסדר החשיבות) – הדוברים והמציגים במפגשים שהגיעו ממשרדי ממשלה אחרים או מגופים פרטיים: תודה לזיו אופק, המייסד והמנכ"ל של CDI-נגב, שהציג בפנינו דרכים חדשניות להנגשת נתונים תוך הגנה על פרטיות האנשים שמאחוריהם. לטלל דולב, המקימה והמנהלת לשעבר של 360°, התכנית הלאומית לילדים ולנוער בסיכון, שדיברה על הדרך שבה התכנית מנחילה תרבות של שימוש בנתונים לאנשים מהשדה החינוכי והטיפול בישראל. טובה מיוחדת מכיר הצוות לפרופ' אדם גמראן (Adam Gamoran), נשיא קרן William T. Grant Foundation, שחלק עמנו תוצאות מעבודתה של ועדת מומחים של האקדמיה האמריקאית הלאומית לחינוך, ועדה שבראשה עמד, ועל הניסיונות לקדם שימוש בנתונים מינהליים לצורכי מחקר בארצות הברית. ד"ר רן וולף, ממעבדת המחקר של יאהו בחיפה, הציג ניתוח סוציולוגי של המושג הפרטיות, ודיבר על הדרך שבה מתייחסות חברות טכנולוגיות פרטיות גדולות, המחזיקות נתונים רבים על אודות משתמשיהן, לנושא השמירה על הפרטיות.

הצוות מודה מקרב לב לדוברים ביום העיון שסיכם את שנת פעילותו של הצוות: לפרופ' דייוויד קפלן (David Kaplan), העומד בראש הקתדרה על שם פטרישה באסק (P. Busk) לשיטות כמותניות במחלקה לפסיכולוגיה חינוכית של אוניברסיטת ויסקונסין-מדיסון. לפרופ' משה יוסטמן, פרופסור אמריטוס במחלקה לכלכלה באוניברסיטת בן גוריון, לפרופ' ויקטור לביא, פרופסור לכלכלה באוניברסיטה העברית בירושלים, מופקד הקתדרה ע"ש וויליאם הבר ופרופסור במחלקה לכלכלה באוניברסיטת ווירג'יניה, אנגליה. לד"ר עמליה רן, חוקרת במכון מופ"ת ומרצה וחוקרת באכסניות שונות בתחום לימודי אמריקה הלטינית. לאלין אטיאס, מנהלת מערך Big Data והתממה במשרד הבריאות, האגף לבריאות דיגיטלית ומיחשוב. לד"ר יובל אילון, חבר סגל במחלקה להיסטוריה, פילוסופיה ולימודי יהדות באוניברסיטה הפתוחה. לעו"ד לימור שמרלינג מגזניק, מנהלת המחלקה לקשרי ציבור וממשל ברשות למשפט טכנולוגיה ומידע (רמו"ט) במשרד המשפטים.

מלבד הדוברים במפגשים הלימודיים השתתפו בהם גם מוזמנים שלא הרצו אך הגיבו והשתתפו בשיחה שהתפתחה סביב ההרצאות. גם למשתתפים אלו אנו מודים על תרומתם: לעו"ד סימה אדיב מהייעוץ המשפטי של משרד החינוך, לפרופ' אודרי אדירקח מבית הספר לחינוך, אוניברסיטת תל-אביב, לרנה אוסיזון מלשכת המדען הראשי במשרד החינוך, לפרופ' רון אסטור מבית הספר לעבודה סוציאלית ובית הספר לחינוך באוניברסיטה של דרום קליפורניה, לפרופ' מיכאל בירנהק, מרצה בפקולטה למשפטים באוניברסיטת תל אביב, לפרופ' רמי בנבנישתי מבית הספר לעבודה סוציאלית באוניברסיטת בר-אילן, אפרת דגני-טופורוב מיד הנדיב, יובל וורגן, ראש תחום חינוך במינהל המחקר והמידע של הכנסת, אתי וייסבלי, ממרכז המחקר והמידע של הכנסת, אסף וינינגר, גם הוא ממרכז המחקר והמידע של הכנסת, לנורית ליפשוט מנהלת תחום מחקרי אורך בראמ"ה, לעו"ד זיוית מוזר, ממונה (קשרי ציבור וממשל) ברשות למשפט, טכנולוגיה ומידע במשרד המשפטים, לד"ר תמי הראל בן שחר, מרצה בפקולטה למשפטים באוניברסיטת חיפה ומנהלת הקליניקות למשפט ושינוי חברתי בפקולטה, לעו"ד רותי הורן לשעבר היועצת המשפטית של הביטוח הלאומי, למירי סבג-אנדבלד ראש אגף מחקר כלכלי, מינהל המחקר והתכנון, המוסד לביטוח לאומי. תודתנו שלוחה גם לעידו לויטה, מהאגף הבכיר לתכנון ואסטרטגיה במשרד החינוך, הילה לנקרי-מיוסט, מהאגף הבכיר לתכנון ואסטרטגיה במשרד החינוך, סופיה מינץ, לשעבר ראש מינהל תקשוב ומערכות מידע במשרד החינוך, גל פישר מנהל תחום חינוך ביד הנדיב, לאבנר קנטור, עמית מחקר במרכז לחקר האינטרנט באוניברסיטת

חיפה, מוביל פרויקט 'פיזה פתוח' להנגשה וניתוח של מבחני פיזה ([opisa.org](http://opisa.org)), עליזה קרובי, ראש תחום איזופד ומרשם השכלה בלמ"ס, ד"ר גיל מור קשת-מאור, מנהלת אגף מדעים במזכירות הפדגוגית במשרד החינוך, ד"ר אנליה שלוסר מבית הספר לכלכלה באוניברסיטת תל-אביב ומזי שרטר, מהאגף הבכיר לתכנון ואסטרטגיה במשרד החינוך.

הצוות מודה לכותבות הסקירה המדעית בנושא מיפוי מאגרי מידע בישראל בתחום החינוך – ד"ר עמליה רן וד"ר ליאת יוספסברג בן-יהושע, ממרכז המידע במכון מופ"ת – שהקדישו שעות רבות לאיסוף המידע, כתיבתו ועריכתו והגישו סקירה מקיפה ואיכותית.

צוות המומחים מודה למנהלת היזמה למחקר יישומי בחינוך, ד"ר אביטל דרמון, ולצוות היזמה – אורי חרוש, אילה ולודבסקי-יובל ועדה פלדור, שמאמציהם, זמנם וכישרונותיהם היו חיוניים ליצירת דוח זה ולעבודת הצוות בכלל.

הדוח עבר הליך מקובל של שיפוט עמיתים בלתי תלוי. עורך הדוח וצוות המומחים מודים על השיפוט, שסייע להבטיח את בהירותו, את איכותו ואת איתנותו. האחריות לתוכנו של מסמך זה מוטלת על עורך המסמך.

### **חברי צוות המומחים:**

פרופ' לאה אחדות – המרכז האקדמי רופין ומכון ון ליר בירושלים (יו"ר)

פרופ' מיכל בלר – מכללת לוינסקי לחינוך

מר נעם זוסמן – בנק ישראל

פרופ' טל ז'רסקי – אוניברסיטת חיפה

פרופ' איריס טבק – אוניברסיטת בן-גוריון בנגב

פרופ' אורלי מנור – האוניברסיטה העברית והדסה

מרכז הפעילות: עודד בושריאן.

## תוכן העניינים

<b>1</b> .....	<b>מבוא וסיכום</b>
1.....	הרקע להקמת צוות המומחים ופעילותו
2.....	סיכום: תובנות מפעילות צוות המומחים
2.....	נתוני אורך, מחקרי אורך והשימוש בהם לגיבוש מדיניות חינוך בישראל ובחינתה
3.....	נתונים מינהליים הנאספים בישראל והגשתם לצורכי מחקר
4.....	שמירה על פרטיות והתממת מאגרי מידע
6.....	מבנה הדוח
<b>7</b> .....	<b>פרק ראשון - נתוני אורך ומחקרי אורך</b>
7.....	נתוני אורך ומחקרי אורך: מבוא קצר
10.....	סיכומי הרצאות מיום עיון בנושא נתוני אורך ומחקרי אורך
10.....	הזדמנויות ואתגרים באיסוף וניתוח של נתוני אורך בחינוך - פרופ' דיוויד קפלן
15.....	הפוטנציאל של מחקרי אורך בתחום החינוך בישראל - פרופ' משה יוסטמן
18.....	השפעות בטווח הארוך של התערבויות חינוכיות בבתי ספר יסודיים ותיכונים בישראל - פרופ' ויקטור לביא
<b>21</b> .....	<b>פרק שני - נתונים מינהליים הנמדדים לאורך זמן בנושא חינוך בישראל</b>
21.....	סקירת מצב: הנגשת נתונים מינהליים בחינוך על ידי הלמ"ס ומשרד החינוך
23.....	תקציר סקירה מדעית: מיפוי מאגרי מידע בישראל בתחום החינוך - ד"ר עמליה רן וד"ר ליאת יוספסברג בן־יהושע
<b>27</b> .....	<b>פרק שלישי - פרטיות, התממה, וניהול סיכונים</b>
27.....	הצדקות תאורטיות לערכה של הפרטיות
28.....	תאוריות סוציולוגיות על פרטיות
29.....	כללים והנחיות רגולטוריות בנושא פרטיות: הדין בישראל
30.....	תפיסת הרגולטור בישראל
31.....	התממה ואנונימיזציה
33.....	שיטות להגנה על הפרטיות ולהתממת קובצי מידע
<b>37</b> .....	<b>רשימת מקורות</b>
<b>39</b> .....	<b>נספחים</b>
39.....	נספח א': סדר יום של המפגשים הלימודיים
44.....	נספח ב': תקצירי קורות החיים של השותפים העיקריים למהלך הלימודי
44.....	חברי צוות המומחים
45.....	הדוברים ביום העיון
47.....	הדוברים במפגשים הלימודיים
50.....	צוות היזמה



## מבוא וסיכום

### הרקע להקמת צוות המומחים ופעילותו

מערכת החינוך בישראל, כמו גם במדינות אחרות, היא גורם מרכזי באספקת שירותים ציבוריים, עמוד תווך בתהליך בניית ההון האנושי של האוכלוסייה, ויש לה השפעה מכרעת על רווחת האנשים ועל צמיחת המשק. מערכות חינוך ניצבות בפני אתגרים מורכבים – לרבות קידום שוויון הזדמנויות והענקת שירותים בהיקף הולך וגדל ובאיכות גבוהה – וכל זאת תחת מגבלת משאבים תמידית.

לצורך הערכה של פעילות מערכת החינוך ושיפורה המתמשך נדרשים מסדי נתונים מקיפים ואיכותיים. בשנים האחרונות גברה בישראל המודעות ליתרונות הגלומים בנתוני אורך כבסיס לקביעת מדיניות, להערכת תכניות התערבות ולמחקר באיכות גבוהה העוסק בהיבטים שונים של מערכת החינוך. במקביל גבר גם הביקוש לנתוני אורך והתרחב השימוש בהם לצד השימוש המסורתי בנתוני חתך. כל זאת מתוך הבנה שרכישת השכלה היא תהליך רב-ממדי ומתמשך, ושתהליך זה מושפע מהחלטות המתקבלות בנקודות זמן שונות על ידי אנשי חינוך, הורים, תלמידים ואחרים. מעקב על פני זמן אחר אותם פרטים עשוי להעמיק את ההבנה של תהליכים הקשורים ברכישת השכלה ובהזדמנויות נוספות לחינוך.

היזמה למחקר יישומי בחינוך הקימה בתמיכת יד הנדיב ובעידודה צוות מומחים שיוביל תהליך למידה בנושא השימוש בנתונים הנמדדים לאורך זמן כמקור מידע למדיניות ולתכניות חינוכיות. חברי הצוות הם: פרופ' לאה אחדות – המרכז האקדמי רופין ומכון ון ליר בירושלים (יו"ר); פרופ' מיכל בלר – מכללת לוינסקי לחינוך; מר נעם זוסמן – בנק ישראל; פרופ' טל זירסקי – אוניברסיטת חיפה; פרופ' איריס טבק – אוניברסיטת בן-גוריון בנגב; פרופ' אורלי מנור – אוניברסיטה העברית והדסה. מרכז הצוות מטעם היזמה הוא עודד בושריאן.

בפני הצוות עמדו שלוש מטרות מרכזיות: א) בחינת דרכים ליצירת תשתיות של נתוני אורך בתחום החינוך ובתחומים המשיקים לו, הנגשתם והרחבת השימוש בהם לצורך מחקר וקביעת מדיניות ב) קידום שיתוף פעולה בין גופים המחזיקים בנתוני אורך במטרה לייצר מסדי נתונים אינטגרטיביים ג) בחינת היבטים טכנולוגיים ומשפטיים-אתיים הכרוכים ביצירת תשתיות של נתוני אורך והנגשתם לציבורים רחבים, תוך שימת דגש על שמירת הפרטיות.

הצוות קיים שלושה מפגשים עם נציגים של משרדי ממשלה, בעיקר משרד החינוך והלשכה המרכזית לסטטיסטיקה (הלמ"ס), אנשי אקדמיה ומומחים אחרים – ששיתפו את החברים בניסיונם ובידע שלהם (סדר היום של המפגשים בנספח ב', תקצירי קורות חיים של המשתתפים בנספח ג'). שני מפגשים הוקדשו לתשתיות הנתונים הנמצאות במשרד החינוך ובלשכה המרכזית לסטטיסטיקה, ובמהלכם נסקרו מסדי הנתונים הקיימים, האופן שבו הם מאורגנים ונשמרים, דרכי ההנגשה שלהם לציבור הרחב ובמיוחד לחוקרים ועוד כיוצא באלה. מפגש נוסף הוקדש בעיקרו להיבטים המשפטיים הכרוכים ביצירת נתוני אורך, בהעמדתם לרשות הציבור ובתכניות לעתיד – תוך שימת דגש על הגנת הפרטיות. כן פגש הצוות מומחים בתחומים של התממה (anonymization) ואבטחת מידע.<sup>2</sup> הצוות פרסם שני קולות קוראים לסקירות: 'מיפוי מאגרי מידע בתחום החינוך' – להכנת סקירה זו נבחרה ד"ר עמליה רן ממרכז המידע במכון מופ"ת – ואבטחת מידע ופרטיות'. להערכת צוות המומחים, לא התקבלו הצעות איכותיות דיין לסקירות נושא השני, ולפיכך הוחלט

<sup>2</sup> בשפה האנגלית קיימים מספר מונחים המתארים שינוי קובץ נתונים על מנת להקשות על זיהוי האנשים שמאחורי הנתונים: anonymization, de-identification, Statistical disclosure control ועוד. חוקרים מסוימים משתמשים במושגים הללו באופן שונה, אולם לא קיימת כרגע משמעות מוסכמת שונה לכל אחד מהביטויים. בעבודה זו נשתמש במונח התממה כתרגום לכל הביטויים הללו.

להשלים את המידע הנחוץ על ידי מספר פגישות ושיחות טלפוניות עם מומחים, וחיבור סקירה לא־מקיפה שהכין מרָז הצוות.

בחורף 2016 ערך צוות המומחים יום עיון לסיכום פעילותו ולהצגת הידע שצבר – לקהל הרחב. יום העיון עסק בשלושה נושאים, בהתאם לנושאים שעלו בתהליך הלמידה: המושב הראשון עסק בנתוני אורך ובשימוש בהם לצורכי מחקר וגיבוש מושכל של מדיניות חינוכית. המושב השני דן בנתונים המינהליים בתחום החינוך הקיימים בישראל ובדרכים (המצויות והרצויות) להנגיש אותם לציבור החוקרים. המושב השלישי כלל דיון בצורך להגן על פרטיות מושאי הנתונים המינהליים ועל הדרכים לעשות כן.

### **סיכום: תובנות מפעילות צוות המומחים**

במהלך הפעילות התגבשו אצל צוות המומחים מספר תובנות, חלקן עלו במפורש מהרצאות הדוברים במפגשים הלימודיים, חלקן מקריאת ספרות מחקרית בנושא, וחלקן מהיכרותם של חברי הצוות עם הנושא ומומחיותם בתחום. בסיכום להלן נציג תובנות אלו, בחלוקה לפי הנושאים שבהם עוסק הדוח הנוכחי. חשוב לציין כי תובנות אלו אינן מבוססות על מחקר מקיף או על סקירה מקיפה של הידע המדעי בנושאים שבהם הן עוסקות. כן אין לראות בתובנות הללו המלצות שיש עליהן תמימות דעים של חברי הצוות. עם זאת צוות המומחים מקווה כי התובנות יכולות לאפשר גיבוש מושכל יותר של מדיניות בנושאים הללו.

### **נתוני אורך, מחקרי אורך והשימוש בהם לגיבוש מדיניות חינוך בישראל ובחינתה**

1. הניסיון שהצטבר מתהליך קבלת ההחלטות על עריכת סקר אורך ("סקר אורך טווח") על ידי הלמ"ס, תהליך שנמשך כעשר שנים, מלמד על הקשיים הרבים הכרוכים בגיוס מקורות מימון ציבוריים לעריכת סקרי אורך בשל עלותם הגבוהה וההתחייבויות ארוכות הטווח שהגופים המממנים צריכים לקבל על עצמם כבר בראשית הדרך. גם הניסיון שהצטבר באירופה ובישראל בפרויקט SHARE מלמד שעם השנים, סקרים אלה הופכים להיות מורכבים יותר ומחייבים מאמצים רבים למציאת מקורות מימון עבור כל גל של ראיונות. כמו כן מתעוררים קשיים הנובעים מנשירה של נסקרים עם הזמן. אנו סבורים שהשאלה אם לערוך סקרי אורך נוספים, בתחום החינוך למשל, רצוי שתיבחן בדיונים הנוגעים ליצירת תשתיות נתונים לאומיות, שבמסגרתם ההחלטה על עריכת סקר אורך תלויה גם בהבטחת מימון אורך טווח לביצועו.

2. חשוב לציין שעל אף שצוות המומחים החליט להתמקד באפשרות של חיבור נתונים מינהליים הנמדדים לאורך זמן על מנת לייצר נתוני אורך, יש תחומים רבים שבהם יש יתרון למחקרי אורך המבוססים על איסוף נתונים פעיל. כך למשל מחקרים המבוססים על איסוף נתונים מאפשרים לבדוק הסברים לתופעה מסוימת בקלות רבה יותר מאשר מחקרים המבוססים על נתונים מינהליים. יתרון נוסף של הסוג הראשון הוא שכיוון שהמשתתפים בסקר אורך נותנים בדרך כלל את הסכמתם לאיסוף הנתונים באופן אקטיבי (בניגוד לאיסוף מינהלי, שבו המדינה לא מבקשת את הסכמת האזרחים), הסיכון לפגיעה בפרטיות קטן בהרבה, לפחות מבחינה משפטית.

3. השימוש במקורות מידע מינהליים ליצירת תשתיות של נתוני אורך בתחומים שונים כגון חינוך, בריאות, עבודה ושכר הולך ומתרחב בקרב קהיליית החוקרים ובקרב מקבלי החלטות במדינות רבות. גם ישראל עשתה כברת דרך לא מבוטלת בכיוון זה. הצוות התרשם שהן אנשי הלמ"ס והן אנשי משרד החינוך הפנימו בשנים האחרונות את חשיבותה של הנגשת נתוני חינוך לצורכי מחקר, ופעלו רבות על מנת לקדם גישה נוחה ויעילה של חוקרים לנתונים האלה. עם זאת הצוות שמע מכמה מקורות כי זמן ההמתנה

לנתונים מחדרי המחקר ועלות קבלתם עדיין גבוהים יחסית, בעיקר מפני שהמדינה אינה מקצה מספיק תקציבים ותקני כוח אדם להנגשתם. צמצום זמן ההמתנה לנתונים והפחתת עלותם עשויים לקדם מאוד את השימוש של החוקרים בהם.

4. יש מקום להמשיך ולשקול את הרחבת התהליך ושיתוף הפעולה בין גופים ציבוריים על בסיס אסטרטגיה כוללת וראייה מערכתית של הצרכים והתועלות, ובהתאם לסדרי העדיפויות במחקר ובקביעת המדיניות. בהיעדר אסטרטגיה כוללת של הממשלה, יש חשיבות לכך שמשרדי הממשלה, ולענייננו משרד החינוך, יקבעו מדיניות ויעדים ליצירת נתוני אורך ולהנגשתם לציבור. אנו מתרשמים שזו הדרך שגם משרד הבריאות מאמץ בימים אלה.

5. לזמינותם של נתוני אורך חשיבות רבה במיוחד למחקר בתחום החינוך. תהליך רכישת ההשכלה הוא מתמשך, מעורבים בו שחקנים רבים (תלמידים, מורים, הורים ועוד), ובמהלכו מתרחשים מעברים בין דרגי חינוך ונצבר הון אנושי. לפיכך יש חשיבות למעקב אחר התלמידים וסביבתם לאורך זמן. יש הטוענים שמורכבותו של התהליך החינוכי לא מאפשרת למעשה לתאר אותו בלי לתת את הדעת לעברךם של המשתתפים בו. זאת ועוד: ההשפעה של התהליכים החינוכיים מתרחשת בטווחי זמן שונים, וגם עוצמתה אינה בהכרח זהה במהלך השנים. יש למשל אירועים שהשפעתם ניכרת מיד (דוגמה: תלמיד שמשתמש בסמים), ויש כאלו שתוצאותיהם יבואו לידי ביטוי לאחר פרק זמן מסוים (דוגמה: שינויים באיכות המורים עשויים להשפיע על תוצאות המבחנים רק לאחר שיעבור זמן). רק נתוני אורך יכולים ללמד אותנו על זמן התגובה.

6. תובנה מעניינת שעלתה מהרצאתו של פרופ' דיוויד קפלן ביום העיון היא שנתוני אורך אינם מוכרחים להיות ברמת הפרט בלבד. נתונים מינהליים הנמדדים לאורך זמן על בתי ספר או אפילו על מדינות (כגון PISA ו-TIMSS) יכולים להיחשב לנתוני אורך המלמדים על אותם בתי ספר או מדינות.

7. חשיבות רבה מיוחסת לשימוש במידע מינהלי לא רק כנתוני אורך העומדים בפני עצמם, אלא גם במסגרת אסטרטגיה כוללת של קישור נתוני אורך עם נתוני חתך (מסקרים או ממקורות מינהליים). קישור שכזה יכול לסייע בהערכת התוצאות של תכניות התערבות, לרבות במסגרת ניסויים מבוקרים. לאסטרטגיה זו יתרונות רבים ועל כן השימוש בה רווח יותר הן בישראל והן ברחבי העולם.<sup>3</sup>

### נתונים מינהליים הנאספים בישראל והנגשתם לצורכי מחקר

1. הצוות הדגיש את תפקידה המרכזי של הלמ"ס הן בתכלול המידע והן בגיבוש סדרי העדיפויות לאיסופו. הדברים חלים גם על מידע שנאסף על ידי גופים ציבוריים אחרים – כפי שעולה מפקודת הסטטיסטיקה. יש מקום להרחיב את ההסכמים לשיתוף פעולה בין הלמ"ס למשרדי הממשלה, בדומה להסכמים שנחתמו עם בנק ישראל והמוסד לביטוח לאומי. מטבע הדברים יהיה צורך להתאים את היקף המשאבים העומדים לרשות הלמ"ס, לרבות לצורך הכנת קובצי נתונים לציבור החוקרים, על מנת להוציא לפועל את המהלכים האלה.

2. הצוות התרשם שמשרדי הממשלה יכולים להפיק תועלת רבה מאיסוף שיטתי של נתונים בתחומי עיסוקם ומאחסונם במערכות מידע מרכזיות אינטגרטיביות. נתונים אלו עשויים ללא ספק לשרת את המשרד לצורך ייעול הפעילות, הערכתה וקביעת יעדי מדיניות. אותם נתונים עשויים לשרת גם את קהיליית החוקרים ואת הציבור הרחב.

<sup>3</sup> ראו למשל Dynarski, 2014

3. בשל השימוש הנרחב במיקור חוץ, משרדי הממשלה יתקשו לקבל תמונה מלאה של פעילותם אלא אם ייקבעו הערוצים שבאמצעותם ספקי השירותים יעבירו למשרדים באופן שוטף מידע מלא על פעילותם.
4. חשיבותה של הנגשת מידע לצורכי מחקר ולטובת גיבוש מושכל של מדיניות מחייבת אותנו לחיפוש מתמיד אחר כלים סטטיסטיים, טכנולוגיים ומשפטיים כדי להגן על המידע ולצמצם ככל האפשר את הסיכון לפגיעה בפרטיות. עם אלה נמנים כמובן שיטות להתממת מידע כמו גם כלים חוזיים-משפטיים וטכנולוגיים שאחדים מהם מוצגים בדוח זה.
5. התובנה העיקרית שאליה הגיע הצוות (ועל כך יורחב בהמשך) היא שגם אם נעשה שימוש באותם כלים, יהיה סיכון לפגיעה בפרטיות של לפחות אחד ממושאי המידע. עם זאת ניתן, באמצעות שימוש בשיטות מוכרות, לגרום לכך שעלות הזיהוי מחדש תהיה גבוהה מאוד במונחי זמן ומשאבים. לפיכך יש מקום לניהול סיכונים – בחינת התועלת הגלומה בהנגשת המידע לציבור החוקרים לעומת סיכון הפגיעה בפרטיות.
6. הצוות תמים דעים בדבר הצורך בבניית תשתיות למערכות גישה מרחוק ("חדר מחקר וירטואלי") שיאפשרו לחוקרים לגשת לנתונים מחדר עבודתם (להבדיל מחדרי מחקר ייעודיים כמו אלו של הלמ"ס והמוסד לביטוח לאומי), בדומה לפתרון שמציע משרד החינוך לחוקרים.
7. פרסום קטלוגים של מסדי נתונים (לרבות ה-metadatas) המצויים ברשות משרדי הממשלה והלמ"ס יכולה לסייע הן לגופים והן לציבור המשתמשים הפוטנציאליים. יש גם מקום שהלמ"ס תכין קטלוג של מסדי נתונים שהוכנו במיוחד עבור חוקרים לעבודה בחדר המחקר ויכולים לשרת חוקרים אחרים.
8. הצוות מצביע על מחסור משמעותי במידע בכמה תת-תחומים בחינוך. קִסר זה נובע מכך שחלק מהמידע לא נאסף, וחלק אחר נאסף אך אינו זמין. הצורך במידע מתבקש גם לאור אינדיקטורים נוספים בחינוך שראוי לפתחם (להרחבה ראו יוסטמן ובקובזה, 2010). להלן התחומים העיקריים שבהם קיימים פערי מידע:
  - תקצוב מערכת החינוך על ידי רשויות מקומיות, עמותות והורים.
  - פעילות רשויות מקומיות במערכת החינוך.
  - פעילות של עמותות במערכת החינוך ומחוצה לה.
  - מידע מפורט על גני ילדים שאינם ציבוריים (תלמידים וגננות).
  - מידע על הנעשה במגזר החרדי (לדוגמה הישגים לימודיים) ובחינוך במזרח ירושלים.
  - השתתפות תלמידים ומורים בתכניות התערבות.
  - מידע על תשתיות במערכת החינוך (מבנים, מחשבים וכיוצא בזה).
  - קישור מורים לתלמידים ברמת המקצוע (תנאי מפתח להערכת איכות ההוראה, אי-השוויון בחינוך ועוד).
  - חינוך בלתי פורמלי (תנועות נוער, מתנ"סים ועוד).
  - מיומנויות "רכות", תהליכי למידה והוראה.
  - עמדות, ערכים וכדומה, הדרושים להשלמת המידע על הישגים בבחינות.
  - נתונים שצה"ל אוסף בעת הגיוס ובמהלך השירות.

