

הישגים במבחנים בין-לאומיים והקשר שלהם לצמיחה כלכלית

נחום בלס, מיכאל דבאוי ואלכס וינרב

נייר מדיניות מס' 04.2023

ירושלים, ניסן תשפ"ג, אפריל 2023

מרכז טאוב לחקר המדיניות החברתית בישראל

מרכז טאוב נוסד ב-1982 ביוזמתם של הרברט מ' סינגר, הנרי טאוב וארגון הג'וינט האמריקאי. המרכז ממומן באמצעות קרן צמיתה שהוקמה על ידי קרן הנרי ומרלין טאוב, קרן הרברט ונל סינגר, ג'ין וג'ון קולמן, קרן משפחת קולקר-סקסון-הלוק, קרן משפחת מילטון א' ורוזלין ז' וולף וארגון הג'וינט האמריקאי.

מרכז טאוב הוא מכון מחקר על-מפלגתי ובלתי תלוי העורך מחקרים איכותיים בנושאי חברה וכלכלה בישראל. הוא מציג בפני מקבלי ההחלטות המובילים בארץ ובפני כלל הציבור תמונה רחבה המשלבת בין הממדים החברתיים והכלכליים בהתוויית המדיניות הציבורית. הצוות המקצועי של המרכז וצוותי המדיניות הבין-תחומיים, הכוללים חוקרים וחוקרות בולטים בתחום באקדמיה ומומחים ומומחיות מובילים בתחומי המדיניות, עורכים מחקרים ומציעים חלופות למדיניות בנושאים חברתיים-כלכליים מרכזיים שעל סדר היום במדינה. המרכז מציג ניתוחים אסטרטגיים לטווח ארוך והערכות של חלופות מדיניות בפני הציבור ובפני מקבלי ההחלטות על ידי כתבות בתקשורת, תוכנית פרסומים פעילה, כנסים ופעילויות אחרות בישראל ובחו"ל.

הפרסומים של מרכז טאוב הם על דעתם ועל אחריותם של מחברייהם בלבד. אין בהם כדי לחייב את המרכז, את חבר הנאמנים שלו, את עובדיו האחרים ואת התומכים בפעולותיו.

כתובת המרכז: רחוב האר"י 15, ירושלים

טלפון: 02-567-1818

פקס: 02-567-1919

דואר אלקטרוני: info@taubcenter.org

אתר אינטרנט: www.taubcenter.org.il

הישגים במבחנים בין-לאומיים והקשר שלהם לצמיחה כלכלית

נחום בלס, מיכאל דבאוי ואלכס וינרב

מבוא

מבחני ההישגים הבין-לאומיים, שהחלו כבר בסוף שנות השבעים של המאה העשרים וקיבלו תנופה ומשנה חשיבות במפנה המאה, עם פרסום תוצאות מבחני טימס (TIMSS) 1999 ופיזה (PISA) 2000, שינו במידה רבה את פני הדיון על מדיניות החינוך, הן בישראל והן בכלל המדינות המפותחות. רבים וטובים – בעיקר מחוץ לתחום החינוך – סומכים את ידיהם על קביעות בדבר הקשר בין ציוני פיזה לתוצאות כלכליות עתידיות. הם רואים בתוצאות המבחנים הבין-לאומיים אינדיקטור מהימן לסיכויי הפרט להשתלב בחברה המודרנית ולסיכויי החברה לשפר את מצבה הכלכלי, הן באופן מוחלט והן יחסית למדינות שכל מדינה שואפת להימנות עימן.¹ כך למשל תיאר בנק ישראל את האפקט של הפיגור בהשכלה בישראל לעומת המדינה עם הציון החציוני ב-OECD:

מבחנים בין-לאומיים כדוגמת פיזה מעידים כי [...] מידת הפיגור גורעת כ-0.4 עד 0.6 נקודת אחוז מקצב הצמיחה השנתי של הטווח הארוך וכחמישית עד רבע מרמת הפריון הכולל של הטווח הארוך (בנק ישראל, 2015).

* שמות המחברים לפי א-ב. נחום בלס, חוקר ראשי וראש תוכנית מדיניות החינוך של מרכז טאוב לחקר המדיניות החברתית בישראל; מיכאל דבאוי, עוזר מחקר במרכז טאוב ודוקטורנט במחלקה לכלכלה, אוניברסיטת בן-גוריון בנגב; פרופ' אלכס וינרב, מנהל המחקר במרכז טאוב.

1 יש כמה סיבות נוספות לעלייה בחשיבות המיוחסת למבחנים האלה. הראשונה שבהן היא סוגיית ה"גאווה" שהתעוררה כאשר הן קובעי מדיניות החינוך והן הציבור קיבלו אותם כמבחנים המבטאים נכונה את רמת מערכת החינוך במדינה בהשוואה למדינות אחרות, והשנייה היא הקישור שנעשה בין ההישגים של המדינה במבחנים הבין-לאומיים לביצועיה הכלכליים. שתי סיבות נוספות אינן קשורות כלל לענייני חינוך, אלא לתרבות המשילות שהתעצמה במהלך שנות התשעים של המאה הקודמת: (1) קובעי מדיניות דרשו מדדים כמותניים, בייחוד מדדים שיאפשרו מעקב והשוואה (לסקירה ראו Diaz-Bone, 2019); (2) כפי שמתואר בספרות על תיאוריית World Polity, מדינות נתונות בלחץ להראות ולהוכיח את נאורותן על ידי השתתפות בכל מיזם של הקהילה הבין-לאומית, בלי קשר לאיכותו או לרלוונטיות שלו מבחינתן. הלחץ מופעל בדרך כלל באמצעות רשתות על-לאומיות של מומחים וחברי ארגונים לא-ממשלתיים, ולכן הוא מורגש ביתר שאת במדינות מפותחות (ראו למשל Beckfield, 2010).

עם השנים הולך וגדל מספר המדינות המשתתפות במבחני פיזה – 79 מדינות השתתפו במבחנים בשנת 2019, לעומת 28 שהשתתפו בהם בשנת 2000 – והשפעתם המתרחבת על מדיניות החינוך בעשרים השנים האחרונות מעוררת ביקורת רחבה: על ההתמקדות החד-ממדית במספר מוגבל של כישורים לימודיים וקוגניטיביים; על בעיות והטיות בבניית המדגמים והמדדים; ועל המנגנונים שאפשרו את מעורבותם הרבה של המבחנים בתהליכי קביעת המדיניות במדינות רבות למרות פגמים מהותיים ומתודולוגיים כה יסודיים (Araújo et al., 2020; Fulge et al., 2016; Zhao, 2020). גם חוקרים ישראלים הוסיפו את קולם לביקורת על עצם הקישור בין ההישגים במבחנים להתפתחות כלכלית עתידית (בלס, 2016; Feniger & Atia, 2019). עם זאת, חלקם מייחסים את הישגיה הנמוכים יחסית של ישראל למציאות החברתית-כלכלית במדינה (יוגב ואחרים, 2009; פניגר ושביט, 2011). המטרה העיקרית של מאמר זה היא להתמקד בתכונה אחת המיוחסת למבחני פיזה – כושרם לנבא שגשוג כלכלי ברמה גבוהה של אמינות ודיוק – המחזקת להערכתנו את דעתם של קובעי מדיניות בישראל באשר לערכם הרב לכלכלת המדינה. במחקר קודם ערך בלס בדיקה ראשונית של השערה זו. הוא הסתמך על בחינת הביצועים הכלכליים של המדינות שהשתתפו במבחן SIMS ב-1985 וסיכם: "הקשר בין ההישגים במבחנים הבין-לאומיים לקצב הצמיחה של הכלכלה, ולמשתנים אחרים המצביעים על איכות החיים בחברה, הוא במקרה הטוב חלש, ולרוב הולך ונחלש ככל שמתרחקים בזמן ממועד המבחן" (בלס, 2016, עמ' 45-46).

במחקר זה הרחבנו את הבדיקה הקודמת שנעשתה בארבעה היבטים. ראשית לכול הרחבנו את היקף הנתונים. בבדיקה הקודמת התייחסנו למבחן SIMS 1985 ולהתפתחויות במשתנים כלכליים וחברתיים שונים עד 2010, ואילו כאן נתייחס למבחני פיזה 2000 ו-2003, וכן למבחני טימס 1999 ו-2003, שכללו מדינות רבות יותר וטווח גילים רחב יותר (מבחני פיזה בודקים את הישגיהם של בני ובנות 15, כלומר בעיקר תלמידי כיתות י', ואילו מבחני טימס בוחנים את הישגיהם של תלמידי ותלמידות כיתות ח', שהם בני 13-14). שנית, הרחבנו את תקופת המיקוד. רבים מהמחקרים שנעשו (ראו בין השאר (Hanushek, 2021; Hanushek & Woessmann, 2008, 2010, 2012, 2015, 2021) התמקדו במבחנים ובצמיחה כלכלית בתקופה שמ-1960 ועד 2000. באותן שנים חוו מדינות רבות תמורות דמוגרפיות וכלכליות, כמתואר בהמשך. ייתכן שהקשר בין הישגי תלמידים ממדינות שונות במבחנים הבין-לאומיים ובין השינוי בתמ"ג לנפש רלוונטי עוד לאותן שנים, אך הוא איננו רלוונטי לתקופה הנוכחית.

שלישית, הרחבנו את היקף המשתנים המוסברים. מחקרים אחרים שעסקו בנושא השוו את הציונים במבחנים בעיקר – אם כי לא רק – לקצב השינוי בתמ"ג לנפש לאורך שנים. אנו נתייחס גם לקשרים שבין הישגי התלמידים מהמדינות השונות במבחנים הבין-לאומיים ובין השינויים בתוחלת החיים ובהשקעה במחקר ופיתוח בשנים 2010-2020.

ההרחבה הרביעית במחקר שלנו היא **מתודולוגית**. מחקריהם של הנושק ו-ווסמן קשרו בין כלל המבחנים שנערכו בשנים 1960–2000 ובין ההתפתחויות הכלכליות באותן שנים. כדי להתמודד עם סוגיית הסיבתיות ההפוכה אנו נתייחס כאן לקשרים שבין התפתחות קצב השינוי בתמ"ג לנפש (וכן בתוחלת החיים ובהשקעה במחקר ופיתוח) בשנים 2010–2019 ובין תוצאות מבחני פיזה וטימס מהשנים 1999–2003 בהשוואה להתפתחות מדדים אלו במהלך שני העשורים שקדמו לכך (1980–1999, ראו דיון בהמשך) או בפיקוח על מדדים אלו והתפתחותם.

ממצאינו מעמידים בספק את הטענה בדבר הקשר החזק והבלתי ניתן לערעור שבין ההצלחה במבחני פיזה בהווה ובין גשוג כלכלי בעתיד, לפחות במדינות הנחשבות למדינות מפותחות (שישראל היא אחת מהן). לאחר בחינה של עשרות רבות של מודלים שונים מצאנו שהציונים של מדינות שונות במבחנים שנערכו בשנים 1999–2003 כמעט אינם תורמים לניבוי קצב הצמיחה של התמ"ג, שיעור ההשקעה במחקר ופיתוח, שיעור השינוי בהשקעה זו ותוחלת החיים באותן שנים. כל מתאם שנמצא בין התפתחות כלכלית לציונים היה נמוך יותר ככל שפרק הזמן בין המצב שנבדק ובין מועד המבחנים היה ארוך יותר. גם כאשר נמצא קשר חיובי כלשהו בין ההישגים במבחנים ובין תוצאות כלכליות לאחר מכן, הקשר בין נתונים כלכליים בעבר ובין הישגים במבחנים שנערכו לאחר מכן היה חזק ומובהק יותר במידה משמעותית. במילים אחרות, העבר וההווה הכלכליים של מדינה מנבאים את הישגי תלמידיה במבחני פיזה טוב הרבה יותר מכפי שההישגים במבחנים מנבאים את עתידה הכלכלי של אותה מדינה.