### שמירה על פרטיות והתממת מאגרי מידע

1. לא קיים מדרג דיכוטומי בין מידע מותמם, מצד אחד, לבין מידע שניתן לזהות מתוכו פרטים, מצד שני. תובנה זו הוצגה הן בהרצאתו של פרופ' ז'רסקי, הן בהרצאתו של עוה"ד בריאן ניגן, והן בהרצאתה של עוה"ד לימור שמרלינג (כולן במפגש הלימודי באוגוסט). התובנה גם היא עולה יותר ויותר בספרות העוסקת בנושא התממה (לדוגמה: HITRUST De-Identification Working Group, 2015; Garfinkel, 2015). קיים כמובן מידע שבוודאות לא מאפשר זיהוי של פרט מסוים (למשל קובץ המציג רק

סיכום או מאפיינים סטטיסטיים של הנתונים), וקיים מידע שבוודאות מאפשר זיהוי כזה (למשל כל קובץ נתונים הכולל מספרי תעודות זהות). אולם רוב קובצי הנתונים יימצאו בטווח שבין שני הקצוות הללו.

2. על מנת לברר היכן במנעד הזה נמצא קובץ הנתונים יש לשקול את הסיכון לזיהוי. גרפינקל (Garfinkel, 2015, p. 6) מציע את המנעד הבא (מהסיכון הנמוך ביותר לפגיעה בפרטיות, לגבוה ביותר):

נתונים שאינם עוסקים בפרטים ← נתונים שעוסקים בפרטים אך לא ניתן לזהות בתוכם פרט מסוים  
 ← נתונים שניתן לזהות כשייכים לאחד מבין X פרטים (K אנונימיות) ← נתונים שניתן לזהות בתוכם פרט מסוים ← נתונים בעלי משתנה מזהה.

פרופ' ז'רסקי בהרצאתו מוסיף כי הסיכון לפגיעה בפרטיות אינו מושפע רק מאופי הנתונים, אלא גם משאלות כגון:

- איזו רמה של מומחיות דרושה על מנת לזהות פרטים מתוך קובץ הנתונים?
- האם ניתן לזהות פרטים מתוך הקובץ אם קיים מידע נוסף בְּרִשׁוֹת המנסה לזהות?
- האם ניתן למצוא בקלות את המידע הנוסף הזה?
- האם ניתן למצוא את המידע הנוסף הזה, אך הדבר כרוך בעבירה על החוק?
- כמה אנשים מתוך קובץ הנתונים עלולים להיחשף לפגיעה בפרטיותם?

3. עו"ד לימור שמרלינג ממינהל רמו"ט, הציגה בהרצאתה במפגש הלימודי באוגוסט מידע שיכול לעזור לקבוע את מידת הסיכון: טבלה המהווה את תפיסת הרגולטור הישראלי לגבי ניהול הסיכון שבחשיפת המידע. טבלה זו, יש לציין, כוללת, בנוסף לשאלות לגבי המידע ולגבי הסיכון בחשיפתו, גם שאלה לגבי הנזק העלול להיגרם לפרטים מחשיפת המידע עליהם:

סיכון גבוה		סיכון נמוך
רבים	מהי כמות האנשים שעליהם יש מידע?	מעטים
רב	מהי כמות המידע שנשמר על כל אדם?	מועט
גבוהה	מהי רגישות המידע הנשמר?	נמוכה
ללא הגבלת זמן	מה משך זמן השמירה של המידע?	זמן קצר
רבה	מהי כמות המידע שמועבר?	לא מועבר
רבים	כמות הממשקים החיצוניים למערכת המידע:	ללא ממשקים
גדול	מה הסיכון לנזק ליחיד מגילוי המידע?	לא קיים או נמוך
קישור חזק לזהות	האם המידע מקושר באופן חזק לזהות?	קישור חלש או לא קיים
ניתן	האם ניתן לעשות במידע שימוש אחר?	לא ניתן
לא מודעים	האם האנשים מודעים לשימוש במידע?	מודעים
לא	האם איסוף המידע נעשה בהסכמת אנשים?	כן
לא	האם השימוש במידע נעשה בהסכמת אנשים?	כן

4. על פי עו"ד בריאן ניסן, היועץ המשפטי של הלמ"ס, גם ברמה הרגולטורית לא קיימת הנחיה חד־משמעית הקובעת שקובץ נתונים מסוים הוא מותמם לחלוטין. למעשה, כל ההנחיות שניתנו כוללות מידה זו או אחרת של שיקול דעת המוענק למי שאמור לקבוע אם הקובץ מותמם דיו או לא. כך למשל, התקנות של האיחוד האירופי דורשות שיינקטו "כל האמצעים הרלוונטיים שיש סבירות שגורם שלישי יעשה בהם שימוש כדי לזהות את היחידה הסטטיסטית". ההנחיה אינה קובעת מהם האמצעים הרלוונטיים או מהי "סבירות". ההנחיה באוסטרליה דורשת הסרה של מידע אשר "יכול, בסבירות גבוהה, לאפשר זיהוי של אותו אדם". גם כאן לא ברור מהי "סבירות גבוהה". ניסוחים דומים מופיעים גם בשאר ההנחיות.
5. גם אם נעשה שימוש בכלים סטטיסטיים, משפטיים, וטכנולוגים להגנה על המידע, יהיה כמעט תמיד סיכון לפגיעה בפרטיות של לפחות אחד ממושאי המידע. אין למעשה אפשרות להבטיח שלא ייתכן זיהוי של אף אחד ממושאי המידע.
6. למרות האמור לעיל ניתן, באמצעות שימוש בכלים המוכרים, לגרום לכך שהעלות, בזמן ובמשאבים, של זיהוי מחדש תהיה גבוהה מאוד. יש להבין את התחום, אם כך, לא כתחום של קביעה מוחלטת של מותר ואסור אלא כתחום של ניהול סיכונים והערכה האם הסיכון לפגיעה בפרטיות (לאחר שננקטו כל הצעדים האפשריים) עולה על הסיכוי שמחקר המסתמך על הנתונים יתרום לטוב הציבורי.
7. ניהול סיכונים מסוג זה חייב להיעשות יחד עם הערכת הסיכויים – חייבים להביא בחשבון את התועלת. יש להעריך מהי התועלת האפשרית בכלל, ולענייננו – התועלת האפשרית ממחקר באופן כללי, וכן מהי התועלת באופן ספציפי מכל מחקר לפי המטרות והחשיבות של המחקר המוצע. נקודה זו משמעותית כיוון שכרגע אין שיתוף שיטתי של חוקרי חינוך ושל קובעי מדיניות בתחום החינוך בתהליך ההחלטה אם וכיצד לאפשר גישה לנתוני חינוך בישראל.

## מבנה הדוח

בעמודים הבאים יוצג עיקר המידע שאסף צוות המומחים במהלך פעילותו. הדוח יתמקד בשלושה נושאים, ולכל אחד מהם יוקדש פרק נפרד. הפרק הראשון יעסוק בנתוני אורך, מחקרי אורך, וחשיבותם למחקרים בתחום החינוך ולעיצובה של מדיניות חינוך יעילה ומועילה. הפרק השני יבחן את הנתונים המינהליים הקיימים בישראל בתחום החינוך, ואת הנגשתם על ידי הגופים השונים המחזיקים בהם, לצורכי מחקר ולצורכי ניהול של מדיניות חינוך. הפרק השלישי יבחן את המצב החוקי והרגולטורי בישראל בנוגע לשמירה על פרטיות מושאי הנתונים בעת הנגשת נתונים מינהליים לצורכי מחקר, וכן יציג גישות ושיטות שונות להערכת הסיכון לפגיעה בפרטיות ולהפחתתו.

## פרק ראשון – נתוני אורך ומחקרי אורך

פעילות הצוות עסקה כאמור בנתוני אורך בחינוך ובאפשרויות המחקריות שנתונים אלו מזמנים. הפרק הראשון בדוח יוקדש אם כן לנתונים כאלה ולמחקרים המבוססים עליהם. בחלקו הראשון של הפרק יוצג מעין מבוא לתחום, שיכלול את ההגדרות של נתוני אורך ומחקרי אורך, ההבדלים בין נתוני חתך לנתוני אורך, הסוגים השונים של נתוני אורך והאפשרויות למחקר שהם מספקים. לאחר מכן יוצגו סיכומי הרצאות, מתוך החלק ביום העיון שעסק בנושא מחקרי אורך, סיכומים שבהם מודגמים היתרונות והאפשרויות של השימוש בנתוני אורך לצורכי מחקר ולצורכי ניתוח מדיניות. כל אחת מההרצאות תציג את הנושא מזווית שונה מעט.

### נתוני אורך ומחקרי אורך: מבוא קצר

**נתוני אורך ונתוני חתך:** באופן כללי קיימים שני סוגים של נתונים – נתוני חתך (Cross-section) ונתוני אורך (Longitudinal). נתוני חתך מתארים אוכלוסייה בנקודת זמן נתונה. במדעי החברה משתמשים בנתוני חתך לעתים קרובות למציאת קשר בין תופעות או התנהגויות שונות של קבוצות באוכלוסייה. נתוני חתך מאפשרים לבחון משתנים רבים בעת ובעונה אחת, אך בדרך כלל קשה להסיק באמצעותם על תהליכים המתמשכים לאורך זמן.<sup>4</sup> לעומת זאת, נתוני אורך, בהגדרתם הרחבה, הם נתונים הנאספים עבור כל משתנה בשתי תקופות או יותר או שהנתונים נאספים עבור אותן יחידות מחקר לאורך זמן. על פי הגדרה רחבה זו כל רצף של תצפיות שנערכות בזו אחר זו יכול להיקרא נתוני אורך.

**סוגים של נתוני אורך:** מקובל בספרות המחקרית לדבר על שלושה סוגים של דרכים שמייצרות שלושה סוגים של נתוני אורך. הסוג הראשון הוא סדרה של מחקרי חתך חוזרים (Repeated cross sectional studies) שכל אחד מהם מתאר את אותה האוכלוסייה בנקודת זמן אחרת. במקרה כזה המדגמים – תת-קבוצות מתוך האוכלוסייה כולה – יכולים להיות שונים מתקופה לתקופה (מדגמים בלתי תלויים) כל עוד הם חלק מאותה אוכלוסייה. סוג אחר של נתוני אורך מושג מסקרים רטרוספקטיביים – סקרים הנערכים בנקודת זמן אחת אך כוללים שאלות לגבי עברו של הנשאל ואוספים, אם כן, מידע על ההיסטוריה שלו. הסוג השלישי של נתוני אורך, שהוא הסוג שאליו התייחסה עבודתנו של צוות המומחים ואליו נתייחס בדוח זה, הוא זה המופק מתוך 'מחקרי פנל' (Panel studies) – נתונים המודדים את אותם הפרטים<sup>5</sup> בשתי נקודות זמן או יותר.

נהוג לציין שני סוגים של נתוני פנל: מדגם פנל מייצג (Representative panel), ומחקר עוקבה (Cohort Study). מדגם פנל מייצג, כשמו כן הוא, אוסף נתונים על מדגם מייצג של האוכלוסייה הנחקרת בשתי נקודות זמן או יותר. ההבדל בין מדגם פנל מייצג לבין מחקר חתך חוזר הוא כמובן שהראשון עוקב אחר אותו מדגם – אותה תת-קבוצה מתוך האוכלוסייה – על פני כל תקופת הזמן. מחקרי עוקבה עוקבים אחר קבוצה של אנשים שחולקים מאפיינים דומים או שחוו אירועי חיים משמעותיים בפרק זמן מוגדר. הדוגמה הנפוצה ביותר למחקר עוקבה היא מעקב אחר קבוצה של פרטים שנולדו בתוך פרק זמן מוגדר. מחקר מעקב כזה נקרא עוקבת לידה (birth cohort). דוגמאות אחרות למחקרי עוקבה הן מחקרים שעוקבים אחר קבוצת אנשים בעלי אותה שנת נישואין, אנשים שנחשפו לתרופה מסוימת ועוד.

<sup>4</sup> אם כי נתוני חתך עשויים לכלול מידע על זיכרונות עבר של מושא הנתונים.

<sup>5</sup> בדרך כלל מדובר במעקב אחר אותה קבוצה של בני אדם לאורך זמן, אבל ייתכן גם מעקב אחר כל אוכלוסיית מחקר אחרת, למשל בתי ספר, כיתות או משפחות.

**היתרונות העיקריים של נתוני אורך ושל מחקרי אורך:** היתרון הראשון ואולי החשוב ביותר של מחקרי אורך על פני מחקרי חתך הוא כמובן **האפשרות לבחון תופעות על פני זמן** וכך לגלות התפתחויות ולזהות דפוסי שינוי על פני זמן. מחקרים כאלו מאפשרים לעקוב לאורך זמן אחר סדרה של אירועים, נסיבות, מאפיינים והתנהגויות של האוכלוסייה הנחקרת.<sup>6</sup> באמצעות המעקב על פני זמן ניתן למצוא גם קשרים בין אירועים שהתרחשו בזמנים שונים, ולא רק אירועים שהתרחשו בסמיכות זה לזה.

מחקרי אורך מאפשרים גם **למדוד טוב יותר את השינויים בקשרים בין המשתנים על פני זמן** או לחילופין לבחון את יציבותם של הקשרים על פני זמן. למשל, אם מיצב חברתי כלכלי משפיע על רמת המשכל בגיל 3 אך הקשר מאבד מעוצמתו בגיל 16, יוכל מחקר אורך לגלות זאת בקלות רבה יותר מאשר מחקר חתך. מחקרי אורך יכולים גם לגלות השפעה מושהית – קשר בין משתנים שנוצר רק אחרי שעובר זמן – למשל כאשר שינוי מסוים שנערך בגיל צעיר משפיע, אבל רק אחרי זמן רב. דוגמה לכך היא המחקר של צ'אטי הנדרן וכץ (Chetty, Hendren, & Katz, 2015) שגילה שמעבר בגיל צעיר לשכונת מגורים ממיצב חברתי כלכלי גבוה יותר משפיע על הכנסתם של ילדים שעברו בגיל צעיר, אף על פי שמחקרים קודמים לא מצאו קשר בין המעבר לבין הישגים בלימודים.

נתוני אורך לא רק מחזקים את התובנות שלנו על תהליכי שינוי וקשרים בין משתנים, אלא גם **מאפשרים, כשמשתמשים במערך מחקר מתאים, לזהות סיבתיות**. בנתוני חתך קיים קושי לזהות סיבתיות, למעט, במקרה של ניסויים מבוקרים (Randomized controlled trials). זיהוי הסיבתיות מאפשר לחוקר להמליץ על התערבות, והוא תנאי הכרחי להערכה מהימנה של הצלחת תכניות התערבות ולגיבוש מדיניות.

מחקר אורך גם מקנה לחוקרים **גמישות** במובן שהוא מאפשר להם להוסיף מוקדי מחקר בעת איסוף הנתונים (למשל להוסיף עוד מידע במהלך עריכת המחקר). החוקר יכול לבחור להוסיף לשאלון שאלות נוספות, וכך לבחון השערות שעלו אצלו לאחר שהתחיל בתהליך איסוף הנתונים.<sup>7</sup>

לזמינותם של נתוני אורך חשיבות רבה במיוחד למחקר בתחום החינוך. תהליך רכישת ההשכלה הוא מתמשך, מעורבים בו שחקנים רבים (תלמידים, מורים, הורים ועוד), ובמהלכו מתרחשים מעברים בין דרגי חינוך, ונצבר הון אנושי. לפיכך יש חשיבות למעקב אחר התלמידים וסביבתם לאורך זמן. יש הטוענים שמורכבותו של התהליך החינוכי לא מאפשרת למעשה לתאר אותו בלי לתת את הדעת לעִבְרָם של המשתתפים בו.

זאת ועוד: ההשפעה של התהליכים החינוכיים מתרחשת בטווחי זמן שונים, וגם עוצמתה אינה בהכרח זהה במהלך השנים. יש למשל אירועים שהשפעתם ניכרת מיד (דוגמה: תלמיד שמשתמש בסמים), ויש כאלה שתוצאותיהם יבואו לידי ביטוי לאחר פרק זמן מסוים (דוגמה: שינויים באיכות המורים עשויים להשפיע על תוצאות המבחנים רק לאחר שיעבור זמן). רק נתוני אורך יכולים ללמד אותנו על זמן התגובה.<sup>8</sup>

**נתוני אורך מהסוג של מדגם פנל מייצג יכולים להיאסף בשתי דרכים עיקריות:** הראשונה היא השימוש **בסקרי אורך** – סקר שמועבר על אותם נבדקים בתקופות זמן שונות. השנייה היא שימוש בנתונים שנמדדו בעבר עבור מדגם של האוכלוסייה, לדוגמה נתונים מינהליים שאוספת המדינה. לסקרי אורך יש יתרונות לא מעטים, אך יש להם גם כמה חסרונות שנדמה כי יקשו על עריכתם בישראל. הנה כמה מהחסרונות הללו: עלות עריכתם גבוהה יחסית (גם בכסף וגם בזמן) והיא דורשת התחייבות למימון על פני זמן רב. התפוקה המדעית מסקרים אלו מגיעה כעבור זמן רב יחסית, ולכן ניתן להתחיל לערוך מחקרים המבוססים עליהם רק לאחר שנאספו לפחות שני "גלים" של נתונים. יתרה מזו, לעיתים יש צורך במספר גדול של גלי ראינות כדי לאפשר ניתוח של שינוי ארוך טווח.

<sup>6</sup> המונח המקצועי הוא ניתוח התפתחותי (diachronic analysis).

<sup>7</sup> אפשרות זו מוגבלת כמובן למחקרי אורך המבוססים על סקרי פנל. כפי שנראה להלן, לא כל מחקרי האורך מבוססים על סקרים.

<sup>8</sup> להרחבה ראו שביט ופניגר, 2007.



דרושה הקפדה יתרה על הפעלת פרוצדורות הדגימה ואיסוף הנתונים ועל קביעת כללי מעקב ברורים; נדרשת מחויבות לאורך זמן רב מהמשתתפים ומהחוקרים. הייצוגיות של המדגם עשויה להיפגם כתוצאה מנשירה (Attrition), מצב שבו לא מצליחים לראיין שוב את אלו שרואיינו בגלים קודמים. עובדה זו מקשה מאוד על שמירת הייצוגיות, והיא גם עלולה ליצור הטיית בחירה אם יותר אנשים בעלי תכונות מסוימות (או שקרו להם מקרים מסוימים) יטו לפרוש ולא להיות זמינים להיסקר. לבסוף, קיימת בעיית הכיסוי של אוכלוסיית היעד, שכן הדגימה נעשית בגל הראשון אך בפועל האוכלוסייה משתנה על פני זמן. למשל, סקרי אורך שנערכים בארצות הברית ושהתחילו בשנות ה-80 אינם מייצגים כבר את ההרכב העכשווי של אוכלוסיית ארצות הברית מבחינת ארצות המוצא, וזאת בגלל ההגירה מדרום אמריקה וירידת שיעור הילודה של הלבנים.

בהשוואה לנתוני פנל, השגת מידע באמצעות נתוני אורך רטרוספקטיביים זולה וקלה יחסית (כיוון שמדובר בסקר בודד – כמו בנתוני חתך). אלא שנתונים אלה סובלים בנוסף לבעיית הייצוגיות גם מהטיות של זכירה – בעיקר של מצבים קוגניטיביים, עמדות, רגשות ומוטיבציה. יש גם בעיה של צורך בסובלנות רבה מצד המרואיין בשל המידע הרב שעליו לספק. אם המחקר מבקש למדוד שינוי על פני תקופה קצרה יחסית (שבועות או חודשים), ייתכן שכדאי לשקול מחקר רטרוספקטיבי עבור חקר אירועים והתנהגויות. אולם אם המחקר אמור לספק מידע על אירועים שקרו לפני שנים, סביר להניח שלא תתקבל תמונה מדויקת בהסתמך רק על זיכרונם של המשתתפים.

בהתחשב בחסרונות אלו ולאחר ששמענו על האתגרים שעומדים בפני הלשכה המרכזית לסטטיסטיקה בעבודתה על הסקר אורך הטווח שבהעברתו הוחל לאחרונה, העריכו חברי הצוות שיהיה קשה למצוא גוף שיפעיל סקר אורך טווח נוסף בישראל. בהתחשב בכך הוחלט לשים את הדגש על האפשרות השנייה להשגת נתוני אורך – **קישור של נתונים מינהליים הנמדדים על ידי המדינה עבור אותם פרטים בנקודות זמן שונות.**

נתוני אורך ממקורות מינהליים נהנים ממספר יתרונות בהשוואה לאלו שהתקבלו מסקרים: הנתונים המינהליים מכסים לרוב את כל האוכלוסייה הרלוונטית, ועל כן הם מייצגים כראוי את כל קבוצות האוכלוסייה לרבות הקטנות שבהן. לעומת סקרי אורך, שבהם קיימת נשירה, בנתוני אורך מינהליים תופעה זו בטלה בשישים. מטבע הדברים הפקת נתונים מינהליים מתבצעת בכל מקרה, ולכן העמדתם לרשות הציבור כרוכה בעלויות תקציביות זניחות. הנתונים המינהליים זמינים לרוב מדי שנה ובתוך פרק זמן קצר ממועד יצירתם, ולעומתם סקרים נערכים לרוב אחת לכמה שנים ותהליך הפקתם ארוך יחסית.

למרות הבחירה שלנו להתמקד בנתונים מינהליים, חשוב לציין שעלולים להיות בהם כמה וכמה חסרונות: הם נוטים להיות פחות עשירים מאלו הכלולים בסקרים, הם אינם כוללים מענה לשאלות סובייקטיביות, ובתחומים מסוימים איכותם עלולה להיות פחותה. מעבר לכל אלה, יצירת נתוני אורך מנתונים מינהליים כרוכה בבעיה המשפטית והאתית של שמירת הפרטיות, ומשום כך גם הנגישות לנתונים אלה עלולה להיות נמוכה.

ברחבי העולם קיימים סקרי אורך פרוספקטיביים ורטרוספקטיביים כבר שנים רבות. דוגמאות בולטות לכך הן ה־National Longitudinal Sample (NLSY) National Child Development Study (NCDS) הבריטי, ה־National Longitudinal Panel Study of Youth (NLSY) Study of Youth (SOEP) הגרמני, ה־British Household Panel Study (BHPS) Study of Income Dynamics (PSID) האמריקאי, ה־British Household Panel Study (BHPS) הבריטי ומחקר היסטוריית החיים הגרמני. חשוב לציין שקיימת מגמה לשלב בין נתוני פנל פרוספקטיביים ורטרוספקטיביים ובכך ליהנות משני העולמות. גם הספרות ממליצה בכיוון זה או בחיבור של נתוני אורך לנתוני חתך.

בישראל נערכו בעבר מעט סקרי אורך, ובעיקר כאלה שחוקרים יזמו באופן עצמאי על אוכלוסיות ספציפיות. כיום נערכים שני סקרי אורך חשובים: הראשון הוא 'סקר משקי בית בישראל – סקר אורך טווח' של הלמ"ס,

שהחל בשנת 2012 ועד כה נערכו בו ארבעה גלים של ראיונות.<sup>9</sup> הסקר מתמקד בתהליכים העוברים על משקי בית בישראל לאורך זמן במעגל החיים, במטרה לבדוק מעברים ושינויים בתחומים שונים, על רקע המאפיינים של משקי בית ונפשות בחברה בישראל, וכדי לסייע בתיעוד של תהליכים חברתיים וכלכליים לשם תכנון מדיניות ומחקר. חלק מהנושאים הנחקרים הם "שאלות גרעין", ואחרים יתחלפו מדי מספר גלים ויהוו "חטיבות מתחלפות". מטבע אופיו יכול סקר האורך של הלמ"ס לשמש פלטפורמה ליצירת נתוני אורך בתחום החינוך. הסקר השני הוא הסקר האירופי של בריאות, הזדקנות ופרישה (Survey of Health, Ageing and Retirement in Europe – SHARE) שאליו הצטרפה ישראל בשנת 2004.<sup>10</sup> זו דוגמה לסקר מעקב בינלאומי הנערך כיום על ידי 27 מדינות אירופיות וישראל, במטרה ליצור מסד רב-תחומי של נתוני אורך על האוכלוסייה המבוגרת (בני 50 ומעלה) כדי לעקוב אחר תהליך ההזדקנות והשלכותיו במגוון תחומי חיים. בישראל נערכו עד כה ארבעה גלים של איסוף נתונים ופורסמו התוצאות של שלושת הגלים הראשונים. ולבסוף, במסגרת סקר זה שולבו נתונים רטרוספקטיביים בנתונים פרוספקטיביים. אחד מגלי הראיונות מוקדש להיסטוריית חיים.

בד בבד עם תהליכי הגלובליזציה שחלו בעשורים האחרונים, הייתה עדנה גם להשוואה בינלאומית של תשומות ותפוקות של מערכת החינוך, ובעקבות זאת גדל מספר הסקרים ומעגל המדינות המשתתפות בסקרים האלו, לרבות ישראל. הסקרים בחנו את תפיסות המורים ושיטות ההוראה (TALIS), אוריינות התלמידים בתחומי דעת שונים (PISA, PIRLS ו-TIMSS) ומיומנויות היסוד של בוגרים (PIAAC). מדובר בסקרי רוחב; בתחום החינוך לא קיימים היום סקרי אורך בינלאומיים.

## סיכומי הרצאות מיום עיון בנושא נתוני אורך ומחקרי אורך<sup>11</sup>

### הזדמנויות ואתגרים באיסוף וניתוח של נתוני אורך בחינוך

#### דיוויד קפלן

בהרצאתי אני מבקש להדגים את האפשרויות והאתגרים של שימוש בנתוני אורך לצורך ניתוח מדיניות. ההרצאה תציג מחקרי אורך שהתבססו על נתונים שנסקרו ב-*Early Childhood Longitudinal Survey* בשנתון שלמד בגני הילדים בשנת 1998 (ECLS-K-1998).<sup>12</sup> זה היה סקר בקנה מידה גדול, והוא היה מבוסס על מדגם מייצג של אוכלוסיית ארצות הברית (באותן שנים). ניתן לטעון שזהו סקר האורך המקיף ביותר שֶׁבָּחן התפתחות יכולות אקדמיות אצל ילדים מגיל בית הספר היסודי ועד גיל החטיבה העליונה (elementary to secondary).

הילדים שנסקרו הגיעו גם מבתי ספר פרטיים וגם מבתי ספר ציבוריים, והיו ביניהם גם ילדים ששהו בגני הילדים ימים מלאים וגם כאלה ששהו בהם ימים חלקיים. הם הגיעו מרקעים חברתיים-כלכליים ואתניים מגוונים. ההורים, המורים ומנהלי בתי הספר והגנים השתתפו גם הם בסקר, וסיפקו מידע על ההתפתחות הקוגניטיבית, החברתית, הפיזית והרגשית של הילדים. בנוסף לכך נאסף מידע על הסביבה הביתית של הילדים, הפעולות החינוכיות בביתם, הסביבה הבית-ספרית והכיתתית שלהם, תכנית הלימודים שלהם ורמת ההכשרה של מוריהם.