נשוב ונדגיש כי מטרתה של עבודה זו אינה לבחון את השאלה אם הישגים לימודיים במתמטיקה, במדעים ובאוריינות יכולים להשפיע על קצב השינוי בתמ"ג, על אי-השוויון בחברה או איכות החיים במדינה. אין חולק על כך שהון אנושי – ואחת היא איך נגדיר אותו ואיך נמדוד אותו – הוא גורם בעל השפעה רבה על תהליכים חברתיים וכלכליים. אין גם חולק על כך שלידע שנרכש במוסדות חינוך רשמיים ולא-רשמיים יש חלק חשוב בהון האנושי. לבסוף, אין חולק אפילו על כך שמבחנים יכולים למדוד את רמתו של הידע הזה. עם זאת, יש להביא בחשבון שהון אנושי הוא רק אחד מן הגורמים המשפיעים על תהליכים חברתיים וכלכליים (Angrist et al., 2021; Woodberry, 2012), שהון אנושי מורכב לא רק מידע אלא גם מניסיון, ממאפיינים רגשיים ולא-אקדמיים של כוח האדם וכדומה, ושמבחנים אמנם יכולים למדוד ידע אך לא את איכות החינוך במדינה. לכן יכולתם של מבחנים בין-לאומיים לנבא התפתחויות עתידיות ברמת הפרט, ובוודאי ברמה הלאומית, מוגבלת ביותר.

מטרתה העיקרית של העבודה היא לבחון את הטענה שרמת ההישגים של תלמידים במדינה מסוימת במבחנים הבין-לאומיים קובעת את מסלול ההתקדמות הכלכלית של אותה מדינה. נוסף על כך אנחנו מבקשים לבחון בעבודה זו את הביסוס העובדתי של

אמירות כגון "עלייה של כך וכך אחוזים בשיעור התלמידים המשיגים רמת ציון מסוימת תביא לידי עלייה מקבילה של כך וכך אחוזים בקצב הגידול השנתי של התמ"ג".²

רקע

עד ראשית המאה העשרים ואחת וכניסתם של מבחני פיזה וטימס לעולמנו החינוכי, ממוצע שנות הלימוד של האוכלוסייה או שיעור המבוגרים בעלי תעודת בגרות או תואר ראשון היו המדדים המקובלים ביותר להערכת רמת החינוך במדינה. מכפלת המדדים האלה במספר אזרחי המדינה ביטאה את ההון האנושי של אותה מדינה.³ גישה זו, שמשלה בכיפה שנים רבות, באה לידי ביטוי אולטימטיבי בעבודתם של בארו ולי (Barro & Lee, 1996), שיצרו מאגר נתונים בין-לאומי של ממוצע שנות הלימוד של אוכלוסיותיהן של מדינות רבות לפי משתנים שונים. מזמן לזמן הם עדכנו את המאגר הזה (Barro & Lee, 2013), והוא שימש ועודנו משמש חוקרים רבים לאמידת רמת החינוך של המדינות המיוצגות בו.

הביקורת העיקרית כלפי מדדים אלו היא שמספר שנות הלימוד הממוצע או שיעור בעלי התואר הראשון במדינות בעלות מערכות חינוך מפותחות אינו שווה ערך לאותו מספר שנות לימוד או שיעור בעלי התואר הראשון במדינות שמערכות החינוך שלהם מפותחות פחות. ביקורת נוספת היא שגישה זו מתעלמת מתרומות אחרות של מערכת החינוך להון האנושי, שאינן קשורות בהכרח לידיע אקדמי ולמספר שנות הלימוד הפורמליות.⁴

לאחר פרסום עבודתם של בארו ולי פרסמו הנושק וקימקו מאמר מפתח (Hanushek & Kimko, 2000) שבו קבעו, בהסתמך על בחינת הישגיהם של תלמידים מעשרות מדינות במבחנים בין-לאומיים שונים, כי הידע שהצטבר אצל התלמידים ובא לידי ביטוי בציוניהם משפיע על התוצאות הכלכליות של מדינותיהם הרבה יותר ממספר שנות הלימוד. הנושק, בשיתוף פעולה פורה ורב-שנים עם ווסמן, המשיך לבסס ולהעמיק את הטענה הזאת. כדי להתגבר על סוגיית המקריות שבבחירת מבחן מסוים לקביעת הקשר שבין המבחנים ובין התפתחויות כלכליות עתידיות, הם חישבו את ממוצע הציונים של תלמידי כל מדינה בכל המבחנים הבין-לאומיים שהשתתפה בהם, וכיילו אותו לעומת נתוני NAEP (התוכנית הלאומית של ארה"ב להערכת ההתקדמות החינוכית). ההסתמכות על נתוני ארצות הברית

2 הנושק אף פרסם תחזיות רבות לשיעור הגידול בתמ"ג בעקבות עלייה בציונים. לדוגמה, בפרסום שלו עם ווסמן מ-2010 הם העריכו שבהיעדר שינויים אחרים, אם צרפת תעלה את הישגיה במבחני פיזה ב-25 נקודות היא תגדיל את התמ"ג שלה בתוך 80 שנה ב-6 טריליון דולר (Hanushek & Woessmann, 2010).

3 אמנם מבחנים בין-לאומיים נערכו כבר במחצית השנייה של המאה העשרים (מבחן FIMS בשנות השישים ומבחן SIMS בשנות השמונים), אך הם זכו לתשומת לב משמעותית רק עם מבחני טימס ב-1999 ופיזה בשנת 2000.

4 החוקר המוביל בתחום זה הוא ג'יימס הקמן (Heckman), שפרסם עבודות רבות בתחום.

מוסברת בהיותה המדינה היחידה שגם השתתפה בכל המבחנים וגם ערכה מבחנים ארציים מכוילים לאורך כל התקופה. הציון הממוצע שמתקבל באמצעות חישוב זה מייצג, להערכתם, את איכותה של מערכת החינוך. לחיזוק עמדתם זו הם מצביעים על היציבות בתוצאות המבחנים.

לאחר כיוול הציונים אמדו החוקרים את הקשר שבין צמיחת כלכלתן של מדינות ובין ציוני תלמידיהן לאורך התקופה (במחקרים מאוחרים יותר הם אמדו את הקשר בין ציונים לאורך התקופה **בהינתן ציונים בתחילת התקופה** לבין צמיחה לאורך התקופה **בהינתן עושר התחלתי**. ראו Hanushek & Woessmann, 2010). בסופו של דבר הם הגיעו למסקנות מפורטות בדבר הקשר החד-משמעי, לדבריהם, שבין רמת הציונים במבחנים הבין-לאומיים ובין סיכוייהן של מדינות להגדיל את התמ"ג לנפש. הנושק, ובשנים האחרונות עם עמיתו ווסמן, הגיע לנוסחאות שקשרו את רמת ההישגים במבחנים הבין-לאומיים לקצב הצמיחה העתידי. את הביטוי העדכני ביותר של עמדתם הם ניסחו כך:

ניתוח אמפירי של הבדלים בשיעורי הצמיחה מוליך למסקנה פשוטה: צמיחת התמ"ג של מדינה בטווח הארוך נקבעת **במידה רבה** על ידי המיומנויות של אוכלוסייתה. יתר על כן, את המיומנויות הרלוונטיות ניתן לאמוד בקלות בעזרת מבחנים מתוקננים של הישגים קוגניטיביים. **שלושה רבעים** מן השונות בגידול שחל בתמ"ג לנפש במדינות רבות במשך התקופה שמ-1960 עד 2000 אפשר לייחס למדדים בין-לאומיים של מיומנות במתמטיקה ובמדעים. **הקשר בין המיומנויות הקוגניטיביות המצטברות, המכונות ההון האנושי של מדינה, לבין קצב הצמיחה שלה בטווח הארוך הוא קשר חזק ביותר.**

[...]

אם מבססים את תחזית התמ"ג העתידי על יחסי הצמיחה בעבר, התוצאות מעידות כי **מאמצים צנועים להביא את כל התלמידים לרמה מינימלית יניבו רווחים כלכליים עצומים**. לשיפור איכותם של בתי ספר יש יתרונות ניכרים לטווח הארוך. המחקר הקיים אינו מבהיר מהי הדרך הטובה ביותר לשיפור איכותם של בתי הספר, אך כמה מדינות מפותחות ומתפתחות הראו כי השיפור הזה אפשרי (Hanushek & Woessmann, 2021, Summary, p. 1). ההדגשות שלנו).

במילים אחרות, מבחני הערכה בין-לאומיים הם מדד תקף ומהימן למגוון הידע והכישורים שמונחים ביסוד ההון האנושי, והם מסבירים כ-75% מהבדלי הצמיחה בין מדינות במחצית השנייה של המאה העשרים. באותו מאמר אמדו החוקרים כי סטיית תקן אחת במבחנים תואמת הבדל משמעותי של כ-2% בקצב הצמיחה השנתי בשנים 1960–2000 (Hanushek & Woessmann, 2021, p. 11).⁵

5 על פני הדברים, הערכה זו אינה עולה בקנה אחד עם העובדה שקצב הצמיחה הממוצע של התוצר לנפש במדינות ה-OECD ירד מ-4% בשנות השישים של המאה הקודמת לפחות מ-2% מתחילת המאה הנוכחית. כפי שנסביר בהמשך, ממצא זה נובע ככל הנראה מנוכחותן של מספר מדינות חריגות שעברו תהליך של צמיחה מואצת באותה תקופה.

הנושק ו-ווסמן התייחסו בהרחבה לשאלת הסיביות גם בספרם מ-2015 (Hanushek & Woessmann, 2015), והשתמשו במספר שיטות המקובלות במחקר הכלכלי האמפירי בניסיון לבסס את מסקנתם שהישגי התלמידים משפיעים על הצמיחה, בניגוד לטענה שהצמיחה משפיעה על ההישגים או לטענה שגורמים בלתי נצפים משפיעים על ההישגים ועל הצמיחה במקביל. בבדיקה על פי אחת השיטות הם התבססו על מאפיינים של מערכות חינוך במדינות שונות (עצמאות בתי הספר, בחינות אקסטרניות לתלמידים וכדומה) כמשתני עזר, על סמך ההנחה שמאפיינים אלו משפיעים על צמיחה עתידית דרך השפעתם על הישגי התלמידים, אך אינם מתואמים עימה בשום דרך אחרת. בבדיקה על פי שיטה אחרת הם אמדו את הקשר בין מגמת הישגי התלמידים ובין מגמת הצמיחה בכל מדינה, בדרך שמביאה בחשבון גורמים קבועים או תהליכים ארוכי טווח שעלולים ליצור מתאם שווא (תרבות לאומית, מוסדות כלכליים וכדומה). בבדיקה על פי השיטה השלישית זנחו החוקרים השוואות בין מדינות והתמקדו בפרטים: הם בחנו את תוצאות שוק העבודה (תעסוקה ושכר) של מהגרים בארצות הברית והשוו בין מהגרים בעלי רקע דומה שרכשו את השכלתם במדינות שונות, על סמך ההנחה שניתן לגזור מסקנות מאקרו-כלכליות מראיות מיקרו-כלכליות. כפי שהעידו החוקרים בעצמם, זיהוי ישיר של קשר סיבתי בסוגיה הנדונה הוא כמעט בלתי אפשרי, ולכל אחת מהשיטות יש מגבלות משמעותיות משלה. למרות זאת הם איתנים בדעתם שההוכחות הנסיבתיות מאפשרות להסיק שישנו קשר סיבתי בין הישגי התלמידים במדינות השונות במבחנים הבין-לאומיים לבין ההתפתחות הכלכלית העתידית במדינותיהם (Hanushek & Woessmann, 2015, p. 80).