<sup>9</sup> להרחבה ראו: הלשכה המרכזית לסטטיסטיקה, משקי בית בישראל, סקר אורך טווח.

<sup>10</sup> להרחבה ראו: מרכז ידע לחקר הזדקנות האוכלוסייה בישראל – סקר הבריאות, הזדקנות ופרישה באירופה.

<sup>11</sup> ההרצאות מסוכמות כאן כנתינתן ביום העיון, ובאישור המרצים.

<sup>12</sup> יש לציין שהמחקרים שהציג פרופ' קפלן אינם מבוססים על נתונים מינהליים אלא על מדגם פנל ועל איסוף פעיל של נתונים.

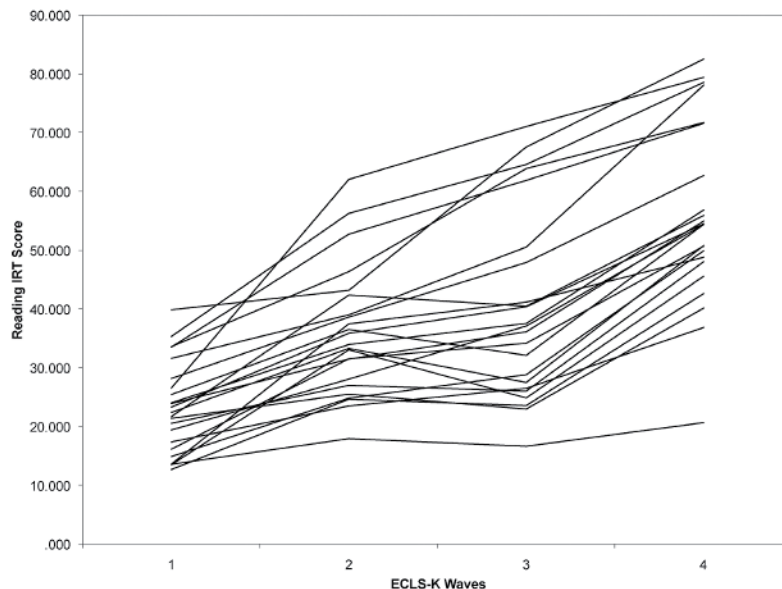
הסקר עקב כאמור אחר הילדים מגיל הגן עד סוף כיתה ח'. מטרתו הייתה לספק מידע אמין ומקיף שיאפשר לתאר ולהבין טוב יותר את התפתחותם של ילדים ואת החוויות שהם עוברים בגילים אלו. כמו כן אמורים הנתונים לאפשר הבנה טובה יותר של הקשר בין החוויות המוקדמות של ילדים בגן לבין התפתחותם המאוחרת יותר, הלימודים והחוויות שלהם בבתי הספר. המידע שנאסף במהלך השנים אפשר לחוקרים ומקבלי החלטות ללמוד איך החוויות שעובר הילד בבית, בבית הספר, בכיתה ובקהילה מעצבות את התפתחותו הפיזית, החברתית והרגשית.

על מנת לתאר את היתרונות שבאיסוף נתוני אורך, שימוש בהם ושימוש בשיטות מחקר ומודלים סטטיסטיים מתאימים נתחיל בדוגמה פשוטה: מחקר (שנציג בפירוט בהמשך) עקב אחר נתוניהם של 20 ילדים שונים מתוך ה־ECLS-K-1998. הילדים נבחרו באופן אקראי. מהנתונים שהתקבלו ייצרנו עקומות המציגות את התקדמות הילדים לאורך זמן. ניתן לבחון את העקומות הללו בשתי דרכים: לבדוק את השינוי בתוצאות אצל אותו ילד לאורך זמן, או לבדוק את ההבדלים בקצב השינויים לאורך זמן אצל ילדים שונים.

בבחינת ההבדלים בין הילדים אנחנו רואים שהם התחילו בנקודות שונות ויכולותיהם התפתחו בנתיבים שונים לאורך תקופת המעקב. היכולות של הילדים גם הגיעו לרמות שונות (אפילו יותר מאשר המקום שבו התחילו) בסופו של המעקב. יש לנו מודלים שמתארים את קצב הצמיחה הממוצע של ילדים, ואנחנו יכולים לבחון את קצב הצמיחה ואת היכולות שלהם בכל אחת מהנקודות שבהן נמדדו היכולות. בנוסף לכך אנחנו יכולים גם לייצר מודלים לא-ליניאריים של התפתחות ולהביא בחשבון קצב צמיחה שונה לכל ילד בכל שלב בחייו.

לאחר שבחנו את היכולת של הילדים בנקודת ההתחלה, את הצמיחה ביכולתם ואת היכולת שלהם בנקודת ה"סיום" (סוף המעקב), ניתן לעבור לרמה הבאה, לנסות להבין את ההבדלים ברמת הצמיחה של כל ילד ולמצוא משתנים המסבירים את ההבדלים הללו. במיוחד יעניינו אותנו כמובן המשתנים שעליהם המדיניות יכולה להשפיע והאופן שבו הם מתקשרים לנתוני הרקע ולנקודת הפתיחה של הילדים. אנחנו יכולים לבחון גם התערבויות ולבחון אם הן משפיעות על עקומות הצמיחה של הילדים.

כל האמור לעיל יכול להיעשות תוך כדי שלוקחים בחשבון את הסביבה הלימודית שבה נמצאים הילדים (השפעת הכיתה או בית הספר למשל). מודלים סטטיסטיים שיש לנו כיום מאפשרים לבחון את התרומה של בית הספר לעקומות הצמיחה של ילדים בהתאם לרקע ולמאפיינים של הילדים.



Change in Continuous Variables over Discrete Time

כאשר יש לנו נתוני אורך איכותיים דיים, מאפשרים לנו המודלים הסטטיסטיים הקיימים לבחון את קצב ההתפתחות של הילדים ולייצר מודלים מורכבים כדי להסביר אותם. אפשר לייצג ציונים ב"סוף" כפונקציה של המצב ההתחלתי וקצב הצמיחה. מודל מורכב יותר יכול להוסיף משתנים כגון מגדר, ולבחון אם יש עדיין מתאם בין נקודת הפתיחה לקצב הצמיחה. אפשר לעשות יותר: להביא בחשבון אירועים שמשפיעים על המשתנה הנמדד בכל נקודת זמן (Time varying variable).

אבל אפשר גם לעשות יותר מזה: אפשר לנתח שני תהליכים של התפתחות שקורים באותו זמן, לבחון כל אחד מהם בנפרד וכן לבדוק אותם במשותף, מבחינת הקשר הקיים בין תהליכי הגדילה שלהם. לדוגמה, מחקר שערכתי יחד עם ננסי ג'ורדן (Jordan & Kaplan, 2008) בעזרת נתוני אורך שאספנו בעצמנו, בוחן את צורת השימוש באצבעות לצורך ספירה אצל ילדים כמנבאת את דיוק הספירה שלהם. בדקנו אם השימוש באצבעות משתנה לאורך זמן ואם צורת השימוש באצבעות משפיעה על דיוק הספירה (וכפועל יוצא – אם כדאי לגמול אותם משימוש באצבעות בגיל מסוים). מצאנו שם שני תהליכים שהתרחשו במקביל: השימוש באצבעות פחת מעט לאורך זמן ואז, בשלב מסוים, נפסק באופן פתאומי. דיוק הספירה, לעומת זאת, עלה באופן לינארי לאורך זמן. השימוש בנתוני אורך מאפשר לחבר בין שני התהליכים הללו ולבחון אותם יחד. יתרה מזו – הוא מאפשר לשלב בינם לבין משתנים אחרים שאנחנו מעוניינים ללמוד עליהם: הבדלי מגדר, הבדלים בין ילדים מרמות חברתיות-כלכליות שונות וכן הלאה. מצאנו שילדים ממיצבים חברתיים כלכליים נמוכים אכן משתמשים באצבעות לאורך זמן רב יותר.

דוגמה מספרית: בחנו נתונים מתוך ה-ECLS-K על 3,575 ילדים. בחנו את האינדיקטורים לרמת הקריאה שלהם בנקודות הזמן השונות, וכן בדקנו משתני רקע חברתי-כלכלי: האם הם שהו בגן הילדים יום מלא או חצי יום ובאיזה גיל החלו לבקר בגן הילדים? שאלות המחקר היו מהו קצב הצמיחה ברמת הקריאה של הילדים המשתתפים במחקר? האם יש הבדל בין ילדים הלומדים יום מלא בגן לבין אלה הלומדים חצי יום והאם ההבדלים הללו שונים כאשר מביאים בחשבון את גיל הכניסה למסגרת ואת הרקע החברתי-כלכלי של הילדים?

ראשית מצאנו שאין הבדל משמעותי בקצב ההתפתחות של כישורי הקריאה בין ילדים המשתתפים במסגרת במשך יום מלא לבין אלה שמשתתפים בה חצי יום בלבד. אבל כאשר הכנסנו למשוואה את הרקע החברתי-כלכלי, התמונה החלה להיראות אחרת. אנחנו רואים הבדל גדול יותר מזה שראינו לפני שהובא בחשבון הגורם החברתי-כלכלי. פירוש הדבר הוא שבעצם יש הבדל בהרשמה ליום מלא או חלקי לפי מיצב חברתי-כלכלי ולפי גיל תחילת הלימודים.

לאחר שהסתכלנו בעקומת הגדילה הממוצעת, מצאנו גם תבניות צמיחה ובה עקומות ייחודיות של נתיבי התקדמות. במחקר המדובר ניתן להבחין שיש חלוקה לשלוש קבוצות מובחנות של מתווה התפתחות. קביעת עקומת גדילה אחת (One size fits all) תחמיץ אם כן ממצאים מעניינים בקשר לכל אחת משלושת הקבוצות. חשוב לנו לדעת אם יש תת-אוכלוסיות שיש להן עקומות גדילה ייחודיות. גם התערבויות חייבות להיות מותאמות לעקומות הגדילה הייחודיות לכל תת-קבוצה באוכלוסייה הנבחנת. עלינו לייצור אפוא מודלים "מעורבים" של צמיחה – כאלה המתאימים לכל תת-קבוצה באוכלוסייה (growth mixture modeling).

מודלים כאלה הם חלק מקבוצת מודלים מעורבים שהרעיון מאחוריהם הוא שכל אוכלוסיית מחקר מורכבת מתת-קבוצות (שמספרן וגודלן משתנה כמובן), ושהמודלים צריכים להתייחס לתת-קבוצות בתוך האוכלוסייה. רעיון כללי זה יכול להיות מיושם גם במודלים סטטיסטיים אחרים. תת-קבוצות באוכלוסייה מחולקות בהתאם למודלים התאורטיים המתאימים לכל מחקר ומחקר (בהתאם למשתנים שלפי התאוריה עשויים להשפיע על המודל), וכן על פי הממצאים האמפיריים (לפי הקבוצות שמצאנו בפועל בתוך האוכלוסייה). כאשר מיישמים את הרעיון למודלים המתארים צמיחה או התפתחות (growth mixture modeling) מחלקים את האוכלוסייה בהתאם לצורה הייחודית של עקומות הגדילה.

במחקר המדובר מצאנו כאמור שניתן לחלק את מתווי עקומות הגדילה לשלוש תבניות צמיחה: התבנית הראשונה, הגדולה ביותר (72% מהילדים), מתאימה לילדים שהתחילו עם ציונים נמוכים יחסית בקריאה, אבל התפתחו באופן מרשים ויציב יחסית במהלך השנים. התבנית השנייה מתאימה לילדים שהתחילו ברמה גבוהה יחסית, ואחרי תקופה של צמיחה הגיעו לרמה קבועה וגבוהה. התבנית השלישית מתאימה לילדים שהתחילו ברמה נמוכה יחסית, ולא הצליחו להתקדם בהשוואה לקבוצות האחרות.

אחרי שערכנו את החלוקה לשלוש עקומות הגדילה השונות, אפשר לבדוק אם יש הבדלים בין הקבוצות ובתוך כל אחת מהקבוצות. ניתן לחלק כל אחת מהקבוצות לילדים שנמצאים חצי יום בגן לעומת כאלו שנמצאים שם יום מלא. אפשר גם לחלק אותן לפי מיצב חברתי-כלכלי. כשעשינו זאת במחקר, עלה ממצא מעניין: בקבוצה החלשה מבחינת הצמיחה (הקבוצה השלישית) יש הבדל בין ילדים שנמצאים יום מלא בגן לבין כאלה שנמצאים בו חצי יום. זה לא נכון בשאר הקבוצות (כששולטים על גיל הכניסה לגן וה-SES, ההבדלים די נעלמים). במילים אחרות נמצא שלשהות של יום מלא בגן הילדים יכולה להיות השפעה חיובית על ילדים הנמצאים במתווה צמיחה איטי יחסית. מה שחשוב כמובן הוא לא הדוגמה הספציפית אלא אפשרויות הלימוד באמצעות מתודולוגיות קיימות כשיש נתוני אורך מתאימים.

בדוגמה שהצגנו ראינו התפתחות שמתמקדת בתוצאות או בציונים "המשכיים" שנמדדו לאורך זמן. ניתן להתמקד גם בנתונים "קטגוריים" שבהם השינוי או ההתפתחות מביאים להגעה (או אי-הגעה) למצב מסוים. כך למשל רכישה (או אי-רכישה) של יכולת מסוימת, או מעבר (או אי-מעבר) משלב התפתחות אחד של פיאיזה לשלב אחר (Piaget, 1964) – הם מעברים ממצב קיום אחד לשני. ה-ECLS-K מאפשר להעריך גם מעברים כאלו. ה-ECLS-K "מתרגם" ציונים בקריאה להסתברויות שהילד יוכל לשלוט בביצוע משימות ספציפיות, והוא גם מספק ציוני מסוגלות דיכוטומיים (יכול/לא יכול). פעולות אלה נעשות על ידי יצירת מקבצים של שאלות בתחום הקריאה שהן בעלות אותה רמת קושי. כאשר הילד מצליח לפתור שלושה מתוך ארבעה פריטים ברמת קושי מסוימת, הוא נחשב כבעל יכולת המתאימה לרמה זו.

בדוגמה שאציג כעת ניסינו (Kaplan & Walpole, 2005) להשתמש בכלים אלו על מנת לעקוב אחר התקדמות הילדים. חילקנו את הבנת הקריאה לשישה שלבים: זיהוי אותיות, זיהוי צלילי התחלת המילה, זיהוי צלילי סוף המילה, זיהוי מילים וזיהוי מילים מתוך הקשר. התוצאות אפשרו לנו לבחון לא רק כיצד מגיעים ילדים ממיצבים חברתיים-כלכליים שונים לרמות מומחיות שונות, אלא גם כיצד משפיעה ההגעה לרמת מומחיות מסוימת על התפתחות הרמה שאחריה וכיצד משפיעים גורמים שונים בסביבת הילד על רכישת המיומנויות הללו.

כעת נבחן את אחד האתגרים המרכזיים באיסוף נתוני אורך בתחום החינוך ושימוש בהם לצורכי ניתוח מדיניות – חיבור בין קובצי נתונים (File matching)<sup>13</sup>. השאלה היא כיצד ניתן לחבר בין קבצים ממקורות שונים כדי לקבל סט נתונים עשיר יותר לצורכי מחקר, ובהנחה שהדבר אפשרי – כיצד שומרים על האיכות, התקפות, והמיומנות של הנתונים בקובץ המחובר? כך למשל כדי לחקור צמיחה במדינות השונות ניתן להוסיף למידע המושג באמצעות מחקר פיז"ה מידע שמצוי בארגון ה-OECD על המדינות האלה. חיבור כזה מכונה יצירת מידע אורך סינתטי (synthetic longitudinal data).

דוגמה למידע אורך סינתטי הוא זה שנוצר מחיבור של ה-ECLS-K עם ה-ECLS-Birth, סקר אורך שעקב אחר ילדים החל מלידתם. מידע אורך ניתן לייצר במקרים כאלו גם אם הנסקרים בכל אחד מקובצי הנתונים אינם אותם נסקרים, וזה היה המצב בחיבור הזה. במקרה כזה יש לערוך חיבור בין אינדיבידואלים מהקבצים השונים בהתחשב במידע רקע עליהם – וחייב להיות ברשותנו מידע רקע רב ומדויק על הנסקרים. צריך גם

<sup>13</sup> האתגר המרכזי השני שעליו התכוון פרופ' קפלן לדבר היה השמירה על הפרטיות, אולם כיוון שביום העיון (וכן בדוח זה) הוקדש חלק נפרד לאתגר זה בחר פרופ' קפלן לדלג על חלק זה של הרצאתו.

להשתמש בשיטות סטטיסטיות כדי לבחון את תקפות המידע הקיים בקובץ הנתונים החדש. אם יש מידע משני קבצים – משתי קבוצות אנשים – נוצרת בעיה קשה של נתונים חסרים: המשיבים מהקובץ האחד לא ענו על השאלות בנקודות הזמן שבהן ענו עליהן המשיבים מהקובץ השני. זאת ועוד: חשוב לתת את הדעת על כל שינוי שחל בהקשר (או בהקשרים) – בעולם שבתוכו נאספו התוצאות, במיוחד במקרה שבו הקבצים נאספו בזמנים שונים. שינויי מדיניות, משברים כלכליים או הגירה יכולים למשל להשפיע מאוד על הנתונים באחד הקבצים, ולהפחית את אמינות המידע בקובץ המאוחד. לבסוף, צריך לחשוב מחדש לאיזו אוכלוסייה מתייחס הקובץ הסינתטי. אי אפשר להניח כמובן מאליו שהקובץ מתאים לאחת מהאוכלוסיות המקוריות, ודאי שלא לשתיהן.

חשוב לציין שכשמתמודדים עם מידע חסר יש לא מעט שיטות סטטיסטיות לעשות זאת.<sup>14</sup> כל אחת מהשיטות תניב תוצאות שונות מעט מהאחרות, ולכן צריך לבחון בשום שכל איזו שיטה מתאימה לאיזה קובץ נתונים. לאחר מכן יש לקבוע קריטריון לרמת התקפות של הקובץ: באיזו מידה הוא אכן משמר את המידע משני קובצי הנתונים המרכיבים אותו? רסלר (Rassler, 2002) מדברת על ארבע רמות של מהימנות:

- שמירה על ערכים ברמת הפרט – הרמה הגבוהה ביותר, שבה אנחנו מצליחים לזהות מה היה עונה פרט מסוים שמופיע בקובץ הנתונים הראשון לו היה עונה על השאלות שבקובץ השני. סביר להניח שאי אפשר להגיע לרמה זו של תקפות.

- שמירה על החלוקה המצרפית של הערכים

- שמירה על מתאמים בין ערכים – רמה חשובה מאוד להפקת מסקנות סטטיסטיות

- שמירה על החלוקה של הערכים השוליים – ערכי הקצה – שהיא הרמה המינימלית לצורכי היסק מחקרי כלשהו

כדי לבחון את השיטות הקיימות לחיבור נתונים ערכנו מחקר "הוכחת ייתכנות" על ידי חיבור של נתונים ממבחני טאליס (TALIS) ופיזה (PISA) בפינלנד. היתרון בשימוש בנתונים של פינלנד הוא שבתי הספר שם השתתפו בשני הסקרים הללו. הורדנו חלק מהמידע מכל אחד מהסקרים וחיברנו ביניהם כדי לקבל נתוני אורך על כל השנים. לאחר מכן בדקנו אם התוצאות שהגענו אליהן דרך השיטות הסטטיסטיות שבהן השתמשנו תואמות את התוצאות שהתקבלו בפועל מכל אחד מהסקרים האלה. בלי להיכנס לפרטים הטכניים, מצאנו כי לפחות אחת השיטות שהשתמשנו בהן כדי להתמודד עם המידע החסר פעלה היטב ושמרה על התקפות של הנתונים בסקרי טאליס ופיזה.

מהם האתגרים שטרם התמודדנו איתם בנוגע לחיבור נתונים?

- טרם התמודדנו עם חיבור של נתונים מרמות שונות, למשל נתונים על ילדים מצד אחד ונתונים על בתי הספר שלהם מצד שני.

- עדיין קיימות שאלות מתודולוגיות הנוגעות למשקלו של כל חלק מהמדגם ולאופן התאמתו לאוכלוסייה הכללית כך שהדגימה תישאר מייצגת לאוכלוסיית המחקר כולה.

- ושוב – עדיין יש אתגרים הנוגעים לשינויים שעלולים לחול בהקשר לאורך זמן, במיוחד אם משתמשים במידע להערכת מדיניות.

מנגד קיימים יתרונות לא מעטים בחיבורי נתונים כאלו, כפי שכבר הוצג לעיל.

<sup>14</sup> רשימה חלקית של שיטות אלו ניתן למצוא במצגת של פרופ' קפלן באתר היזמה למחקר יישומי בחינוך. לא הבאנו את הרשימה כאן על מנת לא להפוך את סיכום ההרצאה לטכני יותר על המידה.

סקרי אורך הם יקרים מאוד, וקשה מאוד למדוד בעזרתם את כל המדדים הרלוונטיים למחקר (גם אם עורכים רוטציה בין שאלות המחקר בשנים שונות). אולי מוטב לחשוב עליהם כעל חלק ממערכת של אינדיקטורים המעידים על רמת הבריאות של מערכת החינוך במדינה. האינדיקטורים הללו צריכים להיות רלוונטיים לתאוריות המדעיות הקשורות למטרותיה של מערכת החינוך במדינה, הם צריכים לאפשר מעקב אחר שינויים (או בחינה של שינויים מתוכננים) במדיניות החינוך, וצריך לעקוב אחריהם ולבדוק שהם אכן משתנים יחד עם שינויים החלים במדדים אחרים הנמדדים עבור מערכת החינוך במדינה (לעקוב אחר תקפותם לאורך זמן). אפשר להשתמש במערכת האינדיקטורים הללו למטרות של גיבוש מדיניות, כולל תחזיות לעתיד.

יחד עם זאת, אם מטרת מערכת האינדיקטורים היא יצירת תחזיות וניבויים, אזי חשוב שסקרי האורך יפקו לא רק מתוך מודל תאורטי אחד אלא מתוך מגוון של מודלים תאורטיים הנוגעים למושאי המדידה שלהם. אחד הדברים שניתן לעשות הוא לשלב בין מודלים סטטיסטיים מגוונים כדי לייצר תוצאות אופטימליות ותחזיות מדויקות לאורך זמן. ניתן גם לכייל את המודלים הללו בהתאם לדיוק של התוצאות שהם חזו לעומת התוצאות בפועל.

## הפוטנציאל של מחקרי אורך בתחום החינוך בישראל

משה יוסטמן

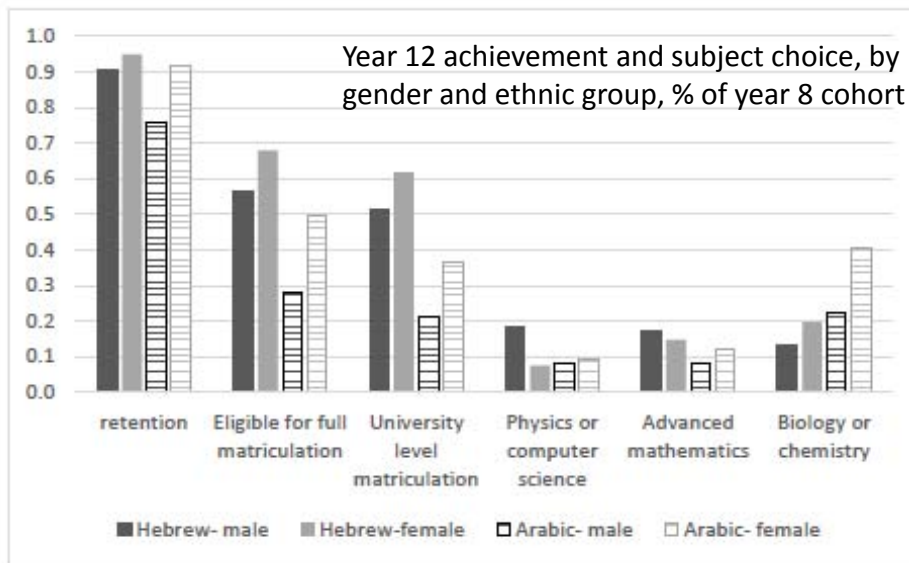
בהרצאה זו אדגים את ערכם של נתוני אורך בשני מחקרים שערכתי לאחרונה – בישראל ובאוסטרליה. מחקרים אלו עשו שימוש בנתוני אורך מינהליים הנמדדים לאורך זמן, שהם כאמור, אלטרנטיבה לסקרי אורך. בישראל נאספים נתוני חינוך רבים, ולעומת זאת קיימים מעט מאוד סקרי אורך ועוד פחות מכך – סקרי אורך בתחום החינוך.

הדוגמה הראשונה שאציג היא מחקר שערכתי עם ד"ר נעמי פרידמן סוקולר ובו נעשתה השוואת פערים מגדריים בבחירת מקצועות מדעיים, הן במגזר היהודי והן בערבי. גילינו שיש הבדלים גדולים בין יהודים לערבים במובן זה. נתוני האורך מאפשרים לשלוט בנתוני היכולת (שבאה לידי ביטוי במיצ"ב), להשוות בין נתונים של תלמידים לבין נתונים של תלמידות בעלות אותן יכולות ולראות מהן הבחירות שהם עושים ועושות אחרי כן (בבחירת מקצועות מוגברים בבגרות).

יש הבדלים מגדריים בין המקצועות של גברים ונשים: נשים מיוצגות פחות במקצועות ההיי־טק והדבר תורם גם לפערי השכר, בישראל ובעולם כולו. ההבדלים הללו משקפים הבדלים בבחירת המקצועות באוניברסיטה ועוד לפני כן – בבחינות הבגרות. בעבר טענו שההבדלים הללו הם תוצאה של פערים ביכולת המתמטית בין בנים לבנות, אולם מחקרים (כולל המחקר שעליו אנחנו מדברים כעת) מראים שאין זה כך ושהפערים הללו נשמרים גם כששולטים ביכולת המתמטיקה של הגברים והנשים (Riegle-Crumb et al., 2012; Friedman-). הסברים אחרים לתופעה זו מתייחסים לפערים הנובעים מיחס שונה של מורים כלפי בנים וכלפי בנות, ולהבדלים בשוק העבודה ובתמריצים – הלא־שוויוניים – שמעסיקים מעניקים לנשים ולגברים על הבחירה שעשו בשעתם במקצועות הבגרות. במחקר הנוכחי אנחנו שולטים גם במיצב החברתי־כלכלי שממנו מגיעים הנסקרים כיוון שידוע שחסכים הנובעים מפערים אלו משפיעים על בנים יותר מאשר על בנות. במחקר זה נערך ניסיון לגלות אם יש הבדלים הנובעים מרקע תרבותי. נבחנו הבדלים בין חינוך ערבי לחינוך עברי במשתנים הבאים:

1. שיעורי נשירה
2. זכאות לתעודת בגרות
3. בחירת מקצועות מוגברים לבגרות

המידע: קובץ נתונים שהוכן בחדר המחקר של הלמ"ס ובו מידע אורך שהושג על ידי חיבור בין נתוני המיצ"ב (מיצ"ב ח') לנתוני הבגרות של אותם תלמידים שהיו בכיתה ח' ב־2002 וב־2003. כיוון שהיה מדובר בנתונים מינהליים, כמות הנתונים החסרים הייתה קטנה יחסית (בהשוואה לנתונים הנאספים באמצעות סקרים), והשתמשו בשיטות סטטיסטיות כדי להשלים פרטים על תלמידים שלא השתתפו בשתי הבחינות. השמטנו רשומות של תלמידים שלא היה לנו ידע על ההשכלה של לפחות אחד מהוריהם. היתרון הגדול של שימוש בנתונים אלו הוא שיכולנו להשתמש במספרים גבוהים מאוד של משתתפים. טעויות הדגימה וההטיה של נשירה מהמחקר קטנות בהרבה מאלה שחלות בעבודה עם סקרים.



Ethnic Difference in Raw Gender Gaps

ממצאים ראשוניים תיאוריים בנוגע להתמדה ולבחירת מקצועות, **בלי לשלוט בגורם של יכולת במתמטיקה:**

- שיעורי התמדה בבית הספר – נשים מתמידות יותר מגברים בכל המדדים: יותר נשים מגיעות ל־12 שנות לימוד מאשר גברים, יותר נשים מגיעות לתעודת בגרות מלאה מאשר גברים ויותר נשים מגיעות לתעודת בגרות איכותית (ברמה המאפשרת כניסה לאוניברסיטה) מאשר גברים. נתון מעניין הוא שהפער הזה בין גברים לנשים, בכל שלושת המדדים, גדול הרבה יותר במגזר הערבי מאשר במגזר היהודי.
- בחירת מקצועות לימוד – אם בוחנים את המדד של בחירה בלימודי פיזיקה ומדעי המחשב רואים תוצאות מעניינות. במגזר היהודי יש פער גדול לטובת הגברים – 20% מהגברים במגזר זה בחרו בלימודים אלו לעומת 7% בלבד מהנשים. החלק המפתיע הוא שבחינוך הערבי המצב הפוך – נשים ערביות בוחרות ללמוד פיזיקה ומחשבים יותר מגברים ערבים. תמונה דומה מתגלה במדד של בחירה ללמוד מתמטיקה ברמה מוגברת של 5 יחידות לימוד – בחברה היהודית יותר גברים בוחרים ללמוד מתמטיקה ברמה מוגברת, ובחברה הערבית יותר נשים. ההפתעה כאן גדולה עוד יותר בהתחשב בכך שהחברה הערבית נחשבת לחברה מסורתית יותר ושמרנית יותר, והיינו מצפים שמאפייני המגדר יהיו מובחנים יותר מסיבה זו.
- בבחירה במקצוע הביולוגיה יש פער לטובת הנשים גם בחברה היהודית וגם בחברה הערבית, הפער גדול יותר משמעותית בחברה הערבית.



מנתוני אורך אנחנו יכולים ללמוד שחלק מהפער בבחירת המקצועות בחברה הערבית (אך לא כולו) משקף פער הישגים לטובת הבנות, פער שמופיע כבר במיצ"ב ח' – הבדלים גדולים בשפה והבדלים פחות גדולים, אבל עדיין משמעותיים, במתמטיקה.

בשלב הבא השווינו את הפער המגדרי בין יהודים לערבים אצל תלמידים ותלמידות בעלי אותם הישגים במיצ"ב. ראינו שגם אחרי ששולטים בגורמי הרקע החברתי-כלכלי והציונים במיצ"ב עדיין ההבדל בין בנות ערביות לבנים ערבים גדול ב-7% מאשר ההבדל בין בנות בחינוך העברי לבנים בחינוך העברי. לעומת זאת כשמדובר בהגעה לבגרות איכותית, אין כבר הבדל בין הפערים. זהו היתרון הגדול של נתוני אורך – הם מאפשרים לנו להבחין בערך המוסף של כל אחד משלבי הלימוד.

בבחירת הבחירה במקצועות מדעיים – פיזיקה ומדעי המחשב – אנחנו רואים שההבדל בין הפער המגדרי בחברה הערבית לבין הפער המגדרי בחברה העברית הוא 10 נקודות אחוז – הבדל משמעותי מאוד. בנוסף לכך קיים הבדל גם בפער המגדרי בבחירת מתמטיקה מוגברת – 4 נקודות אחוז, ובבחירת מקצוע הביולוגיה – 9 נקודות אחוז.<sup>15</sup> עובדות אלה נכונות כאמור אחרי שמביאים בחשבון רקע חברתי-כלכלי וציון במיצ"ב.

נקודה שמן הראוי לציין ויש לה חשיבות להשלכות המחקר – אנחנו לא רואים שהיתרונות הללו של בנות ערביות נשמרים גם ברמת האוניברסיטה. נקודה זו אמנם לא נבדקה במחקר הנוכחי אבל חשוב להזכירה.

דוגמה שנייה למחקר אורך מבוסס נתונים מינהליים הוא זה שנערך באוסטרליה בשיתוף עם ברנדן הואנג (Houng & Justman, 2014). במחקר זה יצרו החוקרים קובץ נתונים מתוצאות מבחני NAPLAN (The National Assessment Program – Literacy and Numeracy), שנערכים באוסטרליה לכל התלמידים בכיתות ג', ה', ז', וט', ומתוצאות מבחני הבגרות שם (בכיתה י"ב). בנוסף יצרו החוקרים סיווג של מיצב חברתי כלכלי בהתאם להשכלת ההורים ומשלח היד שלהם. המחקר התייחס לשנתונים של 2009 ו-2010 (בכיתה ט'). תחילה נעשה שימוש בנתוני 2009 כדי ליצור מודל שמנבא, לפי הישגי התלמיד במבחני ה-NAPLAN ולפי המיצב החברתי-כלכלי של התלמיד, את הסיכוי שלו להגיע להישגים שונים (החציון העליון של התפלגות ההישגים בבגרות, 30% עליונים, ו-10% עליונים). לאחר מכן נבחן המודל הזה על השנתון של 2010 ותחזיותיו הושוּו לביצועים בפועל.

כך למשל בחנו את כל הילדים שהסיכוי שלהם לפי הניבוי להגיע ל-50% הגבוהים היה 20% ובדקו כמה מהם אכן הגיעו לשם. מצאו שכמעט לכל קבוצה (1%, 2%...60%...99%) התחזית אכן הייתה קרובה מאוד למציאות. במילים אחרות, לגבי כל ילד, אם יודעים את תוצאות הבחינות בכתה ט' ואת הסיווג הסוציו-אקונומי שלו, יש לנו הערכה טובה על סיכויי ההצלחה שלו בבגרות.

בדומה למחקרים קודמים רבים מצאנו גם במחקר שלנו שלרקע הסוציו-אקונומי יש השפעה ניכרת על סיכויי ההצלחה, וכן שיש מתאם גבוה בין תוצאות מבחני ה-NAPLAN בכיתה ט' לבין תוצאות הבגרות. ההשלכה של המודל הזה לגבי מדיניות היא שניתן להשתמש במודלים כאלה כדי לתת ייעוץ טוב יותר לילד על המשך דרכו בחטיבה העליונה. כמוכן, כל ילד הוא מקרה בפני עצמו, אולם לילד בעל הישגים נמוכים בכיתה ח' יש באופן כללי סיכויים נמוכים להצליח בבגרות, וזה נתון שטוב להביא בחשבון. ייתכן שיש מקום לילד ולהוריו לחשוב על מסלול חלופי לבגרות, או להעריך נכון יותר את המאמץ הדרוש כדי להצליח.

ניתן להשתמש במודלים כאלה לבחון כיצד התלמיד מתקדם על פני זמן (באוסטרליה נערכים מבחנים בכיתה ג', ה', ז' ו-ט') בהשוואה לתחזיות שלו. אפשרות אחרת: לבחון כיצד מצליחים התלמידים בבתי ספר שונים והאם יש הבדלים בין בתי הספר השונים באחוזי התלמידים שלהם שמצליחים (יחסית לתחזיות של המודל).

<sup>15</sup> חשוב להדגיש שהכוונה אינה שהפער בין בנות ערביות לבנים ערבים גדול ב-10% מאשר הפער בין בנות בחינוך העברי לבנים בחינוך העברי (דוגמה מספרית: 11% לעומת 10%), אלא שהוא גדול ב-10 נקודות אחוז – 20% לעומת 10%.

## השפעות בטווח הארוך של התערבויות חינוכיות בבתי ספר יסודיים ותיכונים בישראל

### ויקטור לביא

בהרצאתי זו אני מבקש להמחיש, בעזרת מחקרים עכשוויים, את חשיבותם של נתוני אורך להערכת הצלחה של התערבויות, את האופן שבו התערבויות כאלה יכולות לעצב את התפתחותם של פרטים, וכן את חשיבותם של מחקרי אורך להערכת מדיניות והשפעתה. המחקרים הללו מבוססים על מעקבים שנערכו בטווחים קצרים יותר (3 שנים, 5 שנים) ועכשיו, בשנה האחרונה, בדקנו את התוצאות בטווח הארוך יותר.

בשנים האחרונות מנסים כלכלנים יותר ויותר לבחון התערבויות ומדיניות לא (רק) לאור הצלחת המטרות הראשוניות של ההתערבויות עצמן והמדיניות עצמה, אלא לאור השפעתן על רווחת הפרטים בטווח הארוך – בבגרותם. האם הן גורמות שיפור ברמת החיים של הפרט ובתרומתו לחברה ולמשפחה שלו? מנסים להבין הקשרים סיבתיים (לא מתאמיים) בין התערבויות ושינויי מדיניות לבין רווחת הפרטים בטווח הארוך. המחקרים בודקים לא רק את מטרות ההתערבות אלא את השפעתה הכוללת על הפרט. יריעת משתני התוצאה רחבה מאוד. יש עבודות רבות המתמקדות בהשפעה של איכות עמיתים, גודל כיתה, גודל בית ספר וכדומה – על פרטים בטווח הארוך.

יחד עם עמיתים למחקר השתמשנו בנתונים מינהליים כדי להוסיף נדבך של התבוננות עיונית יסודית על תוצאות בטווח הארוך של התערבויות. הנתונים מגיעים מהלמ"ס, ממשרד החינוך ומהביטוח הלאומי. המחקרים משתמשים באותם מסדי נתונים:

- קובץ מס בריאות – מאפשר לדעת על כל פרט, לגבי כל שנה מחייו (מאז 1995) אם למד במוסד חינוכי כלשהו, גם על-תיכוני. כל מוסד חינוכי חייב להעביר לביטוח הלאומי מידע על הלומדים בו כיוון שהתלמידים מקבלים הנחה בתשלום מס הבריאות.
- קובץ מס הכנסה – שם אנו יכולים לקבל מידע על הכנסותיו של כל פרט ופרט, האם עבד או לא עבד וכמה חודשים עבד בשנה? מכאן ניתן גם להסיק על הכנסותיו של בן הזוג או בת הזוג של אותו פרט ועל הכנסות ילדיו.
- מרשם האוכלוסין מאפשר לדעת על כל פרט את כתובתו ומצבו המשפחתי וכן לחבר את הנתונים עם נתונים על הכנסת הוריו.

יתרונות הנתונים המינהליים: הם זמינים ומיידים, ועלות השגתם זולה יחסית. החיסרון הוא שלא ניתן להסיק פרטים כגון לאיזה תואר למד הסטודנט ובאיזו מכללה.

מחקר ראשון – השפעת תכנית הבחירה החופשית של בתי ספר על תלמידים בעיר תל אביב. תכנית זו נכנסה לביצוע ב-1993-1994, ובמסגרתה קיבלו תלמידים מרובע 9 בתל-אביב (אוכלוסייה חלשה מאוד) אפשרות לבחור לעצמם בית ספר במעבר מבית הספר היסודי לחטיבת הביניים והתיכון. מחקר קודם שהתפרסם לפני שש שנים בדק איזו השפעה בטווח הקצר והבינוני הייתה לבחירה החופשית הזאת על תלמידים עד תום לימודיהם. התוצאות הראו שהייתה השפעה חיובית משמעותית על שיעור ההתמדה שלהם בלימודים, על שיעור זכאותם לבגרות, על מספר יחידות הלימוד שלמדו, ועל הפחתת החשיפה שלהם לאלומות.

המחקר שערכנו בשנה אחרונה, ואותו אני רוצה להציג כעת, בדק אם אותם שינויים בטווח הקצר והבינוני השפיעו (סיבתי) בטווח הארוך על משתני תוצאה. קיבלנו מספר תוצאות מעניינות:

- בשנים 2000-2011 יש עלייה מתמדת באחוז הלומדים באוניברסיטאות, אולם אין הבדל בין אלה שהשתתפו בתכנית לבין קבוצת הביקורת. בטווח הארוך השפעת התכנית היא אם כן אפסית. תמונה דומה אנחנו רואים גם בסך הכול של שנות הלימוד.

■ בשיעור הלומדים במכללות, לעומת זאת, אנחנו רואים שהתכנית הביאה לעלייה של חמש נקודות אחוז<sup>16</sup> או 25% בלמידה במכללות לעומת קבוצת הביקורת. השפעה זו לא נראתה בשנים הראשונות, אבל אחרי כמה שנים היא נראית.

■ אם בוחנים את השכר הממוצע של משתתפי התכנית, רואים שההשפעה בשנים הראשונות היא אפסית. בטווח הארוך, לעומת זאת, יש השפעה חיובית של כמעט 3000 ש"ח יותר (שהם כ-3%-4%) בשנה.

המחקר השני שהוצג בוחן את ההשפעות לטווח הארוך של ניסוי מבוקר (שנעשה בשיתוף פעולה עם משרד החינוך בשנת 2001–2000) על תכנית שתגמלה מורים על סמך הישגי התלמידים. הישגי המורים נבחנו על סמך הישגים לימודיים של תלמידיהם בבגרות. הניסוי התחיל בינואר והסתיים בבחינות הבגרות ביוני. השתתפו בו בתי ספר שרמתם הייתה דומה לרמה הממוצעת בכלל בתי הספר בישראל אולם ההישגים הממוצעים של תלמידיהם בתחומי המתמטיקה והאנגלית היו חלשים יחסית.

בניסוי השתתפו כל מורי המתמטיקה והאנגלית שהגישו תלמידים לבגרות באותה שנה, ובקבוצת הביקורת היו מורים שהגישו תלמידים לבגרות מבית ספר אחר. בטווח הקצר נמצאו תוצאות חיוביות בשיעור הניגשים לבחינת הבגרות במתמטיקה ואנגלית, בשיעור העוברים לעומת הנכשלים, ובציון הממוצע בבחינות הבגרות.

המחקר שהוצג ביום העיון בחן תוצאות ארוכות טווח במגוון מדדים: למידה במוסד אקדמי, תעסוקה, שכר, גיל חתונה, לידת ילדים והשכלת בת הזוג או בן הזוג. מדדנו גם בעזרת נתוני הביטוח הלאומי את ההשפעה על קבלת קצבאות מביטוח לאומי: דמי אבטלה, הבטחת הכנסה וכיוצא באלו. להלן הממצאים:

■ שיעור למידה באוניברסיטה: מספר הלומדים עלה משנה לשנה. בטווח הארוך נמצאה השפעה חיובית של כמעט 5 נקודות אחוז (או כ-25%) על שיעור הלומדים באוניברסיטאות לעומת קבוצת הביקורת.

■ מספר שנות הלימוד: מצאנו שהתכנית העלתה את מספר שנות הלימוד ב-17 או 18 נקודות אחוז.

■ השכר: גם כאן השפעת התכנית בהתחלה היא שלילית, כיוון ששיעור הלומדים באוניברסיטאות גבוה יותר. לעומת זאת בטווח הארוך נמצאה השפעה חיובית משמעותית של כמעט 6,000 ש"ח נוספים על השכר השנתי הממוצע. חשוב להבין שתוצאה זו יכולה להתקבל רק באמצעות נתוני אורך וכמעט רק באמצעות נתונים מינהליים – לא קיים בישראל מחקר אורך כיום שפעל במשך זמן כה רב.

■ ממצאים נוספים: נמצא שהתכנית מביאה לדחיית גיל הנישואין בכחצי שנה, וכן מביאה לדחיית לידת הילד הראשון, גם אצל בנים וגם אצל בנות.

המחקר מצביע שוב על חשיבותו של שימוש בנתונים מינהליים ושל שילוב נתונים מינהליים עם סקרים לצורך מחקרי אורך. מחקרים אלו יכולים להיות משמעותיים מאוד למדיניות ולשפר מאוד את יכולתו להבין את התוצאות של מדיניות חינוכית ותכניות התערבות.

מחקר שלישי שנערך, הפעם לא כמחקר הערכה של תכנית התערבות, בחן השפעות ארוכות טווח של תוצאות במבחנים בעלי חשיבות גבוהה (High-Stakes) – בחינות הבגרות.

כידוע, בישראל נבחנים התלמידים בתקופת זמן קצרה יחסית במספר גבוה יחסית של מקצועות בבגרות. לבחינות אלו חשיבות רבה לעתיד התלמידים, הן משפיעות על המקצועות שיהיו פתוחים בפניהם ללימודים גבוהים, וכך גם על המשך חייהם. השאלה הייתה אם השימוש במבחנים גורם לכך שאירועים אקראיים (ויכוח עם ההורה, מחלה, ארוחת בוקר לא מלאה) שקורים לפרט ביום הבחינה ישפיעו על התפתחותו

<sup>16</sup> ראו הערה 13 בנוגע למשמעותן של נקודות אחוז.

העתידי עד גיל מבוגר. תשובה חיובית במקרה זה אולי תאמר משהו על ההוגנות והיעילות של השימוש במבחנים כאלו.

על מנת לבדוק השפעה של גורם אקראי ולהבחין בינה לבין היכולת של התלמיד בדקנו מצבים שבהם הפרט נחשף לזיהום אוויר ביום הבחינה. מחקרים אפידמיולוגיים מראים שזיהום אוויר משפיע לרעה על התפקוד הקוגניטיבי של הנחשף לו (ולא ניכנס כאן לשאלה מדוע – במאמר מדובר על כך בהרחבה כמובן). בעזרת הנתונים המינהליים יכולנו לדעת מתי נבחן כל פרט בבחינת הבגרות. החלק השני של הפאזל היה נתונים מהמשרד לאיכות הסביבה, שהראו מדדים של זיהום אוויר בימים השונים של הבחינות. כך התאפשר לנו לבדוק את אופן השפעתה של החשיפה לזיהום אוויר על ההישגים בבחינה.

בהשוואה להישגים בבחינות בגרות אחרות יכולנו לראות שקיים קשר סבתי שלילי מובהק בין רמת החשיפה לזיהום אוויר ובין ההישגים בבחינות הבגרות של הנחשפים לו. הקשר התברר בעיקר מבדיקת ימים שבהם הייתה רמה גבוהה של זיהום אוויר, למשל יום אחרי ל"ג בעומר. ההשפעה, כך הסתבר, קטנה ברמות הנמוכות והבינוניות של זיהום האוויר, אך גדלה משמעותית ברמות הגבוהות שלו.

החלק המשמעותי עוד יותר של המחקר הראה שהחשיפה הזו לזיהום אוויר ביום הבחינה, כשהתלמיד בן 18, היא בעלת השפעה שלילית משמעותית על מדדים שונים בטווח הארוך: כשהתלמידים שנבחנו בגיל 18 הגיעו לגיל 32, נמצא אצלם קשר שלילי מובהק עם לימודים באוניברסיטה ומספר שנות לימוד, וכן השפעה סיבתית שלילית על שכרם בגיל 32.

מדינות כגון דנמרק שוודיה, נורבגיה וארצות הברית מנגישות כיום נתונים מינהליים לחוקרים במהירות וביעילות, כמובן תוך שמירה על פרטיות מושאי הנתונים. כתוצאה מכך ניתן לראות שרוב המחקרים שהתפרסמו בירחוני הכלכלה המובילים מגיעים מארצות אלה. חשיבותם של מחקרי אורך כאלה מבחינת האפשרות להעריך מדיניות ולייצר מדיניות מועילה – רבה ביותר.

תודה למשרד החינוך על המעבדה הווירטואלית שלו, ולשאר הגופים המנגישים נתוני חינוך בישראל. כל הקבצים המינהליים של המשרד נמצאים במאגר ברמה הגולמית ואפשר גם לקשר ביניהם על פי מזהה פיקטיבי לכל אדם. אפשר לגשת לנתונים מרחוק בלי להגיש בקשות לקבצים. זהו מודל שכדאי להעתיקו גם למשרדים אחרים, הוא מקצר מאוד את הדרך לתחילת מחקר. גם הביטוח הלאומי, למרות לא מעט קשיים, מנגיש קבצים מינהליים בחדר מחקר פיזי תוך שמירה על פרטיות ומניעת זליגה של מידע.

## פרק שני -

# נתונים מינהליים הנמדדים לאורך זמן בנושא חינוך בישראל

### סקירת מצב: הנגשת נתונים מינהליים בחינוך על ידי הלמ"ס ומשרד החינוך

ישראל עברה בעשור האחרון כבדת דרך ביצירת נתוני אורך על בסיס קבצים מינהליים ובשימוש בהם מתוך הכרה בחשיבותם לצורכי מחקר ועיצוב מדיניות ובאפשרויות המחקר השונות הנובעות מהם. קיימת היום מוכנות רבה בקרב השותפים ל"סטטיסטיקה הלאומית" (אנשי ביטוח לאומי, מערכת החינוך, גורמי הרווחה, משרד הכלכלה, משרד ראש הממשלה ועוד)<sup>17</sup> לשתף פעולה כדי שיהיה מאגר מידע מקיף גם לצורכיהם של קובעי מדיניות וגם לצורכי מחקר.

מוכנות זו נובעת גם מ"Sustainable development agenda", האג'נדה של האו"ם שנוסחה בינואר 2016 ומציבה 17 מטרות עולמיות לשנים 2015-2030. רוב מדינות העולם, כולל ישראל, חתמו על האג'נדה הזאת. כדי לנטר את ההתקדמות בהשגת המטרות נבנתה מערכת של 240 אינדיקטורים סטטיסטיים. כשליש מהאינדיקטורים ניתנים למדידה כבר עכשיו, בשליש נוסף – ידוע איך למדוד אבל לא מודדים, ובשליש השלישי חסרה גם המתודולוגיה לייצור מדידה. ה"מקל" שמציע האו"ם הוא שאם הגוף המרכזי של הסטטיסטיקה במדינה לא מספק נתונים, האו"ם פונה למכוני מחקר וחוקרים פרטיים כדי לקבלם. מסיבות מובנות קביעה כזאת מתמרצת את המדינה למדוד בעצמה את האינדיקטורים הללו כדי לדווח עליהם בעצמה.

ישראל עובדת על פי המודל האירופי שמבוסס על גוף סטטיסטי לאומי ריכוזי, במקרה של ישראל – הלמ"ס. קיימת היום הבנה שכדי לייצר סטטיסטיקה לאומית טובה, חובה לשתף פעולה גם עם הגורמים שאינם חלק מהגוף הלאומי הזה.

### הנגשת נתוני הלמ"ס לצורכי מחקר

הלמ"ס מתכוונת להנגיש נתונים לחוקרים באמצעות: (1) חדרי מחקר (2) מערכת של גישה מרחוק (שנמצאת כרגע בתהליך של בנייה מחדש).

**חדרי מחקר של הלמ"ס:** החדר הראשון נמצא בירושלים במשרדי הלמ"ס. בנוסף נבנו לאחרונה חדרי מחקר בבנק ישראל ובמשרדי הלמ"ס בחיפה. בימים אלו עובדים על פתיחת חדר מחקר נוסף, במשרדי הלמ"ס בתל-אביב. בחדרי המחקר הפיזיים ניתן לחוקרים מוכרים (שקיבלו אישורים לכך) להשתמש בקבצים ברמת סודיות גבוהה, וכמעט אין מגבלות על הנתונים המונגשים.<sup>18</sup> המגבלות העיקריות על יצירת קבצים בחדרי המחקר היא כוח אדם ותקציב. כרגע אין מי שתפקידו המוגדר הוא יצירת קבצים כאלה. הלמ"ס מקצה אנשים (שיש להם תפקידים אחרים שם) לפי הצורך והזמינות.

**מערכת הגישה מרחוק:** בלמ"ס מפתחים מערכת של גישה מרחוק. נכון לרגע זה (נובמבר 2016) המערכת אינה פעילה כיוון שקרסה תחת עומס הבקשות ממנה. במערכת זו סוג הקבצים יהיו קובצי (Micro) MUC (Data Under Contract) מותממים – נתונים שלא ניתן לזהות באמצעותם אדם כזה או אחר, אבל ייתכן שאם יחברו קבצים שונים, יתאפשר זיהוי כזה. אפשרות כזאת לא קיימת למשל בקובצי (Public Use) PUF (Files), שאינם מאפשרים זיהוי כלל. היות שקובצי MUC אינם נקיים מחשש לזיהוי, תהיה חובה לקבוע

<sup>17</sup> במערכת הבריאות המצב מורכב יותר כיוון שהנתונים מורכבים יותר ונמצאים בידיים רבות יותר, אולם גם שם קיימת הבנה רבה יותר של חשיבות הנגשת הנתונים לצורכי מחקר, ונעשית עבודה מקיפה על מנת לאפשר זאת.

<sup>18</sup> למידע על קובצי הנתונים הקיימים בלמ"ס ראו: רן, 2016.

תנאים שיגבילו את השימוש במערכת לחוקרים מוכרים בלבד. הנחת העבודה היא שקבצים אלו יתנו מענה לליבת המחקר של מדינת ישראל.

זמן ההמתנה לקבלת קובץ מהלמ"ס עדיין ארוך למדי (בין חצי שנה לארבע שנים) וגם המחיר גבוה יחסית. אחרי קבלת הפנייה מחוקר, תהליך קבלת האישור לְגֶשֶׁת לנתונים כולל את השלבים הבאים:

■ קבלת אישורים מאנשי הלמ"ס האלה:

- יועץ משפטי
- מנהל מאגר הנתונים
- ועדת סודיות
- ממונה על אבטחת מידע

■ קבלת אישורים מספקי הנתונים: המוסד שממנו הגיעו הנתונים חייב לאשר את העברתם.

לאחר קבלת האישור ממנה הלמ"ס לעבודה זו איש קשר, שאחראי על תמחור העבודה וחתימה על חוזה מול החוקר, על הכנת נתונים על ידי היחידות הנושאות ועל בדיקת תוצרים לפני העברתם מחדר המחקר. כאמור, חוקרים רבים סבורים שמדובר בתהליך יקר יחסית, שאורך זמן רב מדי.

פעולות הלמ"ס לשיפור יעילות התהליך:

■ בניית קטלוג של מסדי נתונים: הלמ"ס בונה בתחום החינוך קטלוג של מסדי נתונים שכבר קיימים. מדובר במסדים שהלשכה כבר בנתה, והקטלוג יהיה זמין באתר לצורכי החוקרים. קיומו של הקטלוג אמור למנוע עבודה כפולה (במקרה שהחוקר מבקש קובץ נתונים אך זה כבר קיים או קיים ברובו), ולהגדיל את הביקוש לקבצים הקיימים – וכך להפחית את הביקוש לקבצים חדשים שהכנתם דורשת השקעת משאבים נוספים.