השפעתם של הנושק ו-ווסמן התעצמה לאחר שאנדריאס שליכר, העומד בראש פרויקט המבחנים הבין-לאומיים של ה-OECD (מבחני פיזה), שרבים מכנים אותו היום "שר החינוך של העולם", ואיתו חוקרי הבנק העולמי, אימצו במידה רבה את מסקנותיהם בנוגע לקשר בין תוצאות המבחנים הבין-לאומיים ובין הגידול בתמ"ג ברמה הלאומית.⁶

לאחרונה פיתחו נעם אנגריסט ועמיתיו מדד נוסף המשלב בין מספר שנות לימוד להישגים במבחנים הבין-לאומיים. להערכתם, מדד זה מתאר טוב יותר את הקשר שבין הון אנושי לפיתוח כלכלי ומאפשר ניבוי מדויק יותר של ההתפתחות הכלכלית. הם נוקטים עמדת ביניים בין גישתם של הנושק ו-ווסמן ובין עמדות מתנגדיהם, וטוענים כי ההון האנושי מסביר בין חמישית למחצית מן השונות בצמיחתן הכלכלית של מדינות, בעוד הנושק ו-ווסמן סבורים כאמור שמדובר בקרוב לשלושה רבעים (Angrist et al., 2021).

הממצאים והמסקנות של הנושק ו-ווסמן ואימוצם על ידי אנשי ה-OECD והבנק העולמי משפיעים על המדיניות של מערכות חינוך במדינות רבות בעולם, ובהן גם בישראל. כך למשל התבטא מנכ"ל משרד החינוך דאז, ד"ר שמשון שושני, בריאיון שנערך עימו ב-2012:

6 על החשיבות המיוחדת לעבודתו של הנושק תעיד גם זכייתו לאחרונה בפרס המכונה "פרס הנובל של החינוך".

המעגל ייסגר ואתה תראיין אותי בדצמבר עם פרסום התוצאות של המבחנים הבינלאומיים. המבחנים האלו הם לא הכל, אבל עבורי הם ביטוי של מערכת שמציבה מטרות ויעדים מדידים ומוכיחה שכגוף ענק היא יכולה להתרומם אם נותנים לה את האמצעים. אני חושב שזה ייתן לנו יוקרה בינלאומית (מניב, 2012).

לכך אפשר להוסיף גם את הטענה בדוח בנק ישראל שמבחנים בין-לאומיים מעידים כי איכות ההשכלה בישראל מפגרת אחרי המדינה החצינית ב-OECD – פיגור שפוגע בטווח הארוך הן בקצב הצמיחה השנתי והן ברמת הפריון הכולל במדינה (בנק ישראל, 2015). עמדות אלו של גורמים מרכזיים במערכת החינוך ובמערכת הכלכלית בישראל (וגם במדינות אחרות) מדגישות עד כמה הנתונים העולים מן המבחנים הבין-לאומיים מחייבים התייחסות רצינית ומעמיקה לטענות בדבר הקשר המוכח, לכאורה, בין רמת הישגיה של מדינה במבחנים הבין-לאומיים ובין סיכוייה להצליח בעתיד בזירת התחרות הכלכלית הבין-לאומית. זה מה שאנחנו מבקשים לעשות כאן.⁷

ההתנגדות

כמתואר במבוא, עם השנים גברה ההתנגדות להשפעתם הרבה של המבחנים הבין-לאומיים על קובעי מדיניות. בעת האחרונה החוקרים הבולטים בביקורת המתודולוגית על מסקנותיהם של הנושק ו-ווסמן בדבר הקשר הספציפי שבין ההישגים במבחנים הבין-לאומיים להישגים כלכליים הם קומטסו ורפליי, המבקשים "לקרוא תיגר על הגרסאות הקיצוניות של תורת ההון האנושי שמקדמים ה-OECD והבנק העולמי (אלו המניחות כי כישורים קוגניטיביים ואיכות החינוך כפי שהם נאמדים במבחנים בין-לאומיים הם הגורמים בהם הידיעה לצמיחה כלכלית)" (Komatsu & Rappleye, 2019, p. 17).

השניים פרסמו מחקרים אחדים שערערו על התפיסות הקודמות, ובין היתר הראו כי הקשר בין ציונים לצמיחה אינו עקבי לאורך זמן, וכי שינוי בציונים לרוב אינו מתורגם לשינוי בצמיחה (Komatsu & Rappleye, 2017).

חוקר נוסף, בולט ובעל שיעור קומה אף הוא, המבקר את ממצאיהם של הנושק ו-ווסמן הוא הנרי לוי, כלכלן חינוך ותיק ומוערך. במאמר מ-2012 טען לוי שמסקנותיהם של הנושק ו-ווסמן קיבלו הד דרמטי, שקרוב לוודאי תורם לניפוח ניכר של עוצמת הקשר שבין המבחנים להתפתחויות כלכליות עתידיות. לדעתו יש לכך השלכות שליליות, ובהן בין השאר התמקדות-יתר בציוני מבחנים והפיכתם ליעד מרכזי (ואף בלעדי) של בתי ספר ומערכות חינוך (Levin, 2012).

7 יצוין כי לעבודותיו של הנושק יש השפעה גדולה בתחומים נוספים בכלכלת חינוך, ובהם למשל הקשר בין היקף המשאבים המושקעים בחינוך לבין התוצאות, השפעת גודל הכיתה על ההישגים והחשיבות של איכות המורים. בכל העבודות הללו הוא מבטא עמדה הקובעת שהיקף המשאבים המוקצים לחינוך חשוב פחות מאופן ניצולם.

גם חוקרים ישראלים עסקו בנושא. פניגר ועטיה (Feniger & Atia, 2019) בחנו אם יש קשר בין הישגי המדינות השונות במבחן פיזה 2000 לקצב הגידול בתמ"ג לנפש ולא מצאו קשר כזה. הם גם בחנו אם יש קשר בין שינויים בהישגים במבחן פיזה בשנים 2000–2015 ובין קצב השינוי בתמ"ג לנפש בתקופה זו, וגם כאן לא נמצא קשר כזה.

ב-2016 פיתחו שריד ועמיתיו (Sarid et al., 2016) מודל תיאורטי שקושר בין השקעה בהון האנושי ובין צמיחה כלכלית. המודל שלהם חוזה שיפור ברמת התוצר, אך לא השפעה ארוכת טווח על קצב הצמיחה. זאת ועוד, המחקרים ערכו אמידה אמפירית בהתבסס על נתוני מבחנים בין-לאומיים במתמטיקה ובמדעים, וממצאיה הלמו את המודל והראו שאמנם יש קשר בין ציונים לתמ"ג עתידי, אך אין קשר בין ציונים לצמיחה עתידית.

מחקרנו הנוכחי מחזק את מסקנותיהם של המחקרים המייחסים כוח ניבוי מועט (לכל היותר) לתוצאות המבחנים בכל הקשור להתפתחויות כלכליות עתידיות. הוא מראה שההתאמה בין המגמות (בכל אחד מהמשתנים המוסברים שנבדקו) בתקופה שקדמה למבחנים ובין מגמות ההתפתחות העתידיות גדולה יותר מההתאמה שבין הישגים במבחנים לבין מגמות ההתפתחות שניכרות לאחריהם, והתאמה זו היא שמסבירה את עיקר "כוח הניבוי" של הציונים במבחנים.

מעבר לכך נשאלת השאלה הכפולה בדבר התוקף של מבחן פיזה כמדד של חינוך מצד אחד, ובדבר היכולת לקשר בין הישגים לימודיים במבחן כלשהו בתקופה מסוימת ובין התפתחויות כלכליות עשרות שנים מאוחר יותר, מצד שני. שאלה זו רלוונטית במיוחד לתקופה הנבדקת במחקרם של הנושק ועמיתיו, 1960–2000. די אם נציין שבתקופה זו חוו חלק מהמדינות, בייחוד העשירות שבהן, תמורה דמוגרפית משמעותית (כניסת דור ה"בייבי בום" לשוק העבודה, צניחה בשיעורי הפרייון), תהפוכות בחינוך (עלייה חדה במספר מסיימי התיכון ומקבלי התארים האקדמיים) ותמורות כלכליות (מעבר לכלכלה המתבססת על שירותים יותר מאשר על תעשייה, ובסוף התקופה – תחילת כלכלת ההיי-טק). בתוך כך, מדינות רבות שבתחילת התקופה היו מדינות עניות ששיעורי ההשכלה בהן נמוכים מאוד, הפכו עד תחילת המאה הנוכחית למדינות עשירות ומשכילות יחסית ("הנמרים האסייתיים" – הונג קונג, טיוואן, סינגפור וקוריאה הדרומית, מדינות המפרץ, מדינות דרום אמריקה). הזינוק בציונים של תלמידי כל המדינות הללו במבחנים הבין-לאומיים במהלך 1960–2000 לא יחזור על עצמו. לכן אפשר לטעון שאין בסיס אפיסטמולוגי סביר להסיק מתקופת 1960–2000 לשנים האחרונות או לעתיד, ותחת זאת יש להתמקד בשינויים בהתפתחויות הכלכליות בתחומים השונים בפרקי זמן דומים לפני המבחנים ואחריהם. זו בדיוק הגישה שלנו בעבודה זו, שבה אנחנו משווים בין ההתפתחויות הללו בשנים 1980–1999 ו-2010–2019 לבין תוצאות המבחנים בשנים 2000–2003.

נתונים והגדרות

ציוני מבחני פיזה מהשנים 2000 ו-2003 – לרבות תתי-ציונים בקריאה, מתמטיקה ומדעים – נאספו מה-OECD. ציוני מבחני טימס מהשנים 1999 ו-2003 – לרבות תתי-ציונים במתמטיקה ומדעים – נאספו מה-IEA. ציוני טימס ופיזה המשולבים מהווים ממוצע פשוט של ציוני המדינות בשני המבחנים. הציונים המשולבים במתמטיקה ומדעים הם ממוצע פשוט של שמונת תתי-הציונים מארבעת מבחני פיזה וטימס,⁸ והציון המשולב בקריאה הוא ממוצע פשוט של שני תתי-הציונים ממבחני פיזה. נתוני קצב השינוי בתמ"ג לנפש, השקעה במחקר ופיתוח כאחוז מהתוצר ותוחלת חיים נאספו מהבנק העולמי. עבור כמה מהמדינות בחלק מהתקופות המבוקשות חסרים נתונים, ולכן בכל מודל שנציג הציונים נורמלו בהתאם למדגם.

לוח 1. המדינות המשתתפות במבחנים

המבחן	פיזה 2000 (36)	פיזה 2003 (36)	טימס 1999 (36)	טימס 2003 (42)
המדינות המשתתפות	אוסטריה, אוסטרליה, איטליה, אינדונזיה, ארצות הברית, אירלנד, ארצות הברית, בלגיה, ברזיל, גרמניה, דנמרק, הונג קונג, הונג קונג, הונגריה, יוון, יפן, ישראל, לטביה, לוקסמבורג, ליטא, הממלכה המאוחדת, מקסיקו, נורווגיה, ניו זילנד, סלובקיה, ספרד, סרביה, פולין, פורטוגל, פינלנד, צ'כיה, צרפת, קוריאה הדרומית, קנדה, רוסיה, שוודיה, שווייץ, תאילנד	אוסטריה, אוסטרליה, איטליה, אינדונזיה, ארצות הברית, אירלנד, ארצות הברית, בלגיה, ברזיל, גרמניה, דנמרק, הולנד, הונג קונג, הונגריה, יוון, יפן, ישראל, לוקסמבורג, ליטא, הממלכה המאוחדת, מקסיקו, נורווגיה, ניו זילנד, סלובקיה, ספרד, סרביה, פולין, פורטוגל, פינלנד, צ'כיה, צרפת, קוריאה הדרומית, קנדה, רוסיה, שוודיה, שווייץ	אוסטרליה, איטליה, אינדונזיה, איראן, ארצות הברית, בולגריה, דרום אפריקה, הולנד, הונג קונג, הונגריה, טורקיה, יפן, ירדן, ישראל, לטביה, ליטא, מולדובה, מלזיה, הממלכה המאוחדת, מקדוניה, מרוקו, ניו זילנד, סינגפור, סלובניה, סלובקיה, הפיליפינים, פינלנד, צ'כיה, קוריאה הדרומית, קנדה, קפריסין, רומניה, רוסיה, תאילנד, תוניסיה	אוסטרליה, איטליה, אינדונזיה, איראן, ארצות הברית, בולגריה, דרום אפריקה, הונג קונג, הונגריה, טורקיה, יפן, ירדן, ישראל, לטביה, ליטא, מולדובה, מלזיה, הממלכה המאוחדת, מקדוניה, מרוקו, ניו זילנד, סינגפור, סלובניה, סלובקיה, הפיליפינים, פינלנד, צ'כיה, קוריאה הדרומית, קנדה, קפריסין, רומניה, רוסיה, הרשות הפלסטינית, שוודיה, תוניסיה

מקור: נחום בלס, מיכאל דבאי ואלכס יונר, מרכז טאוב | נתונים: IEA, OECD, הבנק העולמי

8 ככלל, הניקוד במבחנים יוצר התפלגות נורמלית של הציונים, אך לרוב ההתפלגות במבחני טימס היא בעלת שוליים רחבים יותר ממבחני פיזה. על כן ציון משולב עבור מדינה שהשתתפה בשני המבחנים יהיה לרוב קרוב יותר לממוצע מציון הטימס שלה ורחוק יותר מהממוצע מציון הפיזה שלה. בחרנו להשתמש בציונים משולבים משום שככל שמספר המבחנים והמקצועות רב יותר, התמונה המתקבלת בדבר ההשפעה של מערכת החינוך על התפתחויות עתידיות רגישה פחות לבעיות המדידה של מבחן כזה או אחר, ועל כן היא גם אמינה יותר.

לוח 2. נתונים תיאוריים על פי מבחן הערכה בין-לאומי

טימס 2003		טימס 1999		פיזה 2003		פיזה 2000		המבחן
472.5 (65.6)		476.1 (69.9)		485.1 (42.4)		469.4 (55.1)		ציון
2019-2010	1999-1980	2019-2010	1999-1980	2019-2010	1999-1980	2019-2010	1999-1980	תמ"ג לנפש, דולרים
\$22,579 (\$21,723)	\$14,348 (\$15,406)	\$20,214 (\$17,180)	\$12,618 (\$12,194)	\$37,332 (\$21,656)	\$25,066 (\$15,683)	\$35,167 (\$25,281)	\$23,293 (\$17,191)	
2.1% (1.6%)	1.7% (2.6%)	2.2% (1.6%)	1.7% (2.6%)	1.7% (1.4%)	1.9% (1.8%)	1.7% (1.3%)	1.9% (1.8%)	שיעור השינוי בתמ"ג לנפש, באחוזים
76.7 (4.9)	70.1 (5.0)	76.7 (4.7)	70.6 (4.9)	79.4 (3.6)	73.6 (4.3)	79.3 (3.5)	73.5 (4.0)	תוחלת חיים, בשנים
2019-2010	1999-1995	2019-2010	1999-1995	2019-2010	1999-1995	2019-2010	1999-1995	השקעה במחקר ופיתוח, אחוז מהתוצר
1.5% (1.2%)	1.2% (0.9%)	1.4% (1.2%)	1.2% (0.9%)	1.8% (0.9%)	1.4% (0.8%)	1.7% (1.0%)	1.4% (0.9%)	
1.4% (5.2%)	2.2% (8.9%)	1.4% (6.1%)	3.1% (8.8%)	0.9% (5.2%)	4.3% (7.1%)	1.1% (5.5%)	5.5% (8.6%)	שיעור השינוי בהשקעה במחקר ופיתוח, באחוזים

הערה: כל תא מציג את הערך הממוצע (לא משוקלל) של המשטנה לאורך התקופה הנתונה עבור המדינות שהשתתפו במבחן, ומתחתיו סטיית התקן בסוגריים.

מקור: נחום בלס, מיכאל דבאוי ואלכס וינר, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לוח 1 ולוח 2 מציגים נתונים תיאוריים עבור כל מבחן, לרבות רשימת המדינות שהשתתפו בו והעושר והצמיחה הממוצעים שלהן בשתי תקופות שונות. עיון ברשימת המדינות שהשתתפו במבחנים השונים (לוח 1) מלמד שמבחיני פיזה כוללים מספר גדול יותר של מדינות מפותחות. הבדל זה בא לידי ביטוי מובהק יותר בלוח 2, שממנו עולה כי המדינות שהשתתפו במבחני פיזה נהנו מתמ"ג לנפש, תוחלת חיים והשקעה במחקר ופיתוח גבוהים יותר בממוצע מאלו של המדינות שהשתתפו במבחני טימס. המדינות שהשתתפו במבחני פיזה נהנו גם מקצב צמיחה גבוה יותר בתקופה שקדמה למבחנים (1999-1980), אך בתקופה שלאחר המבחנים (2010-2019) קצב הצמיחה בהן היה נמוך יותר. הדבר הפוך לגבי המדינות שהשתתפו במבחני טימס. ככלל, ריבוי המדינות המתפתחות שהשתתפו במבחן טימס רומז כי דווקא במבחן זה סביר יותר שתימצא התאמה בין הישגי התלמידים להתפתחות המדינות, שכן למדינות מתפתחות יש לאן להתקדם – הן מבחינת חינוך והשכלה והן מבחינת פיתוח כלכלי – בניגוד למדינות עשירות, שבהן אופק הצמיחה רחב פחות. עם זאת, כפי שנראה בהמשך, הקשר בין ציוני התלמידים בכל אחד מהמבחנים ובין צמיחתן הכלכלית של המדינות המשתתפות הוא קשר רופף ביותר.

מגמות ארוכות טווח וציונים במבחני הערכה בין-לאומיים

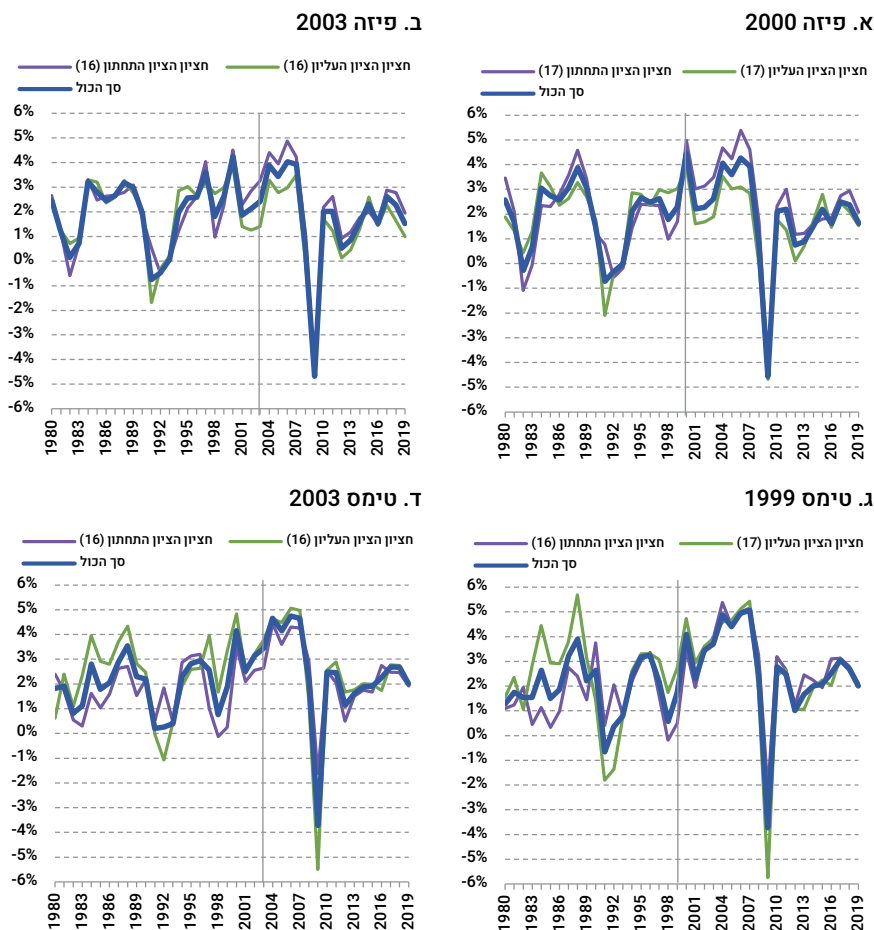
כשלב תיאורי מקדים לניתוח כושר החיזוי הכלכלי של מבחני הערכה בין-לאומיים ראוי לבחון את התפתחותן ארוכת הטווח של מדינות אל מול הישגי תלמידיהן במבחנים שונים. התבוננות זו תתרום לגיבוש אינטואיציה לגבי הקשר התלת-ראשי שבין צמיחה בעבר, הערכת תלמידים בהווה וצמיחה בעתיד.

בתרשים 1 א-ד מוצג קצב הגידול בתמ"ג לנפש של מדינות שונות לאורך התקופה 1980–2019 על פי הישגי תלמידיהן בארבעה מבחנים בין-לאומיים שונים שנערכו בערך באמצע תקופה זו (1999–2003).⁹ כל אחד מארבעת התרשימים מתמקד במבחן אחר. הקווים הסגולים מייצגים ממוצע (לא משוקלל) של קצב הצמיחה במדינות שתלמידיהן השתבצו בחציון התחתון של הציונים בכל מבחן, והקווים הירוקים מייצגים את המדינות שתלמידיהן השתבצו בחציון העליון. לפירוט המדינות שבכל קבוצה ולנתוני קבוצה מצרפיים ראו לוח נ'1 בנספח.

מתרשים 1 א-ד עולה כי אין קשר של ממש בין הציונים ובין מגמות הצמיחה היחסית. זאת ועוד, במבחני פיזה (שני התרשימים העליונים) ניכר כי בעשורים שלאחר המבחן שיעורי הצמיחה יורדים עם העלייה בציוני התלמידים, ובדרך כלל הקווים הירוקים (המדינות בחציון ההישגים העליון) נמוכים מהקווים הסגולים. במבחני טימס, לעומת זאת, המדינות בחציון הציונים העליון נהנו לרוב משיעורי צמיחה גבוהים יותר – למעט בתקופת המשבר של 2008. התאמה זו משקפת ככל הנראה את השפעתו הדיפרנציאלית של המשבר, שבאופן יחסי פגע במדינות מפותחות (שתלמידיהן ציונים גבוהים) יותר משפגע במדינות מתפתחות (שתלמידיהן ציונים נמוכים). בתרשימים ניכרת בבירור הקרבה בין הקווים המציינים את קצב הצמיחה של הקבוצות השונות, ואף נראה שבתקופות מסוימות שיעור השינוי בצמיחת התמ"ג של הקבוצה החלשה מבחינת ציונים גבוה יותר משיעור השינוי בצמיחה של הקבוצה החזקה.

9 תרשים נ'1 בנספח מתאר את גובה התמ"ג לנפש. המגמות שם מצביעות לכאורה על עלייה קלה בתמ"ג לנפש עם העלייה בציונים, אך מגמה זו אינה שונה מן המגמה שלפני עריכת המבחנים.

תרשים 1. שיעור השינוי השנתי בתמ"ג לנפש, 1980–2019, לפי חציון הציון במבחני הערכה בין-לאומיים



הערה: הקו האנכי במרכז התרשים מייצג את השנה שבה נערך המבחן. מספר המדינות בכל קבוצה מופיע בסוגריים במקרא.