■ פרסום meta-data של קבצים: הגדרות נתונים בלמ"ס שונות מאלה הקיימות בגופים אחרים ובהם משרד החינוך. הדבר יוצר קושי לחוקרים ועלול ליצור בלבול ואף להוביל לממצאים שגויים. דוגמה לכך היא מדד מחירי הדיור בישראל: משרד השיכון מודד את המחיר החציוני שבו נמכרה דירה בכל עיר ועיר, השמאי הממשלתי מודד את המחיר הממוצע שבו נמכרו דירות 4 חדרים בארץ, והלמ"ס מודדת מצד אחד את מחירי הדירות כפי שהם עולים מסקר מחירי דירות שערכה, ומצד שני את מחירן הממוצע של דירות לפי עסקה, אזור מגורים וגודל דירה. הלמ"ס רוצה לקחת על עצמה את האחריות לאיחוד המדדים השונים או, למצער, לריכוזם ברשימה אחת בצורה ברורה ונגישה.

■ בניית מערכת האיזופד: מערכת האיזופד (ISOPEd - Integrated System Of Processing Education Data), שנמצאת בבנייה, תהפוך נתונים פזורים למידע שמיש למחקר ולמטרות אחרות. כיום הנתונים באגף חינוך בלמ"ס, ולא רק בו, נמצאים בקבצים פזורים. יש יחידות נושאות שבהן כל עובד אחראי על נושא מסוים, ויש יחידה טכנולוגית. באגפי הלמ"ס מקבל כל חוקר קובצי נתונים בהתאם לצורך שלו. אם עובד בלמ"ס (קל וחומר חוקר חיצוני) רוצה למשל לחבר בין נתוני השכלה לנתוני הכנסה, הוא צריך לקבל את הקבצים הללו ממקורות שונים בתוך הלמ"ס, לברר מה ההגדרות המדויקות של כל אחד מהנתונים, ולערוך את החיבור ביניהם באופן "ידני". מערכת האיזופד תאפשר קליטת קבצים בפורמטים שונים, בדיקה, טיוב, סטנדרטיזציה, קישור וטעינה לתוך בסיס נתונים טבלאי מקושר, והעברה לבסיס נתונים טבלאי מקושר. הדבר יהפוך נתונים פזורים, בלי מטא-דאטה ובלי ארגון, לבסיס נתונים שממנו ניתן להפיק מידע בקלות וביעילות.

- השתתפות ב'מיזם התאומות' של האיחוד האירופי: הלמ"ס נמצא בקשר עם אנשי הסטטיסטיקה הלאומית של דנמרק על מנת ללמוד משיטות העבודה שלהם מבחינת הסטטיסטיקה הלאומית, בכלל זה עבודה משמעותית בנושא החינוך.
- פגישות עם גורמים רלוונטיים כדי לאבחן מהם פערי המידע וגם להגיע להסכם הבנות בין הגופים: הסכמי הבנות קיימים כבר עם בנק ישראל והביטוח הלאומי. המטרה היא לשלב בין מאגרי הנתונים ולבנות מערכות נתונים שתמשנה גם למחקר והערכת מדיניות, וגם, בעתיד, תספקנה יותר ויותר מצורכי הסטטיסטיקה של ישראל באמצעות נתונים מינהליים. בעתיד ננסה לשלב גופים נוספים בעבודה זו, כולל מערכת החינוך שהיא שותף טבעי לתהליך זה.

### הנגשת נתוני משרד החינוך לצורכי מחקר

משרד החינוך מנגיש חלק מהנתונים שברשותו באמצעות חדר המחקר של המשרד, שנמצא פיזית במשרד משרד החינוך בתל אביב אך הגישה אליו אפשרית גם מרחוק: החוקר (שקיבל לכך אישור מבעוד מועד, כמובן) עובד ממחשב מביתו ומתחבר למערכת של משרד החינוך (בלי אפשרות לשמור קבצים למחשבו האישי). המשרד מנסה לקדם שיתופי פעולה עם משרדים אחרים ועם הלמ"ס, אולם קיימות בעיות טכניות, בעיות משפטיות ובעיות של אינטרסים שאינן מאפשרות זאת לפי שעה. נדרשת עבודה פוליטית. המשרד פועל על מנת לרכז את כל איסוף הנתונים דרך מערכות המידע שלו, אם כי ברור (גם לאנשי המשרד) שקיימת עדיין דרך ארוכה עד שהדבר ייעשה.

הדרך של המשרד להנגיש מידע לחוקרים היא כאמור באמצעות חדר מחקר וירטואלי. יש כרגע כ-100 חוקרים שמשתמשים בחדר המחקר, ואדם שעובד בהיקף של חצי משרה כדי לתפעל אותו. קיימת מגבלת גודל: האפשרות של החדר להכיל חוקרים היא מוגבלת, ולכן המשרד ממעט "בפרסום" האפשרות הזו. הנתונים בחדר המחקר מוצפנים באמצעות מזהה ייחודי (unique identifier) והם גם עוברים בדיקה שנועדה לוודא שלא נוצר פרופיל של מאפיינים המאפשר זיהוי של מוסד או אדם מסוים (הוועדה נקראת 'ועדת סודיות ממוחשבת', ובעניין זה ר' הרחבה בפרק על התממת נתונים).

### תקציר סקירה מדעית: מיפוי מאגרי מידע בישראל בתחום החינוך<sup>19</sup>

#### עמליה רן וליאת יוספסברג בן-יהושע

סקירה זו עוסקת במאגרי המידע השוטפים וההיסטוריים, הקיימים בישראל והנוגעים למערכת החינוך הפורמלית בכל הגילים: החל מהגיל הרך, בתי ספר של החינוך היסודי והחינוך העל-יסודי, וכלה במערכת ההשכלה הגבוהה. כן עוסקת הסקירה במערכת החינוך הבלתי פורמלי. הסקירה ממפה את המידע שהופק ממקורות מינהליים, מסקרים שוטפים ומתשתיות מידע הקיימות בידי רשויות ציבוריות, גופים פרטיים ומלכ"רים. הסקירה כוללת מקורות מידע כגון קבצים מינהליים וכן טפסים וסקרים המכילים פעולות סטטיסטיות, אך היא מתמקדת במאגרי מידע המכילים את כל הפרטים הבאים או חלק מהם: מידע על תלמידים והישגיהם, מידע על מורים וכוחות עזר, מידע על מוסדות חינוך, מידע על תקציבים, מידע על תשתיות, חינוך בלתי פורמלי ומועמדים לשירות ביטחון.

<sup>19</sup> הסקירה נכתבה בהזמנת צוות המומחים ובפיקוחו, הסקירה המלאה נמצאת באתר היזמה.

מטרתה של הסקירה היא למפות באופן מקיף ככל הניתן את מאגרי המידע הקיימים בישראל והעוסקים בחינוך. לשם כך התמקדה הסקירה בשלושה סוגי מאגרים: מאגרים פתוחים ונגישים לציבור הרחב, מאגרים חסומים-למחצה הדורשים הרשאות כניסה, ומאגרים מסווגים שאינם פתוחים לקהל.

המידע שהופק ממקורות מינהליים, מסקרים שוטפים ומתשתיות מידע הקיימות בידי רשויות ציבוריות, גופים פרטיים ומלכ"רים – מוצג באופן הבא:

- מיפוי גופים מרכזיים בישראל ותיאור מאגרי הנתונים הקיימים אצלם, בכלל זה נתונים ברמה הארצית או נתונים המבוססים על מדגמים המייצגים את כלל האוכלוסייה הנחקרת.
- מיפוי מאגרי מידע נוספים המצויים בידי רשויות מקומיות גדולות, רשתות חינוך ומלכ"רים.
- מיפוי מאגרים נוספים שנוצרו עבור מחקרים העוסקים בחינוך. מאגרים אלו יצוינו ברשימה ביבליוגרפית של המחקרים האלה.

כל המאגרים שנסקרו במסגרת זו, נבחרו בהתאם לקריטריונים הבאים:

- מאגרים בתחום החינוך, הכוללים רשומות של תלמידים, מורים, מוסדות חינוך וכיו"ב
- מאגרים המכילים פרטים מזהים של יחידת החקירה (כגון מספר ת.ז. של התלמיד; סמל מוסד או שם מוסד וכיו"ב), שבאמצעותם ניתן לקשר בין רשומות ממאגרי מידע שונים
- נתונים הניתנים לעיבוד ממוחשב

בסקירה מופו 120 מאגרי מידע פתוחים ופתוחים-למחצה, ונבדקה רשימה של 35 מאגרים נוספים, סגורים, המצויים בידי משרד החינוך ואינם פתוחים לקהל הרחב. הסקירה תציג בקצרה את מאגרי המידע שמופו באמצעות שאלון. את פרטיו מילא צוות החוקרים מטעם מכון מופ"ת או איש הקשר הרלוונטי בכל אחד מן הגופים. מאגרי מידע הכוללים מידע מסווג או לא נגיש מסיבות שונות יתוארו בקצרה אף הם. מאגרי מידע קיימים שלא נסקרו בגלל חוסר שיתוף פעולה מצד הגופים הרלוונטיים צוינו לפי שם המוסד המחזיק במאגר ואיש הקשר הרלוונטי, ונמנו הנימוקים לכך שלא נסקרו במלואם. על מנת להקל את תהליך החיפוש בקטלוג ממפה זה, מופיעים כל המאגרים בסדר אלפביתי לפי שלוש הקטגוריות שצוינו לעיל. כן ניתן להיעזר ברשימת האינדקס המופיעה בסוף הסקירה כדי לאתר מאגר מידע ספציפי.

לצורך הכנת סקירה זו, נעשה שימוש במספר אמצעי איסוף, על מנת להגיע למאגרי המידע ולתארם. בשלב הראשוני בוצע מיפוי של גופים ומוסדות שיש להם נגיעה לתחומי החינוך. מכתבי פנייה רשמיים מטעם היזמה למחקר יישומי בחינוך של האקדמיה הלאומית למדעים ומטעם צוות החוקרים של מרכז המידע במכון מופ"ת נשלחו לאנשי הקשר ב-70 מוסדות נבחרים. בשלב השני בוצעו מעקב טלפוני והתכתבות באמצעות דואר אלקטרוני עם הגופים השונים, וכן נקבעו שיחות עבודה עם הנציגים הרלוונטיים על מנת להשלים את תהליך איסוף המידע. במקביל לכך אותרו ואופיינו מאגרי מידע, תהליך שהתבצע על ידי חיפוש מעמיק במרשתת באתרים שונים הנוגעים לנושא ועל ידי בחינה של טופסי רישום ומקורות נוספים כגון דוחות שנתיים, תקצירים ומחקרים אקדמיים.

לאחר שלב איסוף החומר התבצעה הערכה של המידע הקיים במאגר, במטרה לבחון אם הוא עומד בדרישות המצוינות לעיל בכל הנוגע לשלמות המידע. מאגרי מידע אשר ענו על הקריטריונים שצוינו לעיל מהווים חלק מסקירה זו וממסד הנתונים הכולל. לבסוף, לאחר הערכת המידע, נסקרו המאגרים הרלוונטיים לפי קריטריונים זהים, והם מהווים את ליבת הסקירה הזו. קריטריונים אלו כוללים נתונים כגון שם המאגר, מחזיק



המאגר, מועד ההתחלה של איסוף הנתונים, מועד אחרון לאיסוף נתונים, תדירות עדכון הנתונים, אופן עדכון הנתונים, אוכלוסייה, יחידת החקירה, מספר רשומות, משתנים, מקור המידע וכיו"ב.

יש לציין, כי סקירה זו עוסקת בנושא רחב ביותר, הנוגע לתחומי חיים רבים ולמספר גופים עצום. מרבית המידע אינו זמין לאדם הפרטי, ולכן נדרשה פנייה בשם האקדמיה הלאומית למדעים והיזמה למחקר יישומי בחינוך אל בעלי תפקידים רלוונטיים המופקדים על מאגרים. תוצאות סקירה זו נשענות אפוא על נכונות שיתוף הפעולה של גופים אלו, שפעלו כמובן בכפוף לחוק חופש המידע. מאגרי מידע של מוסדות וגופים אשר לא השיבו על פניות חוזרות ונשנות מטעם צוות המחקר לא נכללים במסגרת זו בגלל הסירוב לשיתוף פעולה.

חשוב להדגיש שהסקירה הנוכחית אינה מנתחת באופן מחקרי את טיב המאגרים ואינה עוסקת באופי הנתונים הנאספים בהם. מלאכת המחקר עבור סקירה זו מתבססת על חיפוש מאגרי מידע בחינוך באמצעות פנייה ישירה לגורמים הרלוונטיים בכל גוף, ועל מעקב שוטף אחר קבלת המידע או הסירוב להעניקו. כן נעשה חיפוש מקוון באינדקסים שונים ובאתרים רשמיים של הגופים הרלוונטיים באמצעות מילות מפתח שונות כגון מאגרי מידע, חינוך, מאגרים מקוונים, תכניות חינוכיות וכיו"ב. לאחר איסוף החומר והערכתו פורטו המאגרים בהתאם למתווה המוצע לעיל (ראה נספח א'), בסדר אלפביתי.

סקירה זו אינה מתיימרת להציע כיוון מחקרי כזה או אחר, אלא לפרוס בפני החוקרים המעיינים בקטלוג מאגרי המידע המוצע להלן את מגוון הקבצים העוסקים בחינוך, אלה הנאספים בארגונים ובגופים שצוינו לעיל, ובכך לסייע למיפוי המידע הקיים בישראל בתחום החינוך. לשם כך מופו בסקירה זו מאגרים העוסקים בגיל הרך, בגילי בית הספר, במוסדות ההשכלה הגבוהה, בתחומי החינוך הבלתי פורמלי, בישיבות, במסגרות הכשרה מקצועיות וכיו"ב – כל זאת לצד מיפוי של מחקרים בחינוך המכילים קובצי נתונים ומאגרים נוספים.

## פרק שלישי – פרטיות, התממה, וניהול סיכונים

בעשורים האחרונים הולך ומתעצם השיח הציבורי, האתי והאקדמי על אודות פרטיות, הצדקתה המוסרית ומשמעותה המעשית. סיבה אחת לכך היא מהפכת הידע והמידע המאפשרת להגיע למידע רב – גם כזה הנושא אופי אישי – להחזיק בו ולעבד אותו, וכל זה בקלות וביעילות (McFarland, 2012). בפרק זה יוצג נושא הפרטיות משלושה כיוונים שונים אך משלימים:

1. יוצגו מספר הצדקות תאורטיות לצורך בשמירה על פרטיות.
2. יתוארו מודלים סוציולוגיים המנסים להמחיש באילו תנאים עשויים להיווצר הן הצורך בפרטיות והן הנורמה של שמירה על פרטיות.
3. יובאו הנחיות חוקיות ורגולטוריות מישראל ומהעולם (הדין המצוי) הדנות בחובה לשמור על פרטיות בהקשר של מאגרי נתונים.
4. יוצגו גישות וכלים להפחתת הסיכון לפגיעה בפרטיות באמצעות הפחתת הסיכון לזיהוי מושאי המידע.

חשוב לומר שבדומה לפרקים האחרים בדוח הנוכחי, גם בפרק זה איננו מתיימרים להציג סקירה מקיפה של הספרות העוסקת בנושא, אלא תמצית מתוך הרצאות שהוצגו בפני צוות המומחים במהלך שנת פעילות.

### הצדקות תאורטיות לערכה של הפרטיות

בהרצאתו במפגש הלימודי ב-8 באוגוסט הציג פרופ' טל ז'רסקי שלושה סוגים של הצדקות תאורטיות בסיסיות לערך הפרטיות:<sup>20</sup>

1. ערך פנימי (אינטרינזי) לשליטה של האדם על המידע אודותיו: סוג אחד של הצדקות הן כאלה הרואות ערך אינטרינזי בכך שרק האדם עצמו הוא בעל השליטה על המידע הנוגע לו ועל מי שיקבל גישה למידע הזה. ביטוי אחד להצדקה מסוג זה הוא ההנחה שהמידע הנוגע ליחיד הוא רכוש, קניינו. כשם שהפרט הוא השולט על הגישה לרכושו, כן צריכה להיות לו שליטה על הגישה למידע הנוגע לו. ביטוי נוסף לתפיסה הרואה ערך אינטרינזי בפרטיות הוא ראיית השליטה על המידע כחלק מהאוטונומיה של הפרט או כתנאי לאוטונומיה זו. הידיעה שהמידע על אדם מסוים חשוף בפני אחרים עלולה להשפיע על מעשיו של נשוא המידע ועל יכולתו לקבוע בעצמו מה הוא רוצה לעשות. אנשים מתנהגים בצורה שונה כאשר הם חשופים לעין הציבורית. אור השמש הוא אמנם המחטא הטוב ביותר, אבל כשאנשים יודעים שמעשיהם חשופים לעין כול, הם עלולים לשנות את התנהגותם כדי להיות קונפורמיסטים יותר, כדי להתאים עצמם למקובל בחברה. ביטוי שלישי להצדקה מעין זו הוא תפיסת השליטה על המידע כחלק מהכבוד לאדם באשר הוא אדם (Human dignity) – חלק מהחובה להתייחס לאדם כ"מטרה כשלעצמו" ולעולם לא רק כאמצעי להשיג מטרות אחרות.<sup>21</sup>

2. הגבלת בעלי הררה: סוג אחר של הצדקות לצורך בפרטיות נוגע לרצון לרסן גופים גדולים ובעלי כוח על ידי הגבלת כוח המעקב שלהם אחר האזרחים והמידע שיש בידם על אודותיהם. אנחנו לא רוצים לחיות

<sup>20</sup> יש לציין שקיימת תאוריה נוספת – הפרטיות כגישה – שפיתחה פרופ' רות גביון (Gavison, 1980). תאוריה זו אינה מתאימה לדיון בסוגיית הפרטיות במאגרי מידע ולכן לא תוצג כאן.

<sup>21</sup> להרחבה על חובה זו ראה קאנט, 1984.

למשל במדינה שיודעת עלינו הכול, גם אם אנחנו אזרחים שומרי חוק וגם אם המדינה לא משתמשת לרעה במידע שברשותה. חוסר הנוחות הזה קיים גם כשמדובר בחברות פרטיות או ארגונים פרטיים.

המדינה המודרנית היא גוף עצום בכוחו, בעל פוטנציאל לפגוע ולרמוס פרטים. ככל שהמדינה תדע יותר על הפרטים כן יקל עליה להשתמש (והחשש הוא – להשתמש לרעה) בכוח שיש לה עליהם. ההצדקה הזו פחות רלוונטית כאשר מדובר בגופים פרטיים ובחובה שלהם לשמור על פרטיות (פחות רלוונטית אך עדיין משמעותית – ייתכן שיש להגביל גם את כוחם של גופים מסחריים גדולים). אנו רוצים לצמצם חששות אלו ונעשה כן על ידי צמצום יכולתה של המדינה להעביר מידע או להשיג פרטים על אנשים מידע מגופים גדולים אחרים.

3. החשש מפגיעה: סוג שלישי של הצדקות נוגע לחששות מפגיעות ממשיות וקונקרטיות שעלולות להיגרם לפרט אם ייחשף מידע עליו. שימוש במידע אישי עלול להוביל לפגיעות ביחיד. חשיפת מידע מעברו של הפרט עשויה לפגוע במערכות יחסים שלו או בקידומו בעבודה. צד שלישי יכול להשתמש במידע זה על מנת לסחוט את הפרט ולהפעיל עליו מניפולציות. גם כשהמידע נמצא בידי גופים האמונים על שמירתו הוא עלול לזלוג על רקע קָשֶׁל באבטחה ולהגיע למישהו שישתמש בו לרעה. הגבלה של העברת המידע למינימום האפשרי מקטינה את הסיכוי לזליגת מידע ולפגיעה בפרט באמצעות המידע עליו. השאלה היא כמובן מהו אותו מינימום ומיהו בעל הסמכות לקבוע זאת.

שלוש התאוריות האמורות מובילות לתוצאה דומה בהקשר שלפנינו: יש לשמור על פרטיותו של הפרט ולהגביל ככל הניתן את הגישה למידע בעניינו. הדבר כמובן אינו אומר שאין שימוש ללא הסכמה, שהרי גם בזכויות אחרות אנו פוגעים אם יש לנו סיבה טובה מספיק. למשל, לפי זכות הקניין, לפרט אמורה להיות שליטה על הגישה לרכושו. אלא שאם יש סיבה מוצדקת, אנחנו פוגעים בזכות זו (באמצעות מיסים או הפקעה). יחד עם זאת יש להקפיד על כך שפגיעה כזו תיעשה במידה מצומצמת ככל האפשר.

מסיבה זו כאשר המדינה מחליטה לאסוף מידע על היחיד ונשוא המידע אינו מנדב את המידע אלא מחויב לחושפו על פי חוק, נרצה לצמצם ככל האפשר את החשיפה למידע ואת השימוש בו. נאפשר את המשך השימוש במידע רק אם האדם שבו מדובר יסכים, או אם שוב יהיה אפשר להצדיק את הפגיעה על ידי צורך אמיתי. האיזון הנחוץ בין הפגיעה בפרטיות לבין צורכי המדינה והאזרחים מושפע משאלת ההצדקה שבה אנו בוחרים להשתמש בעניין הפרטיות, אולם קצרה היריעה מלעסוק כאן באופי ההשפעה הזו.

## תאוריות סוציולוגיות על פרטיות

ד"ר רן וולף פרש בהרצאתו שני מודלים סוציולוגיים שמטרתם להסביר מדוע ומתי נוצרות נורמות המחייבות שמירה על פרטיות. בניגוד להסברים הפילוסופיים-מוסריים, שמטרתם להצדיק את הצורך בשמירה על הפרטיות, כאן מדובר במודלים (מתמטיים ככל האפשר) שמטרתם לתאר (באופן הניתן לבחינה אמפירית) מתי מייחסים בחברה המודרנית ערך לשמירה על פרטיות ומדוע.

**המודל הראשון מבוסס על התאוריה של הייצוגים החברתיים:** תאוריה מתחום הפסיכולוגיה החברתית שמסבירה כיצד אנשים מבינים זה את זה (Moscovici, 1963). על פי תאוריה זו יש לכל אדם במוחו ייצוגים של אנשים אחרים שדרכם הוא מבין את האנשים, מסביר לעצמו את התנהגותם ומפרש את התקשורת עמם. אם אנחנו מכירים את האדם היטב אזי הייצוג הוא של אותו אדם, ואם אנחנו מכירים אותו באופן שטחי אנחנו משייכים אותו לקבוצת זְהוּת – ייצוג שמשותף ליותר מאדם אחד. הייצוגים הללו נרכשים בתהליך החיברות לאורך השנים.

מתוך התאוריה הזו נובעת ההבחנה כי בעולם המודרני (ובמיוחד במדינות ליברליות) לכל אדם יש זהויות רבות, דרכים רבות שבהן הוא מציג ומייצג את עצמו בפני אנשים שונים. לא תמיד הזהויות הללו תואמות זו את זו – אדם עשוי להתנהג בצורה אחת כלפי ילדיו, בצורה אחרת כלפי עמיתיו לעבודה, ובאופן אחר לחלוטין כלפי חבריו לפעילות הפוליטית במפלגה שבה הוא חבר. הזהות של אדם כעובד סותרת לעתים את זהותו כאב (אבא טוב מגיע הביתה מוקדם, שכיר טוב יוצא מהעבודה מאוחר). במקום שקיימת סתירה כזו – שם נוצרת נורמה של פרטיות העונה על הצורך בהפרדה בין הזהויות. איננו רוצים שילדינו יראו אותנו כשאנו עם חברינו, ואנחנו מתרעמים אם המנהל בעבודה חודר לחיינו הפרטיים. יש לציין כי הסבר סוציולוגי זה משמש גם כהצדקה מוסרית לפרטיות מהסוג הראשון שמנינו לעיל – הצורך בפרטיות הוא אינטרינזי משום שפרטיות מאפשרת לי ריבוי זהויות – ריבוי שהוא הכרחי כדי שנהיה מי שאנו באמת.

**מודל שני – תורת המשחקים:** מודל נוסף המסביר את הצורך בפרטיות מתבסס על המתודה ועל ההנחות התאורטיות של תורת המשחקים. לפי מודל זה הצורך בנורמה של שמירת פרטיות נוצר כאשר קיימים "משחקים" שבהם התוצאה האופטימלית – הטובה ביותר עבור כל אחד מהשחקנים – היא שמירה הדדית על הפרטיות, אבל תוצאה זו אינה תוצאת **שיווי המשקל (שיווי משקל נאש)** אם כל אחד פועל רק על מנת להשיא (to maximize) את האינטרסים שלו. פרטיות במקרה זה מוגדת כמצב שבו לשחקן מסוים יש שליטה על ההחלטה אם לחשוף מידע פרטי של השחקן האחר, אולם הוא בוחר שלא לעשות כן. במקרים רבים שיווי משקל של כיבוד פרטיות אינו יציב ולכן אינו בר-קיימא ללא מכניזמים של הגנת הפרטיות. המודלים הללו מאפשרים לדון בפרטיות במובן של אינטרסים (ולא אידאלים). זוהי הסתכלות מצמצמת על ההתנהגות האנושית. הסתכלות כזאת מניבה מסקנות מעשיות חזקות וברורות יותר (לפחות לפי הנחות מסוימות).

## כללים והנחיות רגולטוריות בנושא פרטיות: הדין בישראל

הכלל הבסיסי בחקיקה הישראלית בנוגע לשימוש במידע מינהלי בכלל ולהעברתו לגופים אחרים בפרט הוא שלא ניתן להשתמש במידע שלא למטרה שלשמה נמסר. בעניין זה יש בפרק ד' לחוק הגנת הפרטיות סעיפים ייחודיים שקובעים למי מבין הגופים הציבוריים מותר להעביר מידע. בנוסף לכך, ללשכה המרכזית לסטטיסטיקה מותר, על פי חוק, לקבל נתונים מזוהים מכל גוף ציבורי בישראל. לפי חוק הגנת הפרטיות מי שנפגעה פרטיותו רשאי להגיש תביעה ללא הוכחת נזק (בישראל ובאירופה) ולקבל עד 50,000 ש"ח. הפגיעה באדם אחד אינה מתבטלת גם אם יש בה תועלת לציבור, וגם אם שאר האנשים בקובץ הנתונים לא נפגעו כלל.

לפני שנציג את חוק הגנת הפרטיות ראוי להזכיר שכאשר מדובר בפעולות של המדינה, יש להביא בחשבון גם את הנחיות חוק יסוד: כבוד האדם וחירותו. חוק זה קובע (לפי פסיקת בג"ץ) שלא ניתן לפגוע בפרטיותו של אדם. אחת ההשלכות של ההכרה החוקתית היא שלא ניתן לפגוע בפרטיות אלא על סמך חוק קיים, וגם במקרה שנעשה שימוש כזה, הוא יעמוד למבחן המידתיות של בג"ץ. קיימים אמנם חוקים שקדמו לחוק היסוד (ולכן לא עמדו למבחן בג"ץ)<sup>22</sup> – אבל גם עליהם מוטלות מגבלות.

חוק הגנת הפרטיות פרק ד', סעיף 24ב' קובע כאמור שגוף ציבורי ימסור מידע רק למטרה שלשמה נמסר לאותו גוף, אלא אם חל אחד משני תנאים:

<sup>22</sup> חוקים שקדמו לחוקי היסוד אינם חייבים לעמוד במבחן בג"ץ (אף על פי שנפסק שגם הם כפופים במידה מסוימת לעקרונות המצויים בחוקי היסוד) אך שינויים בחוקים הללו מצריכים את החוק כולו לעמוד שוב במבחן הבג"ץ.