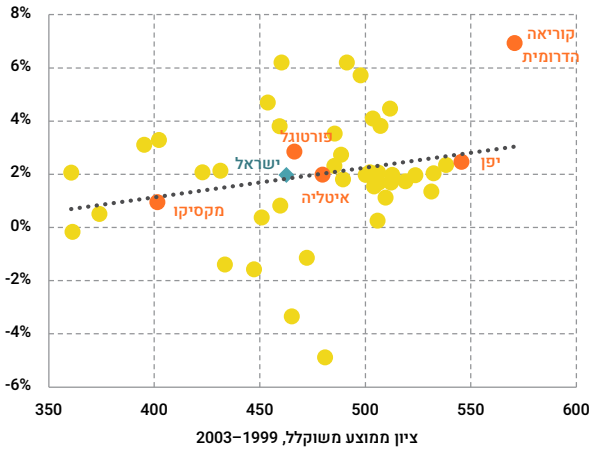
מקור: נחום בלס, מיכאל דבאוי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

מיצוע של העושר וצמיחתו לאורך זמן מבהיר כיצד נראים הקשרים הללו בחתך. תרשים 2 מציג התאמות דו-כיווניות בין הצמיחה הממוצעת בשנים 1980–1999 (בתרשים העליון) ובשנים 2010–2019 (בתרשים התחתון) ובין הציון הממוצע של תלמידי כל מדינה

במבחנים שנערכו בשנים 1999–2003. ניכר כי בעוד ההתאמה בין הצמיחה בעבר ובין הציונים בעתיד חיובית, ככל שישנה התאמה, ההתאמה בין הציונים ובין הצמיחה בשנים 2010–2019 היא שלילית.

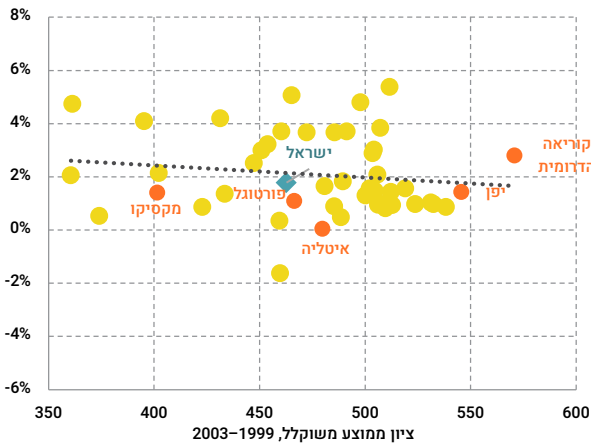
תרשים 2. שיעור השינוי השנתי בתמ"ג לנפש ב-1980–1999 וב-2010–2019 הציונים במבחני 2003–1999

צמיחה שנתית ממוצעת



א. 1999–1980

צמיחה שנתית ממוצעת



ב. 2010–2019

כאמור לעיל, נוסף על בדיקת הקשר בין הישגי המדינות השונות במבחנים הבין-לאומיים ובין קצב הצמיחה של התמ"ג בעבר ובעתיד, בדקנו גם את הקשרים בין הישגי התלמידים ובין תוחלת החיים, ההשקעה במחקר ופיתוח וצמיחתה של השקעה זו – משתנים שלכל אחד מהם יש חשיבות כלכלית וחברתית רבה – לפני המבחנים ואחריהם. לאחר עריכת הבדיקות הסטטיסטיות הדרושות התברר שמלבד בכמה מקרים חריגים אין קשר בין הישגי המדינות במבחנים ובין המשתנים שנבדקו, ואם נמצא קשר שכזה, הרי הוא קשר חלש ולרוב לא-מובהק.

ניתוח אמפירי: כושר החיזוי הכלכלי של ציונים במבחני הערכה בין-לאומיים

מתודולוגיה

כדי להסביר את שיטת הניתוח ניעזר בניסוי מחשבתי: נאמר שבאמצע העשור הראשון של שנות האלפיים ניסו כלכלנים לחזות את עושרן וצמיחתן של מדינות שונות בעשור הבא. עמדו לרשותם נתוני צמיחה ופיתוח כלכלי מהעשורים הקודמים וציונים במבחנים בין-לאומיים שזה עתה התקיימו. עד כמה תרמו נתוני הציונים לדיוק התחזית שלהם? שיטת הניתוח שלנו נועדה לענות על שאלה זו. אמדנו סדרה של מודלים מרובי-משתנים שבהם המשתנה המוסבר הוא ממוצע העושר או הצמיחה של מדינה בשנים 2010–2019, והמשתנים המסבירים הם ממוצע עושרה וצמיחתה בעשור או בעשורים הקודמים וציוני תלמידיה במבחן הערכה בין-לאומי כזה או אחר שנערך בשנים 1999–2003. בהתאם למאפיינים של כל מודל ערכנו מבחן סטטיסטי לבדיקת המובהקות של הציונים כגורם מסביר לצמיחה או עושר עתידיים **בהינתן צמיחה ועושר בעבר**. שיטה זו מכונה "אימות מודל חיזוי" (predictive model validation), ומבחינה רעיונית היא זהה למבחן גריינג'ר¹⁰ (Granger, 1969).

במחקר זה ערכנו מעין שחזור לעבודותיהם המפורסמות של הנושק ו-ווסמן, בכמה הבדלים מתודולוגיים מרכזיים: התמקדנו בתקופה היסטורית עדכנית יותר ובמספר רב יותר של מדדי צמיחה וקידום חברתי-כלכלי; הפרדנו את התקופות שלפני המבחנים ואחריהם; והכללנו באמידה את ערכי העבר של העושר והצמיחה, כך שהשפעתם על הציונים מובאת בחשבון.¹¹ במילים אחרות, שיטתנו בוחנת את השפעת הציונים על הצמיחה לאחר שהבאנו

10 דוגמה לאימות מודל חיזוי ראו אצל Messick, 1995. מבחן גריינג'ר מיושם לרוב במסגרת סדרות עיתיות מתמשכות, ולא בין שתי תקופות בלבד.

11 הנושק ו-ווסמן (Hanushek & Woessmann, 2010) בחנו השפעה של ציונים בהינתן ציונים בתחילת התקופה על צמיחה בהינתן עושר בעבר, אך לא כללו צמיחה בעבר. לכאורה, הקשר בין צמיחה בעבר ובין ציונים עובר דרך ציונים ראשוניים ולכן הם הובאו בחשבון, אך מאחר שיכולה להיות השפעה דיפרנציאלית של הצמיחה בעבר על השינוי בציונים, לדעתנו שיטה זו לוקה בחסר.

בחשבון את נתוני העושר והצמיחה בעבר – שתואמים הן את הציונים והן את הצמיחה העתידית.

לשיטה זו שני חסרונות בולטים: הראשון הוא המדגם המצומצם בשל ההתמקדות במספר קטן של מבחנים, והשני הוא אופק ההשפעה הקצר שאנחנו מאפשרים – התמקדות בהשפעה לאחר 15–20 שנה.¹² כדי להתגבר על החיסרון הראשון אימצנו את הגישה שעליה ממליצה האגודה הסטטיסטית האמריקאית (ASA) והתייחסנו גם לאומדנים לא-מובהקים (מתוך הנחה שכל אחד מהם היה יכול לנבוע ממדגם קטן מדי), ובהמשך הדברים נדון בהם ובמשמעותם (ראו Wasserstein & Lazar, 2016). לחיסרון השני אין לנו מענה אמפירי, אך נדגיש כי האלטרנטיבה – התמקדות בעבר כדי לאפשר אופק השפעה ארוך דיו – בעייתית כשלעצמה, שכן היא מבססת את הדיון על ממצאים שתקפים לעבר, על הקשריו הדמוגרפיים, הכלכליים והחברתיים השונים עד מאוד מהווה, כפי שתואר לעיל.

ממצאים

לוח 3 מציג ממצאים ממספר מודלים ליניאריים שנאמדו בבדיקותינו לצורך המחשה (בהמשך נדון בעשרות מודלים אך נתמקד בהשפעת הציונים בלבד). המשתנה המוסבר הוא שיעור השינוי השנתי הממוצע בתמ"ג לנפש בשנים 2010–2019, והמשתנים המסבירים הם הצמיחה והעושר הממוצעים בתקופה שקדמה למבחנים, היינו 1980–1999, שאותה נכנה להלן "התקופה המוקדמת", וציוני התלמידים במבחן הערכה כזה או אחר שנערך בשנים 1999–2003. ארבע העמודות מימין מתייחסות לתקופה המוקדמת 1980–1989, וארבע העמודות משמאל – ל-1990–1999. בעמודות הימניות לא נלקח שום ציון כמשתנה מסביר, והמדגם שלהן זהה למדגם פיזה, כך שניתן להעריך את תרומתו של ציון זה לניבוי על ידי השוואת הממצאים בין עמודות אלו לעמודות שצמודות להן משמאל. ניתן לראות שהציון עצמו אינו בעל השפעה מובהקת, והוא גם אינו תורם לכושר הניבוי של אף אחד מהמודלים.

12 אתנם בלס (2016) התייחס למבחן SIMS שנערך בשנות השמונים, אך מדובר במבחן שכלל מעט מאוד מדינות.

לוח 3. תוצאות רגרסיה מלאות, מספר מודלים לדוגמה

משתנה מסביר (יחידת שינוי)								משתנה מוסבר: שיעור השינוי השנתי הממוצע בתמ"ג לנפש, 2010-2019 (באחוזים)
(8)	(7)	(6)	(5)	(4)	(3)	(2)	(1)	
תקופה מוקדמת 1999-1990				תקופה מוקדמת 1989-1980				
		0.0554 (0.209)				-0.2398 (0.321)		ציון פיזה 2003-2000 משולב
	0.3731 (0.291)				0.1724 (0.412)			ציון טימס 2003-1999 משולב
0.2777 (0.205)				-0.0003 (0.277)				ציון פיזה וטימס 2003-1999 משולב
-0.04393*** (0.0118)	-0.05681*** (0.0191)	-0.03779*** (0.00993)	-0.03603*** (0.0077)	-0.03944*** (0.0147)	-0.05182** (0.024)	-0.03718*** (0.0202)	-0.04747** (0.0113)	תמ"ג לנפש בתקופה המוקדמת (\$1,000)
0.1262 (0.116)	0.063 (0.099)	0.2418** (0.142)	0.2418** (0.142)	0.1903* (0.11)	0.1322 (0.131)	0.3457** (0.167)	0.3034** (0.185)	שיעור השינוי בתמ"ג לנפש בתקופה המוקדמת (באחוזים)
2.5871*** (0.461)	2.9748*** (0.494)	2.2368*** (0.438)	2.1885*** (0.445)	1.9772*** (0.573)	2.3231*** (0.807)	1.7071** (0.723)	2.0508*** (0.601)	חותך
53	34	39	39	45	30	36	36	מספר תצפיות
0.197	0.246	0.297	0.315	0.138	0.161	0.268	0.279	Adjusted R ²
6.846 (0.001)	3.505 (0.027)	8.31 (0.000)	14.61 (0.000)	6.438 (0.001)	5.615 (0.006)	8.778 (0.000)	10.93 (0.000)	F (p value)

הערה: הלוח מציג ממצאי מודלים האומדים את השפעת הצמיחה, העושר והציונים בעבר על צמיחתן העתידית של מדינות. שש השורות הראשונות מציגות את האומדים להשפעת המשתנים המסבירים על הצמיחה ב-2010-2019. בכל המודלים נעשה שימוש בשגיאות תקן עמידות בפני הטרוסקדסטיות. סטיית התקן בסוגריים.

רמת מובהקות: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

מקור: נחום בלס, מיכאל דבאוי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לוח 4 מציג סיכום מתומצת של כ-60 מודלים ומבחנים סטטיסטיים שאמדו את השפעת הציון – לצד משתנים נוספים – על קצב הצמיחה של מדינות שונות. למען הבהירות מוצגים רק אומדי ההשפעה של הציון על שיעור השינוי השנתי בתמ"ג לנפש. לצורך גיבוש המשתנים הכלכליים המסבירים סיווגנו את המודלים על פי השנים המוגדרות כ"תקופה המוקדמת": השנים 1990–1999 מוצגות בשני הטורים הימניים, 1980–1989 מוצגות בשני הטורים האמצעיים, ו-1980–1999 בשני הטורים השמאליים. השורות מסווגות את המודלים על פי הציון שאת כושר חיזויו בוחן המודל. תוצאות ממודלים נוספים הכוללים תתי-ציונים שונים מהמבחנים או צורות פונקציונליות אחרות למשתנים המסבירים מדווחות בלוח נ'2 בנספח.

המספר המופיע בכל תא מייצג את ההשפעה השולית הממוצעת, באחוזים, משינוי של סטיית תקן אחת (מדגמית) בציון (לרוב כ-50–70 נקודות במבחן; למספרים מדויקים ראו לוח 2). הכוכביות בכל תא מייצגות את רמת המובהקות של הציון כמשתנה מסביר בהתאם למבחן הסטטיסטי הרלוונטי (לפירוט ראו הערה מתחת ללוח). התאים ללא כוכביות מציגים מודלים שבהם הציון אינו משפר את התחזית כלל בהינתן נתונים היסטוריים (המספר שמופיע בו מייצג אפוא הבדל ממוצע כלשהו שאינו מובהק סטטיסטית). גודל המדגם במודל של כל תא מתועד בלוח נ'4 בנספח.

במבט כולל ניכר שכמעט אין אומדים מובהקים סטטיסטית, והציונים היחידים בעלי כושר ניבוי כלשהו הם הציון המשולב במבחן פיזה 2000 וציון הקריאה המשולב (הכולל את פיזה 2000 ופיזה 2003), וגם זאת רק במודל ריבועי ובהיעדר נתוני עושר וצמיחה משנות השמונים. על פני הדברים, ככל הידוע לנו אין ביסוס מחקרי להשמטת שנים אלו או להעדפת המבחנים הללו, ואפילו סביר שהמובהקות הסטטיסטית בשני המודלים האלה מבטאת מתאמים אקראיים (spurious correlations). אם נקבל את האומדנים כפשוטם – ונתעלם משאלת המובהקות הסטטיסטית – נראה כי מרבית האומדנים בלוח חוזים גידול של פחות משליש האחוז בצמיחה בתגובה לשינוי בסטיית תקן אחת בציון, וזהו גם האומדן הממוצע. סטיית התקן של הצמיחה היא 1.3%–1.6% (בהתאם למדגם), כך שההבדלים של סטיית תקן אחת בציונים מסבירים בממוצע עד רבע סטיית תקן בהבדלי הצמיחה בין מדינות, ולרוב אף פחות מזה.

לוח 4. השפעה שולית ממוצעת ומובהקות סטטיסטית של שינוי בסטיית תקן אחת בציון על שיעור השינוי השנתי בתמ"ג לנפש, 2010–2019 (באחוזים)

שיעור השינוי השנתי הממוצע בתמ"ג לנפש בשנים 2010–2019 (באחוזים)						משתנה מוסבר
1999–1980		1989–1980		1999–1990		תקופה מוקדמת
ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	סוג המודל
						ציון מסביר (סטיית תקן)
0.72%	0.04%	0.39%	-0.25%	0.77%*	0.08%	פיזה 2000 משולב
0.57%	0.18%	0.88%	-0.06%	0.55%	0.16%	פיזה 2003 משולב
0.67%	0.01%	0.46%	-0.24%	0.69%	0.06%	פיזה 2003–2000 משולב
0.31%	0.42%	0.00%	0.23%	0.27%	0.44%	טימס 1999 משולב
0.63%	0.52%	0.57%	0.28%	0.59%	0.54%	טימס 2003 משולב
0.39%	0.37%	0.42%	0.17%	0.35%	0.38%	טימס 2003–1999 משולב
0.45%	0.31%	0.43%	0.03%	0.38%	0.32%	מתמטיקה משולב
0.57%	0.00%	0.51%	-0.29%	0.53%*	0.06%	קריאה משולב
0.41%	0.26%	0.34%	0.02%	0.33%	0.28%	מדעים משולב
0.46%	0.27%	0.41%	0.00%	0.37%	0.28%	טימס ופיזה 2003–1999 משולב

תמ"ג ממוצע לנפש בתקופה המוקדמת (דולרים של 2010 בשווי כוח קנייה), שיעור השינוי השנתי הממוצע בתמ"ג לנפש בתקופה המוקדמת (באחוזים)

משתני בקרה

הערה: המספר בכל תא בלוח מייצג את ההשפעה השולית הממוצעת (במונחי סטיות תקן) של הציון במבחן על שיעור השינוי השנתי הממוצע בתמ"ג לנפש בעשור 2010–2019 בהינתן גובה התמ"ג לנפש וצמיחתו בתקופה המוקדמת. הכוכביות מייצגות את רמת המובהקות של משתני הציון במודל (ה-p value של מקדם המשתנה במודל הליניארי וה-p value של מבחן F לשני משתני הציון במודל הריבועי). באמידות נכלל קבוע, ונעשה שימוש בשגיאות תקן עמידות בפני הטרוסקדסטיות.

רמת מובהקות: $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

מקור: נחום בלס, מיכאל דבאוי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לכאורה זוהי השפעה לא מבוטלת, אם כי קטנה משמעותית מהאומדנים של הנושק ו-ווסמן וחופפת בדוחק לקצה התחתון של האומדנים של אנגריסט ועמיתיו (Angrist et al., 2021). אלא שכאן ראוי להזכיר את חוסר העקביות של האומדנים שלנו, שחלקם קטנים משמעותית מהמספרים המוזכרים לעיל. יש אף מקדמים שליליים – ארבעה מעשרת המודלים הליניאריים של 1980–1989 בעמודה השלישית מימין בלוח 4 רומזים שביצועים טובים יותר של תלמידים היו מתואמים עם רמה נמוכה יותר של תוצר לנפש בשנים 2010–2020. לאור כל זה ולנוכח חוסר המובהקות הסטטיסטית שהוזכר לעיל, השורה התחתונה היא שתוצאות המודלים אינן חד-משמעיות או עקביות, והן רגישות עד מאוד להרכב המדינות בכל מדגם (הנובע מהמבחן או מהציון שבהם משתמשים), לתקופה ההיסטורית המובאת בחשבון באמידה ולצורה הפונקציונלית של המודל. כושר הניבוי של

מבחנים לרמת התוצר לנפש הוא במידה רבה תוצאת לוואי (artifact) של מבנה המודל, של הבדלים במדגמים ואולי גם של הטיות בדגימה.

סברה זו מקבלת משנה תוקף בהשמטת מספר מדינות בודדות (עד שלוש לכל רגרסיה) החריגות בצמיחתן או בהישגי תלמידיהן במבחנים (לוח נ'2 בנספח): כמעט כל המודלים שזוהו כמובהקים במדגם המלא מאבדים ממובהקותם הסטטיסטית, מודלים אחרים זוכים למובהקות נמוכה, ומשקלה וכיוונה של ההשפעה החזויה של הציון (בלי קשר למובהקותה) משתנים לחלוטין כמעט בכל מודל. בשיטה זו האומדן הממוצע להשפעתה של סטיית תקן אחת בציון על שיעור השינוי בתמ"ג לנפש הוא כ-0.2%, פחות משישית סטיית תקן של השינוי בתמ"ג. ברוב המוחלט של המודלים הכללת הציון אינה תורמת כלל לכושר החיזוי (ראו לוח נ'5 בנספח).¹³ על כן נראה שהציונים במבחני הערכה בין-לאומיים אינם תורמים כהוא זה לחיזוי הצמיחה העתידית של מדינות.

כאשר מתמקדים במדדים אחרים של פיתוח כלכלי התמונה דומה. לא ניכרת השפעה חיובית מובהקת של ציוני המבחנים על השקעה במחקר ופיתוח, על צמיחת השקעה במחקר ופיתוח או על תוחלת החיים (לוח נ'3 בנספח). האומדנים צנועים בגודלם בלאו הכי, כאשר במרבית המודלים סטיית תקן אחת בציון מתואמת עם שינוי של עד חודשיים וחצי בתוחלת החיים הממוצעת,¹⁴ עם שינוי של עד 0.3% בהשקעה במחקר ופיתוח, ועם שינוי של עד 0.64% בקצב הצמיחה של השקעה במחקר ופיתוח (כלומר 0.300% השקעה תגדל ל-0.302% בשנה שלאחר מכן). כדוגמה לכושר החיזוי הנמוך של הציונים לגבי משתנים אלו נזכיר כי ישראל דורגה בעשור האחרון כאחת המדינות המובילות בשלושת המדדים הללו – על אף ההישגים הבינוניים ומטה של תלמידיה במבחנים בין-לאומיים בתחילת המאה.

13 לוח נ'5א בנספח מציג את ערכי ה-Adjusted R² (מדד מקובל לכושר החיזוי של מודל בכללותו) של מודלים זהים לאלו המוצגים בלוח 2, כאשר הציון כמשתנה מסביר מושמט. לוח נ'5ב מציג את הפער ב-Adjusted R² בין המודלים הכוללים את הציון לאלו שאינם כוללים אותו. במרבית המודלים הציון אינו משפר כלל את כושר הניבוי.

14 בשנים 1970–2019 גדלה תוחלת החיים במדינות עשירות, לרבות בישראל, בכחודשיים עד שלושה חודשים בכל שנה.

דיון וסיכום

בעבודה זו ביקשנו להעמיד לבחינה את הטענה בדבר הקשר החזק והבלתי ניתן לערעור בין הישגיהם של תלמידים במבחני הערכה בין-לאומיים ובין שגשוג כלכלי של המדינה בעתיד. על פי הממצאים שעלו במחקרנו לא ניתן לדחות את השערת האפס שלציון של מדינה במבחן פיזה או טימס אין השפעה ניכרת על צמיחה כלכלית בתקופה שלאחר מכן. במילים אחרות, נראה שציוניהם של בני נוער במבחנים למיניהם אינם טומנים בחובם מידע המאפשר חיזוי של צמיחתן העתידית של מדינותיהם, לפחות לא בטווח של 15–20 שנה. כאמור בפתיח הדברים, לא ביקשנו לערער במחקרנו על חשיבות ההון האנושי כרכיב מרכזי של צמיחה כלכלית או להמעיט בערכו של החינוך כגורם המשפיע על איכות ההון האנושי. גם לא התיימרנו לומר שאין קשר בין הציונים במבחני הערכה בין-לאומיים דוגמת פיזה ובין איכותה של מערכת החינוך, ובוודאי שאין כוונתנו להמעיט בחשיבות תרומתם של המבחנים הבין-לאומיים לחקר החינוך.

עם זאת, לנוכח החשיבות הרבה שמקבלי החלטות וגורמים בעלי משקל בקביעת מדיניות חינוך לטווח הקצר והארוך מייחסים לתוצאות המבחנים הבין-לאומיים, ראינו לעצמנו חובה לבדוק את אמינות הטענה בדבר עוצמת הקשר בין תוצאות המבחנים ובין התפתחות כלכלית בעתיד.

בדיקתנו נבדלה מבדיקות אחרות שעליהן דיווחנו בסקירת הספרות בארבעה היבטים עיקריים: הרחבנו את היקף הנתונים, עדכנו את תקופת המיקוד, הרחבנו את היקף המשתנים המוסברים כך שיכללו גם את השינויים בתוחלת החיים, בהשקעה במחקר ופיתוח ובצמיחתה של השקעה זו, ואימצנו גישה מתודולוגית מתאימה יותר לשאלת המחקר. לאחר בחינה של עשרות רבות של מודלים עלו במחקרנו שלושה ממצאים עיקריים:

1. הציונים של מדינות שונות במבחנים שנערכו בשנים 1999–2003 תורמים 0%–14% להסבר השונות בין מדינות בצמיחת התמ"ג לנפש בשנים 2010–2019, והתרומה הממוצעת מסתכמת בפחות מאחוז אחד – וזאת תוך התעלמות מחוסר המובהקות הסטטיסטית של האומדנים. אילו הגבלנו את עצמנו לאומדנים מובהקים סטטיסטית – בניגוד להמלצות של איגודים סטטיסטיים – לא הייתה התרומה המרבית להסבר השונות עולה על 8%. הגורמים לפער הזה הם הרכב המדינות במדגם של כל מבחן, הגדרת התקופה המוקדמת במודל וצורתה הפונקציונלית של השפעת המשתנים המסבירים.