- פרסום כדין: העברת מידע בין רשויות ממשלה למטרת ביצוע תפקידן של הרשות המעבירה או הרשות המקבלת אינו מוגדר כפגיעה בפרטיות. שר המשפטים הוא הקובע (בתקנות) כיצד ניתן להעביר מידע כזה ומהם התנאים לכך.
- הסכמה של מוסר המידע להעברתו.

לא ניתן כלל להעביר מידע בין גוף ציבורי לפרטי ללא סמכות בדין או הסכמת מושא המידע. עניין זה נאכף בקפידה, דוגמה לכך: בעבר קיבלו הבנקים גישה למרשם האוכלוסין על מנת לוודא שהפרטים שנמסרו להם בידי האזרח נכונים. האגודה לזכויות האזרח עתרה לבג"ץ, ואף על פי שתכלית ההעברה הייתה הטוב הציבורי (מניעת הלבנת הון), ההעברה נפסלה מכיוון שלא נעשתה בהתאם לחקיקה. הבעיה נפתרה אמנם בעקבות חקיקה מיוחדת לעניין זה, אך גם היא הייתה צריכה לעמוד במבחן חוקתי מכיוון שהפרטיות היא זכות חוקתית.

על פי התקנות הנוכחיות, אפשר להעביר מידע בין גופים ציבוריים לצורך מילוי מטרותיהם או במסגרת סמכויותיהם, בכפוף להחלטות ועדה להעברת מידע שתמונה בכל גוף. בתי חולים וקופות חולים מצוינים שם כגופים ציבוריים למטרות מסוימות. גם מוסדות המוכרים כמוסד להשכלה גבוהה נמצאים ברשימה ויכולים לקבל סוג מסויים של מידע לצורכי מחקר. רשימת הגופים המוגדרים גופים ציבוריים לעניין זה נמצאת באתר המועצה להשכלה גבוהה בקישור כאן.

בתקנות (שנכתבו בשנת 1986 ובוצע בהן תיקון בשנת 2005) יש התייחסות להעברת המידע ולדרך ביצועה – בעיקר הקפדה על אבטחת מידע. פרק ד' קובע כללים מוחמרים לעניין "מידע מוגבל" אשר כולל גם מידע על מצב בריאותו של אדם. הנחיות של היועץ המשפטי, המחייבות בדיקה הן של סבירות ההעברה והן של אמצעי האבטחה בגוף המקבל, בקישור כאן.

לאחר הצגת התשתית הכללית בחוק, חשוב לציין כי הפרקטיקה, ואולי גם חוקים רלוונטיים נוספים, צריכים להישען על מסגרת נורמטיבית רחבה יותר. בפועל פרקטיקה זו נקבעת על ידי הרגולטור – מנהל רמו"ט (הרשות למשפט טכנולוגיה ומידע) במשרד המשפטים.

## תפיסת הרגולטור בישראל

עו"ד לימור שמרלינג, מנהלת מחלקת קשרי ציבור וממשל ברשות למשפט, טכנולוגיה ומידע (רמו"ט) במשרד המשפטים, הציגה בפני צוות המומחים את תפיסת הרגולטור הישראלי בנושא פרטיות. ככלל התפיסה היא שהזכות לפרטיות חשובה ובעלת משקל, אך יש לאזנה מול זכויות אחרות. תפיסה זו עולה בקנה אחד עם זו של רגולטורים ומומחים בעולם.

במסגרת תפקידו עוסק הרגולטור בשאלות רבות הנוגעות לאיזון בין הצורך לחשוף מידע שעשוי להביא תועלת לציבור הרחב, לבין הסיכון לפגיעה בפרטיות. העיקרון הבסיסי הוא שלא פוגעים בפרטיות של אדם ללא הסכמתו. כל העברת מידע מהווה סיכון לפגיעה בפרטיות, אלא אם אותו אדם הסכים להעברת המידע, או שהגוף המקבל מוסמך לקבל את המידע מכוח הדין.

קיים הבדל בין גופים ציבוריים והעברה של מידע ביניהם לבין גופים פרטיים והעברת מידע ביניהם. לגבי גופים ציבוריים הכלל הבסיסי מכונה "עקרון צמידות המטרה". לפי עיקרון זה מותר להעביר מידע בין גופים אם ורק אם המידע נאסף לצורך המטרה שלשמה ישתמש בו הגוף המקבל. כשגוף אחד אוסף מידע למטרותיו, אין זה אומר שהוא רשאי להעביר את המידע לגוף ממשלתי אחר, אף על פי ששניהם שייכים לממשלה

ומטרותיהם זהות לכאורה. כל משרד הוא יחידה בפני עצמה ולעתים אף בתוך משרד יש יחידות מובחנות שהעברת המידע ביניהן אינה אפשרית ללא אישור מוקדם. כך למשל במשרד המשפטים לא מאפשרים העברת מידע מהפרקליטות לטאבו, אלא אם ניתנה לכך תשובה מנומקת מדוע המהלך תקין.

בין השאלות שבהן עוסק הרגולטור ניתן למנות גם כאלה שהן טכניות במהותן אך חשובות מנקודת המבט של שמירת הפרטיות, למשל: לכמה זמן שומרים מידע (או מתירים לשמור מידע)? האם ניתן להעביר מידע לשמירה ועיבוד אצל ספק פרטי וכיצד מגינים על מידע שנמצא אצל ספק חוץ?

הניהול התקין מחייב לעבור את תהליך החשיבה שהודגם כאן – ניתוח הסיכונים ומענה להם – כך שכל פרויקט שנבחן ייצא לפועל תוך מזעור הפגיעה בפרטיות. תהליך ניתוח הסיכונים יוצג בהמשך. חשוב להבין שמדובר בתחום עיסוק נרחב, שבמסגרתו כל מקרה ומקרה מחייב שיקול דעת מקצועי.

לגבי מחקר אקדמי – תפיסת הרגולטור היא שנכון לעשות שימושים במידע אישי לצורכי מחקר, אם וכאשר אנו מעריכים שניתן להפיק מהמחקר תועלות חברתיות. קיים צו נפרד שמאפשר העברת מידע מסוגים מסוימים, המפורטים בצו, לגופים אקדמיים למטרות מחקר. עם זאת, תמיד צריך לבחון דרכים להערכת פוטנציאל הפגיעה בפרטיות ולערוך סוג של ניהול סיכונים שבמסגרתו מאזנים בין זכות הפרט לבין טובת הכלל.

## התממה ואנונימיזציה

דרך מרכזית שבה בוחרים גופים ממשלתיים להתמודד עם בעיית הפרטיות בהנגשת קובצי נתונים היא התממה שלהם – שינויים שימנעו הפקת מידע מתוכם על אודות אדם מסוים. בחלק זה נציג תחילה את הנחיותיהן של מספר רשויות רגולטוריות מרכזיות בעולם בנושא התממת נתונים ובשאלה אילו נתונים נחשבים "מותממים". שוב יש לציין כי אין מדובר בסקירה מקיפה אלא רק בסיכום ההנחיות הרגולטוריות שהוצגו במפגשים הלימודיים. לאחר מכן נציג סקירה (לא ממצה) של שיטות להתממת קובצי נתונים ושל כלים קיימים בשוק המאפשרים שימוש בשיטות אלו.

על אף שאינו עוסק ישירות בנתוני חינוך, נראה ש"סטנדרט הזהב" בנושא התממת נתונים הוא ההנחיות שפרסם ראש סוכנות הפרטיות במידע רפואי (Health Information Privacy) במשרד הבריאות האמריקאי. הנחיות אלו פורסמו בהתאם לחוק הנוגע (במקור) להעברת נתוני ביטוח בריאות – HIPPA – (Health Insurance Portability and Accountability Act of 1996), ולפיהן קיימות שתי דרכים להפוך מידע למותמם:

**שיטת קביעת מומחים (Expert determination):** בדרך זו מפעילים מניפולציות סטטיסטיות על המשתנים השונים בקובץ הנתונים, באופן שיביא לכך שהאנשים מאחורי הנתונים לא יהיו ניתנים לזיהוי. ההנחיה הנוגעת לדרך זו מחייבת:

- שהמלאכה תיעשה בידי אדם בעל ידע והכשרה בשיטות סטטיסטיות ובעקרונות המדעיים להפיכת מידע לבלתי ניתן לזיהוי.
- שיטות סטטיסטיות ועקרונות מדעיים אלו יופעלו, וייקבע שנמוך מאוד הסיכון שיעשה שימוש בנתונים – בין כשלעצמם ובין בשילוב נתונים הניתנים להשגה תוך שימוש באמצעים סבירים (Reasonably available information) – כדי לזהות את המידע מאחריהם.
- שהמסמכים והדרכים יוצגו בצורה שתצדיק קביעה כזו.

**שיטת נמל מבטחים (Safe Harbor):** בשיטה זו יש להסיר מתוך המידע שמונה עשר משתנים מזהים שיש בהם כדי לזהות את הפרט, את קרוביו, את מעסיקיו או את בני ביתו.<sup>23</sup> כן מחייבת ההנחיה לוודא שהיישות המחזיקה בנתונים אינה מודעת לאף שימוש אפשרי במידע, בין כשלעצמו ובין בשילוב עם מידע אחר, לזיהוי אדם פרטי שהוא נשוא המידע.

דוגמה נוספת היא ההנחיות שמפרסם הוועד הפדרלי למסחר (Federal Trade Commission) בארצות הברית בעניין ההגנה על פרטיות בהקשר המסחרי. לפי הנחיות אלו חברות אמורות "לנקוט את הצעדים הסבירים וההגיוניים על מנת לוודא שהמידע מותמם. מידע מותמם הוא כזה שלא ניתן להסיק ממנו מידע על צרכן ספציפי או על מכשיר או מחשב שברשותו. החברה אמורה גם להתחייב שמי שהמידע יגיע לידיו לא יזהה אותו מחדש, ואם היא מעבירה את המידע למשתמשים נוספים, עליה לבקש גם מהם להתחייב לכך.

לפי סעיף 502 (4) של חוק ההגנה על מידע חסוי והיעילות הסטטיסטית, שחוקק בארצות הברית בשנת 2002 (CIPSEA – Confidential Information Protection & Statistical Efficiency ACT) "המונח 'צורה ניתנת לזיהוי' פירושו ייצוג כלשהו של מידע המאפשר היסק סביר, ישיר או עקיף, של זהות הנבדק."

דוגמה מחוץ לארצות הברית:

- תקנה (EC) מס' 223/2009 של הפרלמנט ושל מועצת האיחוד האירופי. על פי תקנה זו, "אין להעביר נתונים המאפשרים את זיהוין של יחידות סטטיסטיות, אם ישירות ואם בעקיפין, ובכך חושפים מידע על הפרט. כדי לקבוע אם יחידה סטטיסטית ניתנת לזיהוי, יובאו בחשבון כל האמצעים הרלוונטיים שיש סבירות שגורם שלישי יעשה בהם שימוש כדי לזהות את היחידה הסטטיסטית."
- סעיף 39(3) של מחלקת הסטטיסטיקה והרישום באנגליה, תקנה מס' 2007, בעניין "מידע אישי": לעניין תת-סעיף מס' 2, מידע נחשב למזהה אדם מסוים אם זהותו של אדם זה –
  - א. מצוינת במידע
  - ב. אפשר להסיק אותה מהמידע
  - ג. אפשר להסיק אותה מהמידע כשמשלבים אותו עם מידע אחר, כלשהו, שפורסם
- סעיף 7(1) בהגדרת הסטטיסטיקה באוסטרליה (1983) מתייחס למידע חסוי כ"מידע הנוגע לאדם ואינו בעל אופי אישי או משפחתי ואשר יכול, בסבירות גבוהה, לאפשר זיהוי של אותו אדם."
- סעיף 17(1)(b) של חוק הסטטיסטיקה בקנדה קובע: "כל אדם שהושבע לפי סעיף 6 לא יחשוף או יגרום בודיעין לחשיפה, בכל אמצעי שהוא, של מידע כלשהו שהושג מכוח החוק הזה, באופן שיהיה ניתן, כתוצאה מהחשיפה הזאת, לקשר את הפרטים שהושגו מאדם כלשהו עם אדם, עסק או ארגון שאפשר לזהותם."

<sup>23</sup> שמות, כל תת-חלוקה גאוגרפית שהיא קטנה ממדינה (state), בכלל זה כתובת דואר, עיר, מחוז, רובע, מיקוד, וה-geocodes האקוויולנטים להם, כל רכיבי התאריכים (למעט השנה) עבור תאריכים שקשורים ישירות לאדם הספציפי, בכלל זה תאריך לידה, שנת קבלה (למוסד הרפואי, אבל בהקשר שלנו מתאים גם למוסד הלימודי), תאריך שחרור (כ"ל), וכן כל הגילים שמעל 89 וכל רכיבי התאריכים (בכלל זה שנה) היכולים להצביע על גיל כזה, למעט גילים ורכיבים כאלה שעלולים להצטרף לקטגוריה יחידה של גיל 90 ומעלה, מספרי טלפון, אמצעי זיהוי של רכב ומספרים סדרתיים, בכלל זה מספרים של לוחיות רישוי ורשיונות נהיגה, מספרי פקס, אמצעי זיהוי של התקנים ומספרי סדרות, כתובות דואר אלקטרוני, מען מְשָׁאָבִים אָחִיד במרשתת (URL), מספרי ביטוח לאומי (Social Security), כתובות של פרטוקולים אינטרנטיים (IP), מספרי מסמכים רפואיים, אמצעי זיהוי ביומטריים, בכלל זה טביעות אצבע וקול, מספרים של מוטבֵי ביטוחים רפואיים, צילום פנים מלא וכל תמונה דומה, מספרי חשבונות, כל רכיב נוסף שהוא מזהה ייחודי – מספר, מאפיין או קוד, למעט מה שמותר על פי פיסקה ג' דלעיל, וכן מספרי רשיונות או תעודות.

## שיטות להגנה על הפרטיות ולהתממת קובצי מידע<sup>24</sup>

בחלק זה נציג כאמור סקירה של שיטות להתממת קובצי נתונים ושל כלים קיימים בשוק המאפשרים שימוש בשיטות אלו. הבחנה מקדימה החשובה להבנת חלק זה היא ההבחנה בין משתנים מזהים – כאלה שניתן באמצעותם בלבד לזהות אדם ספציפי מתוך מאגרי הנתונים (למשל מספר תעודת זהות) – לבין משתנים "חצי מזהים" (Quasi identifiers) כאלה שבהם עצמם אין די לזהות אדם מסוים, אבל בעזרת צירוף של מספר משתנים כמותם או של נתונים ממאגר מידע חיצוני – זיהוי כזה אפשרי.

### גישות להערכת רמת האנונימיות

**K Anonymity:** מושג שחשוב מאוד להכיר הוא גישת **K Anonymity** להתממת נתונים. גישה זו מחשבת את הסיכון לזיהוי אדם אחד מתוך קובץ נתונים על ידי בחינת הקובץ ומציאת ערך ה- $K$  שלו (שיכול להיות 1 או יותר). ערך ה- $K$  הוא המספר הקטן ביותר של הרשומות הזוהות בכל הפרמטרים או בחלקן בתוך מאגר הנתונים. ככל שערך ה- $K$  גבוה יותר הסיכון לזיהוי של פרט בתוך מאגר הנתונים הוא קטן יותר. כדי להגדיל את ערך ה- $K$  ניתן לערוך מניפולציות שונות על הנתונים על מנת לקבצם לקבוצות הומוגניות – קבוצות בעלות רשומות זהות בפרמטרים המזהים. אם קיימת רשומה ייחודית – רשומה שניתן בוודאות לדעת שהיא שייכת לאדם מסוים – אזי ערך ה- $K$  יהיה 1. הקביעה איזה ערך  $K$  מספיק גבוה כדי להבטיח רמת סיכון סבירה לפגיעה בפרטיות – נתונה לשיקול דעתו של בעל הנתונים רוב הכלים להתממת נתונים משמשים להגדלת ערך ה- $K$  של קובץ הנתונים.

לגישת **K Anonymity** יש שתי מגבלות לפחות: ראשית, לפעמים די בכך שידוע שאדם מסוים הוא אחד מבין  $X$  מופעים הומוגניים אפשריים כדי לגלות מידע רגיש על אותו אדם (מגבלת ההומוגניות). ניקח לדוגמה קובץ רשומות רפואיות. אם אחת הרשומות היא של המחלה שממנה סובל החולה, ואם אותה מחלה משותפת ל- $X$  רשומות רפואיות, די בכך שאנו יודעים שהאדם שעליו אנחנו מחפשים מידע הוא אחד מבין  $X$  הרשומות הרפואיות הללו כדי שנדע מהי המחלה שממנה הוא סובל.

מגבלה שנייה של גישה זו היא שהיא אינה מונעת מאיתנו לזהות אדם מסוים אם יש לנו ידע מוקדם עליו, גם אם איננו מצליחים לשייך אותו לרשומה ספציפית אחת. למשל, אם אנו יודעים ששכן שלנו מאושפז, ואנחנו מצליחים לצמצם את אפשרויות המחלה שלו לשתיים – שפעת או רגל שבורה – אזי אם שמענו אותו משתעל ומקנח את האף אתמול בחדר המדרגות, נוכל להסיק שהוא חולה בשפעת. (Machanavajjhala, Kifer, Gehrke, & Venkatasubramanian, 2007).

**I Diversity:** גישה זו תוכננה על מנת להתמודד עם מגבלותיה של גישת ה- $K$  anonymity. בלי להיכנס לפרטים הטכניים ולחישוב המתמטי שעליו מבוססת הגישה, ניתן לומר שעל פי גישה זו אין די ביצירת קבוצות הומוגניות של מופעים בתוך קובץ הנתונים, אלא יש לייצר קבוצות שהן הומוגניות מבחינת המשתנים המזהים אך הטרוגניות ככל האפשר מבחינת המשתנים הרגשיים – משתנים שתוקפים ירצו מן הסתם לחשוף. בדוגמה של הרשומה הרפואית ניתן לומר שיש ליצור קבוצות שבהן משתנים כגון השם, אזור המיקוד, או גיל החולה זהים, אולם משתנה המחלה שממנה הוא סובל הוא הטרוגני ככל הניתן (Machanavajjhala, Kifer, Gehrke, & Venkatasubramanian, 2007).<sup>25</sup>

<sup>24</sup> המידע שיוצג בעמודים הבאים נושא אופי מתודולוגי ועשוי שלא להיות רלוונטי לכל הקוראים. קוראים שמידע מסוג זה פחות רלוונטי עבורם מוזמנים לעבור לעמוד 35 להמשך הקריאה.

<sup>25</sup> קיימים פירושים שונים לגישת ה- $I$  diversity, כאלה הדורשים שהערכים המגוונים "יחולקו" בצורה שווה בין הקבוצות, או כאלה הדורשים שהערכים המופיעים הכי הרבה בקובץ לא יופיעו הכי הרבה בקובץ המותמם (Li, Li & Venkatasubramanian, 2002), אולם לא ניכנס כאן להבדלים בין הגישות.



לגישת I Diversity קיימות מגבלות משלה. דוגמה למגבלה אחת כזו היא הסכנה שאף על פי שהערך הרגיש הוא הטרוגני בתוך כל אחת מהקבוצות ההומוגניות, עדיין קיים דמיון בין הערכים כך שהתוקף עדיין יכול להסיק מידע רגיש על האדם הספציפי (בהנחה שהוא יודע לאילו מן הקבוצות ההומוגניות משתייך אותו אדם). כך למשל, אם בתוך קבוצה מסוימת יש המשתנה "שכר" והערכים הם 4,000, 5,000, 6,000 ו-4,000 ש"ח לחודש, ואם אנחנו יודעים שאדם מסוים שייך לקבוצה זו, אזי נוכל להסיק שהוא מרוויח פחות מההכנסה הממוצעת החודשית בישראל (נכון לשנת 2016), גם אם לא נדע כמה בדיוק הוא מרוויח.

מגבלה אחרת היא שהשגת שונות גבוהה בערך הרגיש בתוך כל אחת מהקבוצות יכולה להיות מאתגרת מאוד. למשל אם הערך של המשתנה הרגיש הוא בינארי – חיובי או שלילי ( כמו במקרה של בדיקת נגיף ה-HIV) ואם ההתפלגות של הערכים היא שונה מאוד (למשל – רוב גדול של האוכלוסייה הוא "שלילי" ורק מיעוט קטן "חיובי"), יהיה קשה ליצור קבוצות שהן גם הומוגניות בערכים המזהים וגם הטרוגניות בערך הרגיש (Li, Li & Venkatasubramanian, 2007).

T closeness: על מנת להתמודד עם הבעיות של גישת ה-I diversity הוצגה (עוד באותה שנה) גישת ה-T Closeness. גישה זו מגדירה ראשית קבוצה הומוגנית כבעלת T Closeness קבוצה הומוגנית שבתוכה חלוקת הערכים במשתנה המוגן (המשתנה הרגיש) אינה שונה מחלוקת הערכים של אותו משתנה בקובץ הנתונים כולו ביותר מ-T. קובץ נתונים מוגדר כבעל T Closeness אם כל הקבוצות ההומוגניות בתוכו הן בעלות T.T Closeness, בדומה ל-K, צריך להיקבע על ידי בעל הנתונים, והוא מייצג את רמת הביטחון או רמת ההתממה של קובץ הנתונים.

### שיטות סטטיסטיות להתממת נתונים

ניתן לציין שתי גישות לנושא מניעת הזליגה של מידע פרטי מתוך קובץ נתונים: 1. הסוואה – שינוי או הסתרה של משתנים מזהים ושל משתנים "חצי מזהים" מתוך קובץ הנתונים. 2. יצירת קובץ נתונים סינתטי שלא ניתן לזהות מתוכו אנשים אמיתיים.

החיסרון של כל השיטות הללו הוא שהן פוגעות, במידה כזו או אחרת, בתועלת שניתן להפיק מהמידע באמצעות מחקרים.

### 1. הסוואה של משתנים בשיטות שונות:

- **הסרה:** מחיקה גמורה של המשתנה מקובץ הנתונים. שיטה זו מבטיחה שלא ישתמשו במשתנה כדי לזהות מישהו מתוך הקובץ, אולם מפחיתה את התועלת שניתן להפיק מקובץ הנתונים, בעיקר לצורכי מחקר.
- **הכללה או הקבצה:** ניתן לקבץ ערכים לתוך טווחים וכך להפחית את האפשרות לזיהוי פרטים באמצעותם. למשל, במקום לכתוב את הגיל המדויק אפשר לכתוב 1-10, 10-20, 20-30 וכן הלאה. בשיטה זו ניתן עדיין לערוך גרסיות סטטיסטיות על בסיס המשתנה המוסווה, אולם הרגרסיות יהיו מדויקות הרבה פחות.
- **מניפולציה של המשתנים:** ניתן לערוך על המשתנים מניפולציה מתמטית מסוג כלשהו, למשל להוסיף X לכל ערך או להכפיל כל ערך ב-X. אם השינוי נעשה בזיהרות מספקת ניתן עדיין לשמר תכונות סטטיסטיות של המשתנים, אולם סכנת הזיהוי, אם מישהו יכול להתחקות אחר המניפולציה המתמטית, היא גדולה יחסית.
- **החלפה:** לעתים מחליפים ערכים של משתנים חצי-מזהים בין רשומות. החלפה עשויה לשמר חלק מהתכונות הסטטיסטיות של הקובץ (אם היא נעשית באופן זהיר), אולם אינה מאפשרת עריכת רגרסיות סטטיסטיות כלל.

■ **דגימה חלקית:** שחרור חלק מקובץ הנתונים יכול לשמר את התכונות הסטטיסטיות שלו (תוך הגדלת השונות כמובן) ולהפחית, אם כי לא לבטל, את הסיכון לזיהוי פרטים מתוכו.

## 2. יצירת קובץ נתונים סינתטי:

אפשרות נוספת להתממת נתונים דווחה לצוות המומחים בידי מר זיו אופק.<sup>26</sup> בשיטת התממה זו מייצרים "אוכלוסייה" – קובץ נתונים – בעלת מאפיינים סטטיסטיים זהים: קוהורט של אנשים שמספרים את אותו סיפור סטטיסטי ממש. המידע החדש אמור לכלול בדיוק את אותם מאפיינים סטטיסטיים שכוללת אוכלוסיית המקור. המערכת אינה צריכה לדעת מראש מהי שאלת המחקר, אך היא מחייבת ידע מוקדם של המודל הסטטיסטי שלפיו רוצים לעבוד. היתרון בשיטה זו הוא שהיא חסינה מאוד מפני זיהוי של אדם אחד או יותר מתוך קובץ הנתונים, פשוט מפני שהאוכלוסייה ש"מאחורי" הנתונים אינה אוכלוסייה אמיתית. בצורה כזו ניתן ליצור חדר מחקר בארגון שבו אין צורך לייצר חבילות נתונים חדשות כל פעם שחוקר רוצה לקבל מידע, אלא אפשר לקבל מיידית כל מידע שהחוקרים מחפשים. שיטה זו מאפשרת גם לחבר מאגרי נתונים שונים אם יש לנו גישה לנתונים המקוריים (מהביטוח הלאומי ומהלמ"ס למשל).

על פי סקר מומחים שערך משרד הבריאות (במסגרת תהליך בירור של המשרד לקראת מערכת הנגשת נתונים המתוכננת אצלו), חסרונה של מערכת כזאת הוא שחוקרים רבים מעוניינים להתבונן על קובץ הנתונים האמיתי לפני הבחינה של מאפייניו הסטטיסטיים.<sup>27</sup> הסיבה לכך היא שרק מהתבוננות בקובץ עצמו ניתן לזהות בעיות אפשריות שקיימות בו, ולטייב את הנתונים. הדוגמה שהוצגה בפנינו היא של קובץ נתונים שבו העמודה "גובה" הציגה ערך לא הגיוני בחלק מהמקרים. הסיבה לכך הייתה שחלק ממושאי הקובץ היו נכים בכיסא גלגלים ולא היה ניתן למדוד בקלות את גובהם. אנומליה זו לא הייתה מופיעה בקובץ נתונים סינתטי, והתוצאות של כל מחקר המסתמך עליו היו מוטות.

## כלים להתממת נתונים

בשוק המוצרים קיימים מספר מוצרים המאפשרים התממה של נתונים על ידי הסוואה של משתנים. במאמר סקירה משנת 2012 ציינו וינוגראדוב ופסטסיאק (Vinogradov & Pastyak, 2012) מספר כלים המאפשרים זאת. גם כאן חשוב לציין כי האמור להלן הוא **לקט דוגמאות בלבד**, לקט שמטרתו לעדכן את הדיון באפשרויות העומדות לרשות הגורמים הציבוריים. בשום מקרה אין הכתוב מהווה סקירה של המוצרים הקיימים בשוק וקל וחומר שאינו מהווה המלצה על המוצרים המוצגים בו:

■ **SAP Test Data Migration Server:**<sup>28</sup> כלי להעברה ולשליפה מהירה של מידע (עסקי בעיקר) ממערכות עסקיות אחרות של חברת SAP. אף שהכלי מיועד בעיקרו להעברה וניהול של מידע, יש לו מספר פונקציות היכולות לשמש לצורכי אבטחת מידע והתממה. הכלי מאפשר לבעל המידע לשלוט בגישה למידע באמצעות מערכת של ניהול הרשאות מוגנות בסיסמה. הוא מאפשר גם הצפנה של הנתונים על פי מפתח שהחברה מספקת או על פי מפתח שבעל המידע קובע. ולבסוף הוא מאפשר למחוק, להסתיר או להחליף שדות מסוימים בסימנים, על פי כללים שקובע בעל המידע. האפשרות להחליף שדות כוללת גם את האפשרות להחליפם בערך הנראה כערך אמיתי המתאים לאותו שדה.

<sup>26</sup> מר אופק משווק מוצר הנוקט גישה זו לצורך השמירה על הפרטיות, ובמסגרת זו הופיע בפני הצוות.  
<sup>27</sup> הניסוח המדויק שקיבלו אנשי משרד הבריאות היה כי החוקרים רוצים "לחוש" את קובץ הנתונים. כשביקשנו הבהרות לגבי ניסוח מעורפל זה, קיבלנו את הדוגמה שתוצג מיד.

<sup>28</sup> SAP היא חברה מרכזית (אולי הגדולה בעולם) בתחום תכנות לריכוז, החזקה, שליפה ועיבוד של מידע עסקי.

- **Accenture Clone and Test HCM**: כלי המיועד לייצר "סביבת מבחן" לנתוני SAP. סביבת מבחן היא מאגר נתונים מקביל למאגר הנתונים המקורי אך נפרד ממנו. הכלי מיועד לביצוע מניפולציות ללא השפעה על המאגר המקורי (למשל לצורכי סימולציה). כמו הכלי הקודם גם כלי זה מאפשר הצפנת מידע אוטומטית ויעילה יחסית, המבוססת על כללים שקובע המשתמש. כך מאפשר הכלי הסתרה של שדות מסוימים, מחיקה שלהם או שינוי הערכים המופיעים בהם. הכלי מוגבל לקובצי נתונים של מערכות SAP בלבד ולכן, כרגע, לנתונים עסקיים בלבד.
- **GASPARIN Software Solutions, hr.dat.copy**: מוצר המיועד להעברת מידע אישי, בעיקר קובצי כוח אדם, בין מערכות מידע שונות. המוצר מאפשר לייצר סביבת מבחן תוך הצפנה של הנתונים, וכן לשנות פרטים בתוך הקבצים ולמחוק או לשנות שדות בעייתיים. הכלי מותאם לסביבת עבודה של חברת IBM בלבד ואינו מתאים לכל סביבת עבודה.
- **Data Sync Manager for HCM from EPI-USE**: כלי המיועד ליצירת מאגר נתונים לצורכי מבחן (Test data) ומאפשר ליצור סנכרון בין מאגר נתונים לצורכי מחקר לבין מאגר הנתונים המקורי. הכלי מאפשר הצפנה של שדות בתוך קובץ הנתונים – הן במאגר הנתונים המקורי והן במאגר המבחן – על פי מפתח שנקבע בידי המשתמש. ההעתקה, הסנכרון וההצפנה נעשים במהירות יחסית לשאר הכלים שהוצגו לעיל.
- **Oracle Enterprise Manager Data Masking Pack**: מוצר זה מיועד באופן ספציפי לאבטחת מידע בקובצי נתונים. הוא מאפשר מספר אפשרויות מעבר לאלו שמציעים המוצרים האחרים שתוארו לעיל. בין היתר הוא מאפשר זיהוי אוטומטי וסימון של שדות שעשויים להיות בעייתיים (מובן שיש עדיין צורך בעין אנושית על מנת לוודא את הזיהוי). כן הוא מאפשר קיבוץ של ערכי נתונים לתוך טווחים (גיל 0-12, 13-20, וכן הלאה), והפעלה של מניפולציה מתמטית על השדות (כולם, או על פי הנחית המשתמש). כל אלו נוספים כמובן על האפשרויות להצפין את הנתונים או את חלקם ולמחוק, לשנות או להחליף שדות מסוימים בערכים דומים.

### כלים חוזיים וטכנולוגיים לשמירה על פרטיות

חשוב לתת את הדעת על כך שמלבד התממת הנתונים, קיימים גם כלים חוזיים, משפטיים וטכנולוגיים לשמירה על פרטיותם של מושאי הנתונים. הכלי החוזי הוא בפשטות יצירת חוזה בין הצדדים לפיו יקבלו גישה לנתונים רק אלה שיתחייבו לא לחשוף את האנשים שמאחורי הנתונים. ברור שכלי זה משמעותי יותר ככל שניתן לאכוף את ההתחייבות הזו בקלות רבה יותר. לדוגמה: הלמ"ס מאפשר העברת המידע בתוך חדר המחקר לחוקר מוכר בלבד, המתחייב שלא לבצע פעולות מסוימות כגון חשיפת מושאי המידע או העברת המידע לצד שלישי. הסכמים דומים עשויים להיכרת גם בהקשרים אחרים – ואנו צריכים להעריך איזה משקל יינתן להם תוך התחשבות בעיקר בסנקציות שניתן להפעיל בגין הפרה.

כלים טכנולוגיים – הכוונה לשימוש בכלים המאפשרים לחוקר לגשת אל הנתונים אך מקשים עליו לעשות בהם שימוש אסור. דוגמאות לכך הן הגבלות גישה כך שהחוקר יכול להשתמש במידע אך לא להעתיקו למחשבו האישי, או מתן אפשרות לחוקר לייצר שאילתות מסוימות בלבד. השימוש בכלים אלו אינו מבטיח הגנה מלאה, אך הוא מקשה מאוד על פגיעה בפרטיות והופך ניסיון לפגוע בה ליקר יותר וקשה יותר. הממשק בין הכלים האלה לבין הכלים החוזיים והכלים הסטטיסטיים הוא המאפשר שמירה מלאה יחסית על הפרטיות.

## רשימת מקורות

- יוסטמן, מ' ובקובזה ג' (עורכים) (2010), *קווים מנחים לרענון מערך האינדוקטורים לחינוך בישראל: סיכום, מסקנות והמלצות*, היזמה למחקר יישומי בחינוך, האקדמיה הלאומית הישראלית למדעים, ירושלים.
- רן, ע. (2016) מיפוי מאגרי מידע בישראל בתחום החינוך. ל' יוספסברג בן-יהושע (עורכת), סקירה מוזמנת כחומר רקע לעבודת צוות המומחים בנושא "שימוש בנתונים הנמדדים לאורך זמן כמקור מידע למדיניות ולתכניות חינוכיות בישראל", היזמה למחקר יישומי בחינוך.
- שביט, י. ופניגר י. (2007), *מחקרי אורך במערכת החינוך הישראלית*, נייר עבודה שהוגש למדען הראשי במשרד החינוך.
- קאנט, ע. (1984) *הנחת יסוד למטפיזיקה של המדות*, תרגום: מ. שפי, הוצאת מאגנס, תשמ"ד 1984.
- Chetty, R., Hendren, N., & Katz, L. F. (2015). The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment.
- Dynarski, S. (2014), Building Better Longitudinal Surveys (on the Cheap) Through Links to Administrative Data, National Academy of Education.
- El Emam, K., Jabbouri, S., Sams, S., Drouet, Y., & Power, M. (2006). Evaluating common de-identification heuristics for personal health information. *Journal of Medical Internet Research*, 8(4), e28.
- 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/HiTrustDeIdentificationPresentation.pdf](https://hitrustalliance.net/documents/de_id/HiTrustDeIdentificationPresentation.pdf)
- Houng, B., & Justman, M. (2014). NAPLAN scores as predictors of access to higher education in Victoria.
- 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., & 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.

## נספחים

### נספח א': סדר יום של המפגשים הלימודיים

סדר היום של המפגש הלימודי הראשון, יום שני, כ"ט אייר, 6.6.2016<sup>29</sup>

התכנסות	10:00–9:30
<b>דברי ברכה ופתיחה – פרופ' מנחם יערי, יו"ר ועדת ההיגוי של היזמה</b>	10:10–10:00
מבוא: היתרונות של נתוני אורך לצורכי מחקר ומדיניות <b>פרופ' לאה אחדות, יו"ר צוות המומחים</b>	10:30–10:10
<b>מושב ראשון: נתוני חינוך בלמ"ס, סקר האורך של הלמ"ס, ואפשרויות לעתיד יו"ר: מר נעם זוסמן, חבר צוות המומחים</b>	13:30–10:30
סקר משקי בית אורך טווח בראייה ובתכנון העתידיים של הלמ"ס, והפוטנציאל הגלום בו ליצירת נתוני אורך בתחום החינוך <b>גב' נורית דוברין, גב' נרדית שטיין-קאפח וגב' אוריה לוי, תחום הסקר אורך הטווח, למ"ס</b>	11:00–10:30
תכניות עתידיות ויחזמות עצמאיות של הלמ"ס ליצירת נתוני אורך בתחום החינוך על בסיס מידע מינהלי <b>מר יוסי גידניאן, מנהל אגף א' (חינוך, השכלה וכוחות הוראה), למ"ס</b>	11:30–11:00
שימוש בנתוני אורך ליידוע מדיניות חינוך בישראל: מחקרים לדוגמה, אפשרויות ומגבלות <b>מר דוד מעגן, ראש תחום (חינוך גבוה וכוח אדם בהוראה) למ"ס</b>	12:00–11:30
האסטרטגיה הכללית של הלמ"ס ליצירת תשתיות מידע בחינוך במסגרת הסטטיסטיקה הלאומית <b>מר יואל פינקל, המשנה לסטטיסטיקן הממשלתי</b>	12:30–12:00
<b>הערות ושאלות של משתתפי המפגש</b>	13:00–12:30
הפסקה (תוגש ארוחת צהרים קלה)	13:30–13:00
<b>מושב שני: נתונים מינהליים שאוספים משרד החינוך וראמ"ה: יו"ר: פרופ' מיכל בלר, חברת צוות המומחים</b>	17:00–13:30
מה נעשה ומה דרוש לצורך עיצוב מדיניות ולצורך מחקר בחינוך (איסוף, שיתוף, והנגשה)? <b>ד"ר איתי אשר, מ"מ המדען הראשי, משרד החינוך</b>	14:00–13:30

<sup>29</sup> קישורים למצגות מהמפגש הלימודי הראשון, 6.6.2016

14:30–14:00	איסוף והנגשה של נתוני אורך בתחום החינוך בישראל ושיתופם בין משרדי הממשלה <b>מר פנחס קליין, מנהל כלכלה ותקציבים, משרד החינוך</b>
15:00–14:30	אפשרויות לשיתופם ולהנגשתם של נתוני משרד החינוך לצורכי מחקר ועיצוב מדיניות <b>ד"ר חיים גת, מנהל מינהל תקשוב, טכנולוגיה ומערכות מידע, משרד החינוך</b>
15:30–15:00	נתונים שראמ"ה אוספת והאפשרויות לשיתופם ולהנגשתם לצורכי מחקר וניתוח מדיניות <b>ד"ר חגית גליקמן, מנכ"לית ראמ"ה</b>
16:00–15:30	הנתונים הדרושים לצורך תכנון אסטרטגי במשרד החינוך <b>ד"ר מיכל טביביאן, אגף בכיר לתכנון ואסטרטגיה, משרד החינוך</b>
16:30–16:00	הערות ושאלות של משתתפי המפגש
17:00–16:40	דברי סיכום: <b>פרופ' לאה אחדות, יו"ר צוות המומחים</b>

### סדר היום של ישיבת הצוות בהשתתפות מומחים חיצוניים יום ראשון, ד' תמוז, 10.7.16<sup>30</sup>

יו"ר: פרופ' לאה אחדות, יו"ר צוות המומחים

11:30–11:00	התכנסות והצגת המשתתפים
12:45–11:30	הטכנולוגיה לשילוב בין מערכות מידע בעולם הרפואה, והאפשרויות להתאימה לעולם החינוך <b>מר זיו אופק, מייסד ומנכ"ל CDI-נגב</b> דיון
14:00–12:45	על סקרי אורך בחינוך בארה"ב ומה ניתן ללמוד מהם – תכנית ומתודולוגית (באנגלית) <b>פרופ' אדם גמראן, נשיא קרן William T. Grant Foundation.</b> דיון
14:30–14:00	הפסקה והמשך שיחה עם אדם גמראן, ארוחת צהריים קלה בקפטריה
15:45–14:30	הצגת מערכת האיזופד של הלמ"ס ושימושיה <b>גב' עליזה קרובי ומר עדנאן מנצור</b> דיון

<sup>30</sup> קישורים למצגות מיישיבת הצוות, 10.7.2016

**סדר היום של המפגש הלימודי השני, יום שני, ד' אב, 8.8.2016<sup>31</sup>**

התכנסות	9:50–9:30
פתיחה: פרופ' לאה אחדות, יו"ר צוות המומחים	10:00–9:50
מבוא: המצב החוקי ודילמות אתיות בנושא פרטיות ושימוש בנתונים מינהליים פרופ' טל ז'רסקי, חבר צוות המומחים	10:30–10:00
מושב ראשון: הנגשת נתוני אורך מזהים בראי המשפט והאתיקה יו"ר: פרופ' טל ז'רסקי חבר צוות המומחים	13:30–10:30
שיתוף נתונים במגבלות הצורך החוקי והאתי בשמירה על הפרטיות גב' לימור שמרלינג, ראש אגף קשרי ממשל במינהל רמו"ט, משרד המשפטים	11:30–10:30
האפשרויות והמגבלות המשפטיות בהנגשת נתוני הלמ"ס לחוקרים עו"ד בריאן ניגן, היועץ המשפטי של הלמ"ס	12:30–11:30
נתוני משרד החינוך ושיתופם עם חוקרים וקובעי מדיניות, מבט משפטי ופרקטי עו"ד דורית מורג, היועצת המשפטית של משרד החינוך	13:30–12:30
הפסקה (תוגש ארוחת צהרים קלה)	14:00–13:30
מושב שני: נתונים מינהליים שאוספים משרד החינוך וראמ"ה יו"ר: ד"ר איריס טבק, חברת צוות המומחים	17:00–14:00
שילוב ושיתוף של נתונים ממקורות שונים – תכנית 'התחלה טובה' כמקרה בוחן גב' טלל דולב, משרד הרווחה	15:00–14:00
טכניקות וטכנולוגיות המאפשרות שיתוף נתונים מינהליים במגבלת השמירה על הפרטיות ד"ר רן וולף, אוניברסיטת חיפה	16:00–15:00
הפסקה	16:15–16:00
סיכום פרופ' לאה אחדות, יו"ר צוות המומחים	16:45–16:15

<sup>31</sup> קישורים למצגות מהמפגש הלימודי השני, 8.8.2016



**סדר יום העיון, יום שני י"ט בכסלו, 19.12.2016<sup>32</sup>**

התכנסות והרשמה	09:00–08:30
<b>דברי פתיחה: פרופ' לאה אחדות</b> , המרכז האקדמי רופין, יו"ר צוות המומחים	9:15–9:00
<b>מושב ראשון: הפוטנציאל והחשיבות של מחקרי אורך בתחום החינוך יו"ר: פרופ' לאה אחדות</b> , המרכז האקדמי רופין, יו"ר צוות המומחים	11:15–09:15
הזדמנויות ואתגרים באיסוף וניתוח של נתוני אורך בחינוך (באנגלית, יתאפשר תרגום סימולטני) <b>פרופ' דיוויד קפלן</b> אוניברסיטת מדיסון-וויסקונסין	9:55–9:15
הפוטנציאל של מחקרי אורך בתחום החינוך בישראל <b>פרופ' משה יוסטמן</b> אוניברסיטת בן-גוריון בנגב	10:35–9:55
השפעות בטווח הארוך של התערבויות חינוכיות בבתי ספר יסודיים ותיכונים בישראל <b>פרופ' ויקטור לביא</b> , האוניברסיטה העברית בירושלים	11:15–10:35
הפסקה – קפה ועוגה	11:40–11:15
<b>מושב שני: שימוש בנתונים מינהליים למטרות מחקר יו"ר: ד"ר איריס טבק</b> , אוניברסיטת בן גוריון בנגב, חברה בצוות המומחים	13:20–11:40
נתונים מינהליים בתחום החינוך בישראל: תהליך הסקירה ותוצאותיה <b>ד"ר עמליה רן</b> , מכון מופ"ת	12:10–11:40
פעולות משרד הבריאות כדוגמה להנגשת נתונים מינהליים לחוקרים בישראל <b>גב' אלין אטיאס</b> , משרד הבריאות	12:40–12:10
סיכום: הפער בין המצוי לרצוי ולאפשרי <b>מר נועם זוסמן</b> , בנק ישראל, חבר בצוות המומחים	12:50–12:40
תגובות ושאלות מהקהל: <b>מר יואל פינקל</b> , הלשכה המרכזית לסטטיסטיקה	13:20–12:50
ארוחת צהריים	14:00–13:20
<b>מושב שלישי: איזון בין הגנת הפרטיות לבין הטוב הציבורי שאפשר להפיק ממחקרי אורך יו"ר: פרופ' טל ז'רסקי</b> , אוניברסיטת חיפה, חבר בצוות המומחים	15:45–14:00
טובת הכלל מול זכויות הפרט: התבוננות אתית-מוסרית <b>ד"ר יובל אילון</b> , האוניברסיטה הפתוחה	14:30–14:00

<sup>32</sup> קישורים למצגות מיום העיון, 19.12.2016

<p>התועלת במחקרי חינוך לעומת הסיכון לפגיעה בפרטיות – בראי מושג המידתיות  <b>פרופ' רענן סלוציאנו קינן</b>, המכון הישראלי לדמוקרטיה</p>	15:00–14:30
<p>בין סיכוי לסיכון בעבודת הרגולטור בישראל, שלושה מקרי מבחן  <b>עו"ד לימור שמרלינג</b>, מנהלת מחלקת קשרי ציבור וממשל, מינהל רמו"ט</p>	15:30–15:00
<p>הגנת הפרטיות כשאלה של ניהול סיכונים – תגובה לדברי לימור שמרלינג  <b>פרופ' טל ז'רסקי</b>, אוניברסיטת חיפה, חבר צוות המומחים</p>	15:45–15:30
<p>סיכום ותובנות מיום העיון ומעבודת צוות המומחים:  <b>פרופ' מיכל בלר</b>, <b>מכללת לוינסקי</b>, חברה בצוות המומחים</p>	16:00–15:45

## נספח ב': תקצירי קורות החיים של השותפים העיקריים למהלך הלימודי

### חברי צוות המומחים

(לפי סדר א"ב של שמות המשפחה)

**לאה אחדות**, יו"ר צוות המומחים. פרופ' אחדות היא חברה בצוות התכנית לכלכלה וחברה במכון ון ליר בירושלים וכן היא פרופסור-חבר וראש המחלקה לכלכלה ומנהל, בית הספר לכלכלה ומנהל עסקים, המרכז האקדמי רופין. במהלך עבודתה בשירות הציבורי מילאה תפקידי ניהול ביחידות מחקר של המוסד לביטוח לאומי, ניהלה את המכון למחקר כלכלי-חברתי של ההסתדרות, שימשה סמנכ"לית למחקר ולתכנון של המוסד לביטוח לאומי והייתה חברה במינהלת המוסד. בתקופה זו עמדה פרופ' אחדות בראש ועדות שעסקו בפיתוח תשתיות המידע והמחקר ובפיתוח מיזמים חברתיים בקהילה. היא הופקדה על הכנת דוח העוני ועל הסקירה השנתית של הביטוח הלאומי, השתתפה בוועדות ממשלתיות וציבוריות כחברה או כמשקיפה וייצגה את ישראל בפורומים בינלאומיים. פרופ' אחדות הייתה חברה בחבר המנהלים של הפרויקט הבינלאומי למחקר ההכנסות (LIS) ושימשה נציגת הביטוח הלאומי בוועדת המחקר של הארגון הבינלאומי של מוסדות הביטוח הסוציאלי (ISSA). מאז 2004 פרופ' אחדות שותפה גם לפרויקט סקר המעקב האירופי SHARE בתחומי בריאות ופרישה.

בעלת תואר ראשון בכלכלה וסטטיסטיקה ותואר שני בכלכלה – שניהם מהאוניברסיטה העברית בירושלים.

**מיכל בלר**, נשיאת מכללת לוינסקי לחינוך. תחום התמחותה העיקרי הוא מדידה והערכה חינוכית. הייתה המנכ"לית המייסדת של הרשות הארצית למדידה והערכה בחינוך (ראמ"ה), ועמדה בראשה מאז הקמתה, בשנת 2005. עד כניסתה לתפקידה זה הייתה פרופ' בלר מנהלת מחקר בכירה במחלקת המחקר והפיתוח של מכון המבחנים הגדול בעולם – Educational Testing Service (ETS) בפרינסטון, ניו-ג'רסי. במשך כשמונה שנים עמדה פרופ' בלר בראש המרכז הארצי לבחינות ולהערכה (מיסודן של האוניברסיטאות בישראל). קודם לכן הייתה חברת סגל אקדמי בדרגת פרופסור-חבר במחלקה לחינוך ופסיכולוגיה באוניברסיטה הפתוחה, שם גם הקימה וניהלה את מרכז שה"ם (שילוב טכנולוגיות בלמידה). פרופ' בלר פרסמה מאמרים רבים בתחום המדידה וההערכה החינוכית. היא השתתפה בכנסים בינלאומיים רבים, שבהם ייצגה את ישראל, והיא משמשת חברה בוועדות היגוי מקצועיות בארץ ובעולם. בעלת תואר שלישי מהמחלקה לפסיכולוגיה באוניברסיטה העברית בירושלים, 1983.

הייתה חברה ועדת היזמה לנושא 'מדידה והערכה בחינוך'; פרשה מן הוועדה עם מינויה לתפקיד מנכ"לית ראמ"ה.

**נעם זוסמן**, כלכלן בחטיבת המחקר של בנק ישראל.

**טל ז'רסקי**, פרופסור-חבר בפקולטה למשפטים באוניברסיטת חיפה. חבר בוועדת ההיגוי לשירותי תשתיות של איגוד האינטרנט. תחומי העניין שלו הם דיני קניין, משפט וטכנולוגיה, דיני תקשורת, דיני פרטיות ומסחר אלקטרוני. מחקריו מתרכזים בנושאי אינטרנט, פרטיות ברשת, תקשורת ונושאים אחרים הקשורים למשפט וטכנולוגיה.

בעל תואר דוקטור למשפטים (J.S.D.) מאוניברסיטת קולומביה, ארצות הברית, 2004.

**איריס טבק**, מרצה בכירה במחלקה לחינוך באוניברסיטת בן-גוריון בנגב, ראש המסלול לתכניות לימודים והוראה. ד"ר טבק היא העורכת הראשית (עם ג'וש רדינסקי) של כתב העת *Journal of the Learning Sciences*, ונשיאה לשעבר של האגודה הבינלאומית למדעי הלמידה (ISLS). מחקרה עוסקים בידע מושגי, מיומנויות חשיבה וזהות מקצועית. היא בונה מודלים של ראייה מקצועית (professional vision), ועל בסיסם מפתחת גישות הוראה וטכנולוגיות למידה. במסגרת זו היא פיתחה את גישת המורה כשותף (teacher as partner), ואת המושג פיגומים סינרגטיים (synergistic scaffoldings). במטרה להנגיש את

המדע, היא בוחנת כיצד ניתן להשביח את השימוש במידע מקוון לקבלת החלטות מבוססות ראיות. מחקרים אלה מתקיימים במסגרת מתודולוגית רחבה של מחקר עיצוב (design-based research) ומחקרי מעבדה. בעלת תואר שלישי במדעי הלמידה מאוניברסיטת נורת'ווסטרן (ארצות הברית), 1999. הייתה חברת ועדת היזמה לנושא 'שפה ואוריינות'.

**אורלי מנור** נמנית החל משנת 1990 עם הסגל האקדמי של בית הספר לבריאות הציבור ורפואה חברתית של האוניברסיטה העברית והדסה. בשנים 2009–2012 כיהנה כמנהלת בית הספר, ובשנת 2012 קיבלה דרגת פרופסור מן המניין. מאז 2014 מכהנת פרופ' מנור כיו"ר הוועד המנהל של המכון הלאומי לחקר שירותי הבריאות ומדיניות הבריאות. מנור פרסמה למעלה מ-150 מאמרים מדעיים, קיבלה מענקי מחקר רבים וזכתה בפרסי האוניברסיטה על הוראה ומחקר. היא הייתה פרופסור-אורח באוניברסיטת פנסילבניה, במכון קרולינסקה בשטוקהולם ובאוניברסיטת ז'נווה. משנת 2010 עד היום עומדת פרופ' מנור בראש התכנית הלאומית למדדי איכות לרפואת הקהילה.

בעלת תואר ראשון בכלכלה וסטטיסטיקה ותואר שני בסטטיסטיקה – שניהם מהאוניברסיטה העברית בירושלים – ותואר שלישי בסטטיסטיקה מאוניברסיטת שטוקהולם, 1985.

## הדוברים ביום העיון

(לפי סדר ההופעה, דוברים שאינם חברי צוות המומחים)

**דייוויד קפלן (David Kaplan)** עומד בראש הקתדרה על שם פטרישה באסק (P. Busk) לשיטות כמותניות במחלקה לפסיכולוגיה חינוכית של אוניברסיטת ויסקונסין-מדיסון. ד"ר קפלן מחזיק בשני מינויים נוספים באוניברסיטת ויסקונסין: במחלקה למדעי רפואת ההמונים ובמרכז לדמוגרפיה ואקולוגיה, ואף נושא תואר עמית-כבוד של אוניברסיטת אוקספורד, שבה הוא משתתף במחקר הנערך במחלקה לחינוך. ד"ר קפלן הוא חבר האקדמיה הלאומית לחינוך בארצות הברית, חתן פרס המבולדט למחקר, עמית באגודת הפסיכולוגיה של ארצות הברית (חטיבה 5), ושימש עמית מחקר בקתדרה על שם ג'ין גריפית' (J. Griffith) במרכז הלאומי לסטטיסטיקה חינוכית.

בעל תואר שלישי בחינוך מאוניברסיטת UCLA, 1987.

**משה יוסטמן**, פרופסור אמריטוס במחלקה לכלכלה באוניברסיטת בן גוריון, כיהן בעבר כדיקן הפקולטה למדעי הרוח והחברה. תחומי המחקר העיקריים שלו הם כלכלה פוליטית של חינוך, שוויון הזדמנויות ונגישות להשכלה, מבנה מערכת החינוך ומקורות המימון שלה, השפעת מערכת החינוך על המשק ומדידה בשדה החינוך. כן חקר בהרחבה נושאים במדיניות תעשייתית-טכנולוגית, ונושאים בפיתוח אזורי. עמית בכיר במכון ון ליר בירושלים משנת 2007 ומרכז תחום חברה אזרחית מ-2012.

בעל תואר שלישי בכלכלת עסקים מאוניברסיטת הרווארד (ארצות הברית), 1982.

היה יו"ר ועדת היזמה לנושא 'רענון מערך האינדיקטורים לחינוך בישראל'. בשנים 2011–2013 היה חבר ועדת ההיגוי של היזמה.

**ויקטור לביא**, פרופסור לכלכלה באוניברסיטה העברית בירושלים, מופקד הקתדרה ע"ש וויליאם הבר, ופרופסור במחלקה לכלכלה באוניברסיטת ורוויק, אנגליה. תחומי המחקר שלו הם כלכלת עבודה, כלכלת חינוך וכלכלת פיתוח. פרופ' לביא חבר זה שנים רבות בגופים מובילים לחקר הכלכלה, במיוחד באנגליה ובארצות הברית. כן שימש מרצה-אורח באוניברסיטאות מן המובילות בעולם, והיה שותף למחקרים רבים בתחום הכלכלה. הוא חבר פעיל בגופי מחקר רבים העוסקים בניתוחים כלכליים, וכן באגודה הכלכלית האמריקאית, בחברה האקונומטרית, בחברה המלכותית לכלכלה בבריטניה ובאגודת כלכלני העבודה.