2. בכל הנוגע להשקעה במחקר ופיתוח הציון במבחנים כמשתנה מסביר משמעותי נמצא בפחות מ-5% מהמודלים, ועבור שינוי בהשקעה במחקר ופיתוח הוא נמצא מובהק בפחות מחמישית מהם (עם השפעה זניחה בגודלה בכולם).

3. אין ראיות לכך שהציונים משפרים את כושר החיזוי של מודלים לניבוי צמיחה כלכלית או שיפור בתוחלת החיים בהינתן נתוני עבר.

התוצאות האלה אינן מפתיעות. המבחנים הבין-לאומיים מוגבלים ביכולתם לנבא התפתחויות עתידיות, הן ברמת הפרט והן ברמה הלאומית, גם משום שהידע אינו אלא רכיב אחד בלבד בהון האנושי וגם משום שיש יכולת המדידה שלהם מוגבלת. לפיכך לדעתנו אין לייחס חשיבות רבה מדי לתוצאות המבחנים ולראותן כאינדיקטורים להתפתחויות עתידיות בתחומי הכלכלה והחברה. אמנם למדידה יש תפקיד חשוב בניהול ובהערכה של מערכות חינוך, אבל ציון בודד במקצוע כלשהו אינו יכול לשקף את איכות מערכת החינוך בישראל, ורמת תוקף התוכן שלו (content validity) אינה גבוהה דייה לצורך השוואה משמעותית למערכות חינוך במדינות אחרות.

העומדים בראש מערכת החינוך בישראל צריכים לפעול לאורך של המטרות ארוכות הטווח כפי שנקבעו בחוק חינוך ממלכתי. ההישגים במבחנים הבין-לאומיים נותנים ביטוי חלקי בלבד למטרות אלו; הם יכולים להצביע על כיווני התפתחות עתידיים, לתרום לידע על היבטים שונים של מערכות חינוך, לסייע לנו להבין כיצד פועלות מערכות חינוך שונות בתנאים שונים וכיצד הן מתמודדות עם בעיות שהן דומות בעיקרן (פערים לימודיים בין מגזרים ומגדרים, קשיים בהבנת הנקרא, יעילות ניהולית וכדומה). ואולם יכולתם של המבחנים להצביע על התפתחויות כלכליות וחברתיות עתידיות מוגבלת ביותר. מסוכן מאוד להפוך שיפור של ציון ומקום בדירוג ליעד מרכזי של מערכת החינוך ולמכשיר של התמודדות פוליטית.

מקורות

- בלס, נ' (2016). **מה ערכם של המבחנים הבין-לאומיים למערכת החינוך?** מרכז טאוב לחקר המדיניות החברתית בישראל.
- בנק ישראל (2015). **הקשר בין איכות ההשכלה לצמיחה: ישראל בהשוואה לעולם**. בנק ישראל.
- יוגב, א', ליבנה, ע', ופניגה, י' (2009). סינגפור במקום ככרוך? מבחני הישג בין-לאומיים והגלובליזציה של יעדי החינוך. **מגמות, מו(3)**, 355–337.
- מניב, ע' (2012, 10 ביוני). **פריחים ושושני: מנכ"ל משרד החינוך מדבר. מקור ראשון**.
- פניגה, י', ושבטי, י' (2011). **המחיר הדמוגרפי: שיעורי הילודה והישגים במבחנים בין-לאומיים**. מרכז טאוב לחקר המדיניות החברתית בישראל.
- Angrist, N., Djankov, S., Goldberg, P.K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, 592, 403–408.
- Araújo, L., Costa, P., & Crato, N. (2020). **Assessment background: What PISA measures and how**. In N. Crato (Ed.), *Improving a country's education: PISA 2018 Results in 10 countries* (pp. 249–263). Springer.
- Barro, R. J., & Lee, J. W. (1996). International measures of schooling years and schooling quality. *The American Economic Review*, 86(2), 218–223.
- Barro, R. J., & Lee, J. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104, 184–198.
- Beckfield, J. (2010). The social structure of the world polity. *American Journal of Sociology* 115(4), 1018–1068.
- Diaz-Bone, R. (2019). Statistical panopticism and its critique (Special issue on governing by numbers — key indicators and the politics of expectations). *Historical Social Research / Historische Sozialforschung*, 44(2), 77–102.
- Feniger, Y., & Atia, M. (2019). Rethinking cause and effect: Analyzing economic growth and PISA scores. In R. Gorur, S. Sellar, & G. Steiner-Khamsi (Eds.), *World yearbook of education 2019: Comparative methodology in the era of big data and global networks* (pp. 96–110). Routledge.
- Fulge, T., Bieber, T., & Martens, K. (2016). Rational intentions and unintended consequences: On the interplay between international and national actors in education policy. In K. Mundy, A. Green, B. Lingard & A. Verger (Eds.), *The handbook of global education policy* (pp. 453–469). John Wiley & Sons.
- Hanushek, E. A. (2021). Addressing cross-national generalizability in educational impact evaluation. *International Journal of Educational Development*, 80, 102318.
- Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, 90, 1184–1208.

- Hanushek, E. A., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46(3), 607–668.
- Hanushek, E. A., & Woessmann, L. (2010). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes*. OECD Publishing.
- Hanushek, E. A., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267–321.
- Hanushek, E. A., & Woessmann, L. (2015). *The knowledge capital of nations: Education and the economics of growth*. MIT Press.
- Hanushek, E. A., & Woessmann, L. (2021). Education and economic growth. In *Oxford research encyclopedia of economics and finance*. Oxford University Press.
- Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3), 424–438.
- Komatsu, H., & Rappleye, J. (2017). A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth. *Comparative Education*, 53(2), 166–191.
- Komatsu, H., & Rappleye, J. (2019). Refuting the OECD-World Bank development narrative: Was East Asia's "Economic Miracle" primarily driven by education quality and cognitive skills? *Globalisation, Societies and Education*, 17(2), 101–116.
- Levin, H. M. (2012). More than just test scores. *Prospects*, 42(3), 269–284.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Sarid, A., Eckstein, Z., & Tamir, Y. (2016). *Do cognitive skills impact growth or levels of GDP per capita?* CEPR Discussion Paper No. 11425.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose. *The American Statistician*, 70(2), 129–133.
- Woodberry, R. D. (2012). *The missionary roots of liberal democracy*. *American Political Science Review*, 106(2), 244–274.
- Zhao, Y. (2020). Two decades of havoc: A synthesis of criticism against PISA. *Journal of Educational Change*, 21(2), 245–266.

נספח

לוח נ'1. ממוצע וסטיית תקן של משתנים כלכליים נבחרים, 1980–2019, על פי רבעוני ביצועים במבחני הערכה בין-לאומיים שונים

מבחן	חציון הציון	מדינות	שיעור השינוי בתמ"ג לנפש (אחוזים)	תמ"ג לנפש (דולרים)	תוחלת חיים (שנים)	השקעה במחקר ופיתוח (אחוז מהתוצר)
פיזה 2000	תחתון	איטליה, אינדונזיה, בולגריה, ברזיל, גרמניה, הונגריה, יוון, ישראל, לטביה, ליטא, מקסיקו, ספרד, פולין, פורטוגל, צ'ילה, רוסיה, תאילנד	2.2% (3.9%)	\$16,251 (\$11,670)	73.5 (5.1)	1.2% (1.0%)
	עליון	אוסטריה, אוסטרליה, איסלנד, אירלנד, ארצות הברית, בלגיה, דנמרק, הונגריה, יפן, נורווגיה, פינלנד, צ'כיה, צרפת, קנדה, שוודיה, שווייץ	1.8% (2.5%)	\$38,840 (\$16,055)	77.4 (3.5)	2.2% (0.8%)
פיזה 2003	תחתון	איטליה, אינדונזיה, ארצות הברית, ברזיל, הונגריה, טורקיה, יוון, לטביה, מקסיקו, נורווגיה, סלובקיה, ספרד, פולין, פורטוגל, רוסיה	2.1% (3.6%)	\$21,557 (\$18,384)	73.1 (5.3)	1.1% (0.7%)
	עליון	אוסטריה, אוסטרליה, איסלנד, אירלנד, בלגיה, גרמניה, דנמרק, הולנד, יפן, נורווגיה, פינלנד, צ'כיה, צרפת, קנדה, שוודיה, שווייץ	1.7% (2.5%)	\$42,172 (\$15,014)	78.3 (3.0)	2.3% (0.7%)
שינוי פיזה (2000/2003)	תחתון	אוסטריה, אוסטרליה, איטליה, אינדונזיה, איסלנד, אירלנד, ארצות הברית, הונגריה, יוון, יפן, מקסיקו, נורווגיה, ספרד, צרפת, קנדה, רוסיה	1.8% (3.0%)	\$31,292 (\$18,459)	76.0 (5.5)	1.7% (0.8%)
	עליון	איסלנד, בלגיה, ברזיל, גרמניה, דנמרק, הונגריה, לטביה, פולין, פורטוגל, פינלנד, צ'כיה, קנדה, שוודיה, שווייץ	1.9% (3.1%)	\$30,535 (\$17,172)	75.9 (4.2)	1.7% (1.0%)
טימס 1999	תחתון	אינדונזיה, איראן, דרום אפריקה, טורקיה, ירדן, ישראל, מולדובה, מלזיה, מקדוניה, מרוקו, הפיליפינים, צ'ילה, קפריסין, רומניה, תוניסיה	2.3% (4.4%)	\$7,745 (\$7,938)	70.6 (5.9)	0.8% (1.0%)
	עליון	אוסטריה, איטליה, ארצות הברית, בולגריה, הולנד, הונגריה, יפן, לטביה, ליטא, סלובניה, סלובקיה, פינלנד, קוריאה הדרומית, קנדה, רוסיה, שווייץ	2.5% (3.7%)	\$25,125 (\$15,709)	75.4 (4.4)	1.8% (1.0%)

לוח נ'1 (המשך). ממוצע וסטיית תקן של משתנים כלכליים נבחרים, 1980–2019, על פי רבעוני ביצועים במבחני הערכה בין-לאומיים שונים

מבחן	חציון הציון	מדינות	שיעור השינוי בתמ"ג לנפש (אחוזים)	תמ"ג לנפש (דולרים)	תוחלת חיים (שנים)	השקעה במחקר ופיתוח (אחוז מהתוצר)
טימס 2003	תחתון	איטליה, אינדונזיה, איראן, בולגריה, דרום אפריקה, ירדן, ישראל, מולדובה, מקדוניה, מרוקו, נורווגיה, הפיליפינים, צ'ילה, קפריסין, רומניה, תוניסיה	2.0% (3.9%)	\$14,603 (\$18,421)	72.3 (6.5)	1.1% (1.2%)
	עליון	אוסטריה, ארצות הברית, הולנד, הונגריה, יפן, לטביה, ליטא, מלזיה, סלובניה, סלובקיה, קוריאה הדרומית, רוסיה, שוודיה	2.4% (3.6%)	\$26,806 (\$15,858)	75.2 (4.8)	1.8% (1.1%)
שינוי טימס (1999/2003)	תחתון	אוסטריה, אינדונזיה, בולגריה, דרום אפריקה, הולנד, הונגריה, יפן, סלובניה, סלובקיה, צ'ילה, קפריסין, רוסיה, תוניסיה	2.1% (3.5%)	\$17,467 (\$15,776)	73.1 (6.3)	1.1% (0.8%)
	עליון	איטליה, איראן, ארצות הברית, יפן, ירדן, ישראל, לטביה, ליטא, מלזיה, מקדוניה, מרוקו, הפיליפינים, קוריאה הדרומית, רומניה	2.5% (4.5%)	\$15,657 (\$15,093)	72.9 (5.5)	1.6% (1.4%)