בעל תארים אקדמיים מהאוניברסיטה העברית ומאוניברסיטת שיקגו, ותואר שלישי מאוניברסיטת שיקגו, 1979.

**עמליה רן**, מרצה וחוקרת בתחום של לימודי אמריקה הלטינית. ד"ר רן מתמחה בתרבות וספרות של אמריקה הלטינית וביהדות היבשת, והרבתה לפרסם בנושאים אלה. היא חברה במספר ארגונים אקדמיים, בהם האגודה ללימודים לטינו-אמריקניים והאגודה ללימודי היהדות באמריקה הלטינית, שבה היא מכהנת כחברה בוועד המנהל. ד"ר רן הייתה חברת סגל באוניברסיטת נברסקה, ושירתה כשליחת הסברה מטעם הסוכנות היהודית באוניברסיטת מרילנד. בשנת 2009 שבה ארצה במסגרת תכנית משרד הקליטה להשבת מדענים וחוקרים לישראל. בשנת 2014 הקימה עסק המספק שירותי תרגום ועריכה ספרותיים ואקדמיים בשפות ספרדית, אנגלית ועברית. ד"ר רן גם מרצה ומנחה סדנאות בנושאי תרבות עסקית, ספרות, קולנוע ויהדות באמריקה הלטינית, יחסי ישראל והתפוצות, נשים ותרבות. בעלת תואר ראשון ושני מהאוניברסיטה העברית בירושלים ותואר שלישי מאוניברסיטת מרילנד בארצות הברית, 2007.

**אלין אטיאס**, מנהלת מערך Big Data והתממה במשרד הבריאות, האגף לבריאות דיגיטלית ומיחשוב. בין השנים 1995 ל-2014 עסקה ביעוץ, הקמה וניהול של מחסני נתונים רחבי היקף ותשתיות לתבונה עסקית (Business Intelligence), בעיקר במגזר הפרטי. בתפקידה האחרון ניהלה את מחלקת BI ותשתיות נתונים בחברת הביטוח הפניקס.

בעלת תואר בהנדסת מערכות מידע מהטכניון, ותואר שני בניהול מאוניברסיטת בוסטון בארצות הברית, 2000.

**יואל פינקל**, המשנה לסטטיסטיקן הלאומי בלשכה המרכזית לסטטיסטיקה (למ"ס). בתחילת דרכו בלשכה היה כלכלן ביחידת מדד המחירים לצרכן, ומאז התקדם לתפקיד מתפקיד במידרג הלשכה: סגן מנהל יחידה, מנהל יחידה, מנהל אגף, סגן מנהל הלמ"ס וכיום המשנה לסטטיסטיקן הממשלתי. כמשנה לסטטיסטיקן משמש פינקל שותף בכיר בהובלת הלמ"ס למעמדה המקצועי המוערך בארץ ובחול. בהשבת הדרג הבכיר של שירות המדינה, בלימוד משותף בחברותה וביצירת קשרי עבודה בין האנשים בדרג הזה רואה פינקל תנאי הכרחי ליישום רפורמה אמיתית בשירות המדינה ולשיפור השירות למען הציבור במדינת ישראל. בעל תואר ראשון בכלכלה ומינהל עסקים, תואר שני במינהל עסקים, שניהם מהאוניברסיטה העברית, ותואר שני במינהל ציבורי (1994) מאוניברסיטת הרווארד כעמית של קרן וקסנר.

**יובל אילון**, חבר סגל במחלקה להיסטוריה, פילוסופיה ולימודי יהדות באוניברסיטה הפתוחה. ד"ר אילון הרבה לפרסם בנושאי אתיקה ומשפט, והשתתף בכנסים מדעיים בארץ ובעולם. תחומי המחקר שלו הם אתיקה, מטא-אתיקה, פילוסופיה פוליטית ופילוסופיה של הספורט. בעל תואר שלישי בפילוסופיה מהאוניברסיטה העברית, 2003.

**לימור שמרלינג מגזניק**, מומחית ברגולציה ואכיפה של הגנת הפרטיות במאגרים ממוחשבים של מידע אישי. עו"ד שמרלינג מגזניק היא מנהלת המחלקה לקשרי ציבור וממשל ברשות למשפט טכנולוגיה ומידע (רמו"ט) במשרד המשפטים, וקודם לכן הייתה מנהלת מחלקת רישוי ופיקוח במשרד. מאז 2009 היא עובדת ברמו"ט, שם ניהלה את האכיפה המינהלית והאסדרה מכוח חוק הגנת הפרטיות, חוק שירותי נתוני אשכנזי, וחוק החתימה האלקטרונית. היא עסקה בחקירות מחשב, בהטמעת עקרונות של אבטחת מידע והגנת סייבר, בניהול תקין של מאגרי מידע ובהליכי הטלת סנקציות מינהליות בנושאים כגון צמידות מטרת שימוש במידע, הסכמה ואבטחת מידע. עו"ד שמרלינג מגזניק משתתפת מטעם רמו"ט בוועדות ציבוריות בתחומים הקשורים להגנת הפרטיות כגון הוועדה המייעצת המפקחת על תקופת המבחן של המאגר הביומטרי, ועדות בנושא הנגשת מאגרי מידע לציבור והעברות מידע בין גופים ציבוריים, תשתיות מידע, מיחשוב ענן, רשת ממשלתית ומאגר נתוני אשכנזי. בשנת 2008 שימשה מזכירת המועצה לשידורי כבלים ולוויין במשרד התקשורת ויועצת ליו"ר. לפני כניסתה לשירות הציבורי עסקה עו"ד שמרלינג מגזניק ביעוץ וייצוג משפטי בתחומים דיני חברות, חוזים, קניין ובנקאות, והחזיקה בתפקידי ניהול מוצר וניהול פרויקטים בתעשיית ההייטק. בעלת תואר ראשון במשפטים, תואר שני במשפט ציבורי, ותואר שני בספרות – כולם מאוניברסיטת תל אביב.

## הדוברים במפגשים הלימודיים

(לפי סדר א"ב של שמות המשפחה)

**זיו אופק**, המייסד והמנכ"ל של CDI-נגב, מרכז המצוי בחוד החנית של החדשנות הדיגיטלית ופועל ללא מטרות רווח. המרכז הוקם הודות למאמצים משותפים של אוניברסיטת בן גוריון בנגב וכמה מן היזמים המובילים בישראל בתחום הבריאות הדיגיטלית ביניהם זיו אופק, יזם עתיר ניסיון ומייסדה של חברת dbMotion שנרכשה בידי Allscripts האמריקאית תמורת 235 מיליון דולר. אופק הוא ממציא המודל והמערכת של dbMotion לשיתוף מידע סמנטי בסביבה הטרגונית ומבזרת ולניהול אוכלוסיות מטופלים, פתרון המותקן ופועל בהצלחה בקרב עשרות אלפי משתמשים במאות רבות של ארגונים רפואיים בארץ ובעולם. לפני עבודתו ב־dbMotion כיהן אופק כסגן־נשיא למחקר ופיתוח ב־Ness-ISI, חברה־בת של Ness Technologies. במסגרת זו יזם, תכנן וניהל בין היתר את פיתוחה ויישומה של פלטפורמת שיתוף המידע עבור שירותי בריאות כללית, שאומצה מאוחר יותר על ידי משרד הבריאות כפלטפורמה הלאומית לשיתוף מידע רפואי בישראל. לזיו אופק ניסיון עשיר – תאורטי ומעשי – של למעלה מ־20 שנה בתחום מערכות מידע רפואיות וארגוניות, מערכות תומכות־החלטה, Big Data ומערכות מידע למנהלים בארץ ובעולם. בעל תואר B.Sc. בהנדסת תעשייה וניהול, התמחות במערכות מידע, מאוניברסיטת בן גוריון בנגב.

**איתי אשר**, ממונה ניסויים ומחקרים חינוכיים וממלא מקום המדען הראשי במשרד החינוך. במסגרת עבודתו בלשכת המדען הראשי מתמקד במיוחד בתחומי התוכן הבאים: טיפוח מיומנויות המאה ה־21, קידום שוויון הזדמנויות בחינוך ואוטונומיה בית ספרית. בעבר שימש כראש תחום מחקר ופיתוח במכון אבני ראשה וכעמית הוראה באוניברסיטה העברית בירושלים ובאוניברסיטת בר אילן. במסגרת פוסט־דוקטורט בהוראת המדעים באוניברסיטת בן גוריון בנגב ניהל ד"ר אשר את הצוות הישראלי בפרויקט המחקר האירופי CoReflect, שעסק בלמידת חקר של סוגיות סוציו־מדעיות. ד"ר אשר הוא בוגר בי"ס מנדל למנהיגות חינוכית. בעל תואר שלישי מהאוניברסיטה העברית בירושלים בתחום החישוב והעיבוד של מידע במוח.

**יוסי גדיניאן**, מנהל אגף חינוך, השכלה וכוחות הוראה בלשכה המרכזית לסטטיסטיקה. בעבר ניהל את אגף הדמוגרפיה בלשכה.

בעל תואר ראשון במתמטיקה וסטטיסטיקה ותואר שני בדמוגרפיה, שניהם מהאוניברסיטה העברית בירושלים.

**חגית גליקמן**, מנכ"לית הרשות הארצית למדידה והערכה בחינוך (ראמ"ה). בשנים 2006–2013 ניהלה את האגף למחקר ופיתוח בראמ"ה ומשנת 2013 היא משמשת כמנהלת הרשות. בתפקידה עוסקת ד"ר גליקמן במגוון רחב של מחקרים בתחום החינוך; בכלל זה, מבחנים וסקרים רחבי היקף, מחקרי רוחב ואורך בנושאים השונים, הערכת אפקטיביות של תכניות חינוכיות ופיתוח מתודולוגיות סטטיסטיות ופסיכומטריות. בעלת תואר שלישי בסטטיסטיקה מהאוניברסיטה העברית בירושלים, 2000.

**אדם גמראן** (Adam Gamoran), נשיא קרן William T. Grant Foundation, שאליה הצטרף בספטמבר 2013. קודם לכן היה מופקד הקתדרה על שם ג'ון מק'ארתור ללימודי סוציולוגיה ומדיניות חינוך באוניברסיטת ויסקונסין־מדיסון, שם ניהל את מרכז ויסקונסין למחקר חינוכי ועמד בראש תכנית ההכשרה הבינתחומית לתלמידים מתקדמים במדעי החינוך. פרופ' גמראן, בעל ניסיון בן שלושה עשורים במחקר, ניהל מגוון רחב של מחקרים המתמקדים באי־שוויון בחינוך וברפורמות בית־ספריות. בקרן גרנט הוא מוביל יזמה חדשה לתמיכה במחקר על תכניות, קווי מדיניות ופרקטיקות שנועדו לצמצם אי־שוויון, והוא ממשיך במאמציה המתמשכים של הקרן להבין את השימוש בראיות מחקריות בעת קבלת החלטות. פרופ' גמראן הוא המחבר הראשי של ספרים חשובים בהוראת מתמטיקה ומדעים, בשינויים במערכות חינוך ובריבוד סוציו־אקונומי בחינוך. פרופ' גמראן היה יושב הראש של הוועדה המייעצת בנושא ההערכה הכלל־ארצית של חינוך לקריירה ולחינוך טכני מטעם משרד החינוך של ארצות הברית, וכיום הוא יושב הראש של הוועדה לחינוך ומדע מטעם מועצת המחקר הלאומית (NRC). הוא חבר נבחר של האקדמיה הלאומית לחינוך ושל האקדמיה

הלאומית למדעים ואמנויות בארצות הברית. הנשיא אובמה מינה אותו פעמיים לחברות בוועד הלאומי למדעי החינוך.

**חיים גת**, מנהל מערכות מידע, הוא האחראי על המחקר והסטטיסטיקה במינהל התקשוב של משרד החינוך. ד"ר גת מופקד בלשכת המדען הראשי על פיתוח וחישוב של המדד החברתי-כלכלי של משרד החינוך (מדד הטיפוח), והוא אחראי גם על חדר המחקר הווירטואלי של משרד החינוך. במסגרת עבודתו עוסק חיים גת בבניית מודלים סטטיסטיים, תחזיות, מדדים וסקרים ובניתוח נתונים כולל, באמצעות שיטות של כריית נתונים. בעבר לימד ד"ר גת באוניברסיטה העברית ובחברת IBM-SPSS קורסים בשיטות מחקר ובשיטות איסוף וניתוח של נתונים. חיים גת הוא גם יועץ עצמאי בתחום המחקר, ופסיכולוג קליני מומחה. בעל תואר שלישי בפסיכולוגיה מהאוניברסיטה העברית.

**נורית דוברין**, מנהלת סקרים חברתיים בלשכה המרכזית לסטטיסטיקה. בעשור האחרון היא מנהלת את הסקר החברתי של הלמ"ס בנושאים סוציאקונומיים מגוונים: דיור, השכלה, תעסוקה, מדידה רב-ממדית של עוני, בריאות ואורח חיים, הסדרי פנסיה, איכות הסביבה, שירותי ממשל ועוד. במהלך השנים כתבה ניירות עבודה בנושאי הסקר השונים: ניתוח גורמים לקידום במקום העבודה – ההיבט המגדרי, מצבם של בעלי מוגבלויות בישראל, שוויון הזדמנויות בהשכלה ועוד. נורית דוברין מנהלת גם את סקר אמון הצרכנים שעליו מבוסס מדד הצרכנים החודשי ואשר מופק בהתאם למדריך הבינלאומי של ה-OECD. דוברין מופקדת גם על סקר ארוך טווח שעוקב אחר פרטים ומשקי בית לאורך שנים, על מנת לזהות דינמיקה בתחומי התעסוקה, הדיור ועוד. הסקר הוא מיזם משותף של משרד האוצר, משרד החינוך, המוסד לביטוח לאומי, בנק ישראל והלמ"ס, והוא ייחודי באופן ניהולו ובתוצריו.

בעלת תואר מוסמך בסוציולוגיה ואנתרופולוגיה, ובדמוגרפיה – מהאוניברסיטה העברית בירושלים.

**טלל דולב**, המקימה והמנהלת של 360°, התכנית הלאומית לילדים ולנוער בסיכון. התכנית הוקמה בשנת 2008, והגב' דולב מנהלת אותה מאז. את התכנית מנהלים במשותף חמישה משרדי ממשלה (רווחה, חינוך, קליטה, בריאות, ביטחון פנים), הגוינט והמרכז לשלטון מקומי, והיא אחת התכניות הבין-משרדיות הבולטות בארץ. התכנית פרוסה ב-180 רשויות, בהיקף תקציבי של כ-200 מיליון שקל בשנה, והיא משרתת כ-60,000 ילדים ב-1,500 תכניות ומסגרות. כחלק מהתכנית נערך לראשונה מיפוי של כלל הילדים בסיכון בישראל, ומופעלת מערכת מידע בין-משרדית, יחידה במינה, המאפשרת מעקב אחר השינויים במצב הילדים ואחר התקדמותם. כחלק מתפקידה גב' דולב חברה בפורום הנהלת משרד הרווחה והשירותים החברתיים ובוועדות משרדיות ובין-משרדיות, והיא מסייעת בתכנון וביצוע של החלטות ממשלה שונות הנוגעות למשרד הרווחה וגם למשרדים שותפים. לפני שנתמנתה למנהלת התכנית הייתה טלל דולב שותפה בהקמתו של מרכז אנגלברג לילדים ונוער במאירס גוינט, מכון ברוקדייל. היא ניהלה את המרכז ועסקה בו בגיבוש אסטרטגיה וביצוע מחקרים שתוצאתם היא הגברת רווחת הילדים ובני הנוער בישראל. בתפקידה זה כיהנה בוועדות ממלכתיות שונות ובהן ועדת גולדברגר לגיבוש תכנית לרפורמה בוועדות 'תכנון טיפול והערכה'. הגב' דולב הייתה יועצת לוועדה לזכויות הילד בכנסת ולוועדת ראש הממשלה לילדים ולנוער בסיכון (ועדת שמיד).

בעלת תואר ראשון ושני בעבודה סוציאלית מהאוניברסיטה העברית.

**רן וולף**, ד"ר וולף מועסק במעבדת המחקר של יאהו בחיפה. בעבר היה חבר סגל בחוג למערכות מידע באוניברסיטת חיפה. את שנת הבת-דוקטורט שלו עשה באוניברסיטת מרילנד. לד"ר וולף עשרות מאמרים מדעיים הנוגעים לכריית נתונים ולממשק בין כריית נתונים לבין הגנה על הפרטיות. בנוסף לכך ד"ר וולף חוקר ומלמד סוציולוגיה ואתיקה של הפרטיות. במסגרת עבודתו ביאהו, ד"ר וולף מייץ לוועדת הפרטיות הארגונית בנוגע לשחרור נתוני משתמשים לצורכי מחקר.

בעל תואר שלישי מהחוג למדעי המחשב בטכניון.

**מיכל טביאן-מזרחי**, סמנכ"לית אסטרטגיה ותכנון במשרד החינוך, תפקיד שהיא ממלאת מאז הוקם האגף, בשנת 2014. בתפקידה הקודם הייתה ראש תחום בכירה באגף לממשל וחברה (תכנון מדיניות) במשרד ראש הממשלה, שבו פעלה לקידום רפורמות בעבודת הממשלה ולפיתוח תורת התכנון הממשלתית ומעגל התכנון השנתי. לפני כן עבדה כחוקרת בכירה במרכז המחקר והמידע של הכנסת, שם הייתה שותפה בעיצוב תהליכי חקיקה ומדיניות, והתמקדה בנושאים מתחומי העבודה, הבריאות והממשל. עבודת הדוקטורט שלה עסקה בנושא 'דינמיקה של שינוי מוסדי הדרגתי: בחינת השינויים בהעסקת עובדים בשירות המדינה 1950–2000'.

בעלת תואר שלישי בסוציולוגיה ואנתרופולוגיה מהאוניברסיטה העברית.

**מנחם יערי**, יו"ר ועדת ההיגוי של היזמה. נשיא (אמריטוס) האקדמיה הלאומית הישראלית למדעים. פרופסור (אמריטוס) לכלכלה, האוניברסיטה העברית בירושלים. חבר האקדמיה הלאומית האמריקנית למדעים ולאמנויות, החברה הפילוסופית האמריקנית והאקדמיה למדעים ברלין-ברנדנבורג. חתן פרס ישראל לכלכלה (1987) וחתן פרס רוטשילד למדעי החברה (1994). חתן פרס א.מ.ת. למדעי החברה (כלכלה) לשנת 2012.

כיהן כחבר ועדת ההיגוי של היזמה למן היווסדה, ומאז 2008 הוא יושב הראש שלה.

**אוריה לוי**, עובדת הלשכה המרכזית לסטטיסטיקה, מרכזת מקצועית בכירה בתחום הסקר ארוך הטווח הנערך בלשכה ואחראית בין השאר על פרק החינוך בסקר. במסגרת לימודיה לתואר שלישי היא כותבת עבודה בנושא 'הגירה ורווחה סובייקטיבית של ילדים בישראל', המתבססת על נתונים מתוך הסקר ארוך הטווח. בעלת תואר ראשון בסטטיסטיקה, סוציולוגיה ואנתרופולוגיה ותואר שני בדמוגרפיה מהאוניברסיטה העברית בירושלים.

**עדנאן מנסור**, ראש תחום מעקבים ומהלכי לימודים בלשכה המרכזית לסטטיסטיקה. התחום מבצע בעיקר מעקבי אורך בחינוך – כניסה של מסיימי תיכון ומכינות קדם-אקדמיות למוסדות להשכלה גבוהה, מהלך לימודים של סטודנטים וכניסה לשוק העבודה של בוגרי המוסדות להשכלה גבוהה. בנוסף לכך מפתחים בתחום סטטיסטיקה בנושא מחקר ופיתוח בהשכלה הגבוהה. במסגרת זו מתפרסמים נתונים על הוצאות למחקר ופיתוח, עבודות בתרדוקטורליות ועוד. התחום גם מרכז את העבודה מול החוקרים שמתעניינים בנתוני חינוך לצורך עבודה בחדר המחקר של הלמ"ס.

בעל תואר ראשון בסטטיסטיקה, תואר ראשון במשפטים, לימודי תעודה בראיית חשבון לבעלי תואר אקדמי ועומד לסיים תואר שני בכלכלה.

**דורית מורג**, היועצת המשפטית של משרד החינוך.

**דוד מעגן**, מנהל תחום סטטיסטיקה של חינוך גבוה וכוחות הוראה בלשכה המרכזית לסטטיסטיקה. במסגרת עבודתו ביצע מעגן שורה של מחקרים שהוצגו בכנסים מקצועיים, על היבטים שונים של כוח אדם בהוראה, הישגי תלמידים ומעקבי אורך על אוכלוסיות שונות במערכת החינוך. בנוסף לכך, מעגן הוא עמית מדיניות (חינוך) בתכנית מדיניות חינוך במרכז טאוב לחקר המדיניות החברתית בישראל. הוא שימש חבר בוועדה של משרד החינוך לבדיקת הכשרת מורים למגזר הערבי, והיה חבר בוועדת היגוי מטעם ראמ"ה לסקר בינלאומי של ה-OECD Teaching and Learning International Survey (TALIS). כן שימש משקיף בתת-ועדת שכר מורים של ועדת דוברת (כוח המשימה הלאומי לקידום מערכת החינוך). מעגן גם ייצג את ישראל בקבוצת עבודה של ה-OECD, INES Network on Labour Market, Economic and Social Outcomes of Learning.

בוגר תואר שני בסוציולוגיה מהאוניברסיטה העברית וכיום לומד לקראת תואר שלישי בחוג לחינוך באוניברסיטת בריאלין.



**בריאן ניגן**, היועץ המשפטי וממונה חופש המידע של הלשכה המרכזית לסטטיסטיקה (למ"ס). עו"ד ניגן הוא גם מרצה מן החוץ בנושא 'מדיניות מידע וסוגיות באתיקה של המידע' במכללה האקדמית לחינוך ע"ש דוד ילין. כחלק בלתי נפרד מעבודתו בלמ"ס, עו"ד ניגן חבר בוועדת הסודיות ובוועדת אבטחת מידע והעברת מידע בין גופים ציבוריים לבין הלמ"ס. כן הוא ממונה על ההיבטים המשפטיים של הנגשת נתוני פרט לחוקרים על ידי הלמ"ס. בריאן ניגן יוזם גם פעילות 'הנגשת נתוני פרט לחוקרים' בפרויקט טווינינג (Twinning) הנוכחי של האיחוד האירופי עם Statistics Denmark. בשנים 2011–2013 היה נציג הלמ"ס בצוות מומחים של ה-OECD בנושא שיתוף פעולה בינלאומי בהנגשת נתוני פרט לחוקרים. עו"ד ניגן יוזם ויוצר תנאי שימוש חדשים לאתר האינטרנט של הלמ"ס לפי עקרונות הממשל הפתוח (open government). בעל תואר ראשון במדעי המדינה מאוניברסיטת קליפורניה, סנטה ברברה (ארצות הברית) ותואר ראשון במשפטים מהאוניברסיטה העברית בירושלים.

**פנחס קליין**, כלכלן ומרפז בכיר (עיבוד וניתוח סטטיסטי) באגף כלכלה וסטטיסטיקה, מינהל כלכלה ותקציבים במשרד החינוך. פנחס קליין הוא חלק מצוות הפועל בניהולו של יואב אזולאי, מנהל האגף. הצוות מופקד על גיבוש ועדכון תמונת המצב הסטטיסטית של משרד החינוך, ומשמש אינטגרטור של גופים סטטיסטיים בארץ ובעולם בתחום החינוך. אנשי הצוות אחראים גם על בנייה, פיתוח והנגשה של כלים אופרטיביים על ידי שימוש בנתונים סטטיסטיים – הן ליחידות המשרד על פי צורכיהן והן לכלל הציבור. בתפקידו הקודם בלשכה המרכזית לסטטיסטיקה עסק פנחס קליין בתחום כלכלת חינוך. בעל תארים אקדמיים בתעשייה וניהול, כלכלה ומינהל עסקים.

**עליזה קרובי**, ראש תחום איזופד ומרשם השכלה באגף חינוך והשכלה בלשכה המרכזית לסטטיסטיקה. תחום זה אחראי על תהליך השינוי של שיטת העבודה על נתוני חינוך: מעבודה על קבצים בודדים במחשבים האישיים ותוצרים בנושא אחד בשנה אחת – לעבודה מול בסיס נתונים טבלאי מקושר, המאפשר את שליפת ההיסטוריה ההשכלתית של אדם מכל המקורות המינהליים שבהם הופיע במהלך השנים, זאת תוך העלאת איכות הנתונים ואחידותם.

בעלת תואר ראשון מהפקולטה לחקלאות, מזון וסביבה של האוניברסיטה העברית ובוגרת המסלול להסבת אקדמאים למתכנתי מערכות מידע, במכללת ההייטק ג'ון ברייס.

**נרדיט שטיין-קאפח**, מרכזת סקר ארוך טווח בלשכה המרכזית לסטטיסטיקה. בעבר ריכזה את סקר ההכנסות של הלמ"ס.

בעלת תואר מוסמך בסטטיסטיקה רשמית מהאוניברסיטה העברית בירושלים.

## צוות היזמה

**אביטל דרמון**, מקימת היזמה והמנהלת שלה מאז 2004. עסקה בהוראת ביולוגיה בירושלים, יסדה וניהלה את המרכז הארצי למורי ביולוגיה של האוניברסיטה העברית ומשרד החינוך, וניהלה את בית הספר למנהיגות חינוכית של מכון מנדל. כיהנה כחברה בוועדות מקצוע ותכניות לימודים של משרד החינוך (ביולוגיה, מדע וטכנולוגיה) ובוועדות בנושא מורים.

בעלת תואר שלישי בביולוגיה, 1985, ותעודת הוראה, 1988 – שניהם מהאוניברסיטה העברית בירושלים. בוגרת בית הספר מנדל למנהיגות חינוכית, 1994.

**עודד בושריאן**, מרפז אקדמי ביזמה למחקר יישומי בחינוך של האקדמיה הלאומית למדעים. במסגרת תפקידו ריכז את ועדת הקונסנסוס בנושא קשרי הורים מורים בסביבה משתנה, וכן ריכז צוותי מומחים בנושאים 'התערבויות טיפוליות בילדים בעלי בעיות או הפרעות בהתנהגות', 'אי-שוויון וחינוך' ושימוש בנתונים

מינהלים הנאספים לאורך זמן ובסקרי אורך לצורכי מחקר וגיבוש מדיניות מושכלת בתחום החינוך. קודם לכן שימש עוזר-מחקר ועוזר-הוראה במחלקה למדע המדינה באוניברסיטה העברית, ולאחר סיום לימודיו היה מתרגל בקורס מימון בתכנית המנהלים של בית הספר למדיניות ציבורית באוניברסיטה העברית ובקורס סטטיסטיקה לתלמידי מדעי החברה של האוניברסיטה הפתוחה. בעל תואר ראשון בתכנית המשולבת לפילוסופיה, כלכלה ומדע המדינה, 2007, ותואר שני במדע המדינה – שניהם מהאוניברסיטה העברית בירושלים, 2011.

---

**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