הערה: כל שורה מייצגת את הערכים הממוצעים של קבוצת המדינות על פי חציון הישגים במבחן הרלוונטי. כל תא מכיל את הערך הממוצע ומתחתיו סטיית התקן בסוגריים. כל קבוצת מדינות מחושבת כממוצע (לא משוקלל) של המדינות בחציון הישגים במבחן הרלוונטי. במבחנים מסוימים יש מדינות שנכללות בשני החציונים. מקור: נחום בלס, מיכאל דבאי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לוח נ'2. מובהקות סטטיסטית והשפעה שולית ממוצעת של שינוי בסטיית תקן אחת בציון על שיעור השינוי בתמ"ג לנפש, ללא חריגים

בנקודות אחוז

שיעור השינוי השנתי הממוצע בתמ"ג לנפש בשנים 2010-2019 (באחוזים)				משתנה מוסבר
1989-1980		1999-1990		תקופה מוקדמת
ריבועי	ליניארי	ריבועי	ליניארי	סוג המודל
				ציון מסביר (סטיית תקן)
				קריאה
-0.06%	-0.25%	0.43%	-0.06%	פיזה 2000
0.23%	0.03%	-0.02%	-0.20%	פיזה 2003
0.13%	-0.19%	0.39%	0.00%	פיזה 2003-2000
				מתמטיקה
0.37%	-0.02%	0.21%	0.13%	טימס 1999
0.31%	-0.10%	0.52%	0.01%	פיזה 2000
0.80%	0.19%	0.58%	0.23%	טימס 2003
-0.14%	0.11%	0.14%	-0.13%	פיזה 2003
0.45%	0.06%	0.53%**	0.16%	מתמטיקה משולב
				מדעים
0.60%	0.06%	0.17%	-0.03%	טימס 1999
0.08%	-0.22%	0.47%*	0.09%	פיזה 2000
0.96%**	0.38%	0.65%*	0.28%	טימס 2003
-0.03%	0.08%	0.06%	-0.17%	פיזה 2003
0.44%	-0.05%	0.52%*	0.13%	מדעים משולב
				ציונים משולבים
0.51%	0.02%	0.17%	0.00%	טימס 1999
0.13%	-0.24%	0.56%*	0.06%	פיזה 2000
0.51%**	0.02%	0.17%	0.00%	טימס 2003
-0.03%	0.08%	0.06%	-0.17%	פיזה 2003
0.69%	-0.07%	0.35%	0.00%	טימס 2003-1999
0.21%	-0.22%	0.49%*	0.04%	פיזה 2003-2000
0.47%**	0.02%**	0.54%**	0.12%**	פיזה/טימס 2003-1999

משתני בקרה תמ"ג ממוצע לנפש בתקופה המוקדמת (דולרים של 2010 בשווי כוח קנייה), שיעור השינוי השנתי הממוצע בתמ"ג לנפש בתקופה המוקדמת (אחוזים)

הערה: המספר בכל תא בלוח מייצג את ההשפעה השולית הממוצעת (במונחי סטיות תקן) של הציון במבחן על שיעור השינוי השנתי הממוצע בתמ"ג לנפש בעשור 2010-2019 בהינתן גובה התמ"ג לנפש וצמיחתו בתקופה המוקדמת. הכוכביות מייצגות את רמת המובהקות של משתני הציון במודל (ה-p value של מקדם המשתנה במודל הליניארי וה-p value של מבחן F לשני משתני הציון במודל הריבועי). באמידות נכלל קבוע, ונעשה שימוש בשגיאות תקן עמידות בפני הטרוסקדסטיות.

מדינות שהיו מרוחקות יותר מ-2 סטיות תקן (מדגמיות) מהממוצע (המדגמי) של המבחן הרלוונטי או של המשתנה המוסבר הושמטו מהמודל.

רמת מובהקות: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

מקור: נחום בלס, מיכל דבאוי ואלכס יורנר, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לוח נ'3. מובהקות סטטיסטית והשפעה שולית ממוצעת של שינוי בסטיית תקן אחת בציון על תוחלת חיים (בשנים), על השקעה במחקר ופיתוח (בנקודות אחוז) ועל השינוי בהשקעה במחקר ופיתוח (באחוזים), ללא חריגים

שנינו בהשקעה במו"פ (אחוזים)		השקעה במו"פ (אחוז מהתוצר)		תוחלת חיים (שנים)				משתנה מוסבר
1999-1995		1999-1995		1989-1980		1999-1990		תקופה מוקדמת
ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	סוג המודל
								ציון מסביר (סטיית תקן)
								קריאה
-0.37%	0.86%	-0.14%	-0.01%	-0.2014	0.1522	0.0949	0.189	פיזה 2000
0.11%	0.68%	-0.07%	-0.04%	0.3552	0.1282	-0.2535	-0.2147	פיזה 2003
0.53%**	1.10%	-0.10%	-0.01%	-0.4016	0.323	-0.2416	0.0524	פיזה 2003-2000
								מתמטיקה
-0.22%	-0.36%	0.13%	0.14%	0.5671	0.3837	0.1133	0.073	טימס 1999
0.45%	1.22%	0.07%	0.11%	-0.3256	-0.0563	0.3056	0.1867	פיזה 2000
0.51%	0.23%	0.15%	0.13%	0.1815	0.7701	0.1359	0.2646	טימס 2003
1.91%	1.65%	0.04%	0.05%	0.4928	0.6276	-0.0471	-0.2497	פיזה 2003
1.16%	**1.36%	0.13%	0.08%	0.571	0.7792	0.2646	0.2193	מתמטיקה משולב
								מדעים
0.41%	0.13%	0.20%	0.16%	0.3457	-0.0007	0.1544	0.3972	טימס 1999
0.55%	1.18%	-0.01%	0.05%	-0.2734	0.0371	0.0358	0.06	פיזה 2000
0.71%	0.15%	0.34%**	0.16%	0.875	0.5275	0.4153	0.2253	טימס 2003
1.24%	1.41%	-0.04%	0.00%	0.5355	0.6111	-0.1223	-0.2575	פיזה 2003
0.85%*	1.08%*	0.07%	0.09%	0.5307	0.6532	0.2842	0.1917	מדעים משולב

לוח נ'3 (המשך). מובהקות סטטיסטית והשפעה שולית ממוצעת של שינוי בסטיית תקן אחת בציון על תוחלת חיים (בשנים), על השקעה במחקר ופיתוח (בנקודות אחוז) ועל השינוי בהשקעה במחקר ופיתוח (באחוזים), ללא חריגים

שינוי בהשקעה במו"פ (אחוזים)		השקעה במו"פ (אחוז מהתוצר)		תוחלת חיים (שנים)				משתנה מוסבר
1999-1995		1999-1995		1989-1980		1999-1990		תקופה מוקדמת
ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	סוג המודל
0.09%	-0.06%	0.17%	0.16%	0.5505	0.2084	0.2536	0.3901	ציונים משולבים
0.23%	1.19%	-0.04%	0.06%	-0.3123	0.0517	0.0584	0.08	טימס 1999
0.09%	-0.06%	0.17%	0.16%	0.5505	0.2084	0.2536	0.3901	פיזה 2000
1.21%	1.39%	-0.04%	0.00%	0.5318	0.6127	-0.1215	-0.2565	טימס 2003
0.14%	-0.06%	0.21%	0.18%*	0.6637	0.5712	0.1765	0.1428	פיזה 2003
1.01%*	1.60%*	-0.08%	0.03%	-0.2341	0.498	-0.052	0.1397	טימס 1999-2003
0.74%*	1.14%*	0.09%	0.11%	0.6199	0.8057	0.3088	0.2216	פיזה/טימס 2003-1999
תמ"ג ממוצע לנפש בתקופה המוקדמת (דולרים של 2010 בשווי כוח קנייה), שינוי שנתי ממוצע בהשקעה במחקר ופיתוח בתקופה המוקדמת (אחוזים)		תמ"ג ממוצע לנפש בתקופה המוקדמת (דולרים של 2010 בשווי כוח קנייה), השקעה ממוצעת במחקר ופיתוח בתקופה המוקדמת (אחוז מהתוצר)		תמ"ג ממוצע לנפש בתקופה המוקדמת (דולרים של 2010 בשווי כוח קנייה), תוחלת חיים ממוצעת בתקופה המוקדמת (שנים)				משתני בקרה

הערה: המספר בכל תא בלוח מייצג את ההשפעה השולית הממוצעת (במונחי סטיית תקן) של הציון במבחן על המשתנה שמופיע בשורה העליונה בעשור 2010-2019, בהיתן גובה התמ"ג לנפש והמשתנה הרלוונטי בתקופה המוקדמת. הכוכביות מייצגות את רמת המובהקות של משתני הציון במודל (ה-value p של מקדם המשתנה במודל הליניארי וה-value p של מבחן F לשני משתני הציון במודל הריבועי). באמידות נכלל קבוע, ונעשה שימוש בשגיאות תקן עמידות בפני הטרוסקדסטיות.

מדינות שהיו מרוחקות יותר מ-2 סטיות תקן (מדגמיות) מהממוצע (המדגמי) של המבחן הרלוונטי או של המשתנה המוסבר הושמטו מהמודל. רמת מובהקות: *p < 0.10; **p < 0.05; ***p < 0.01.

מקור: נחום בלס, מיכאל דבאוי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לוח נ'4. גודל מדגם בכל סדרת מודלים, על פי תקופה מוקדמת לאיסוף נתונים ומבחן הערכה בין-לאומי

שיעור השינוי בהשקעה במחקר ופיתוח (אחוזים)	השקעה במחקר ופיתוח (מהתמ"ג)	תוחלת חיים בלידה (שנים)			שיעור השינוי בתמ"ג לנפש (אחוזים)			מדד
		1999-1980	1989-1980	1999-1990	1999-1980	1989-1980	1999-1990	
1989-1980	1999-1990	1999-1980	1989-1980	1999-1990	1999-1980	1989-1980	1999-1990	תקופה מוקדמת
								קריאה
								פיזה 2000
26	31	36	32	36	36	31	36	פיזה 2003
26	31	36	32	36	36	31	36	פיזה 2003-2000
26	31	36	32	36	36	31	36	מתמטיקה
								טימס 1999
26	31	36	32	36	36	31	36	פיזה 2000
23	28	30	25	30	30	24	30	טימס 2003
23	28	30	25	30	30	24	30	פיזה 2003
23	28	30	25	30	30	24	30	מתמטיקה משולב
23	28	30	25	30	30	24	30	מדעים
								טימס 1999
29	34	39	34	39	39	33	39	פיזה 2000
22	26	32	23	32	32	22	32	טימס 2003
22	26	32	23	32	32	22	32	פיזה 2003
22	26	32	23	32	32	22	32	מדעים משולב
17	23	29	21	29	29	20	29	ציונים משולבים
								טימס 1999
17	23	29	21	29	29	20	29	פיזה 2000
								טימס 2003
22	28	34	25	34	34	24	34	פיזה 2003
36	44	53	43	53	53	42	53	טימס 2003-1999
31	36	41	34	41	41	33	41	פיזה 2003-2000
36	44	53	43	53	53	42	53	פיזה/טימס 2003-1999
36	44	53	43	53	53	42	53	

הערה: כל תא בלוח מציג את מספר התצפיות בסדרת מודלים על פי המשתנה המוסבר, התקופה המוקדמת שלפיה נמדדו ערכי העבר העצמאיים של אותו משתנה ומבחן ההערכה הבין-לאומי שנכלל במודל. מקור: נחום בלס, מיכאל דבאוי ואלכס וינר, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

לוח נ'5א. Adjusted R² למודלים ללא ציוני מבחנים

שיעור השינוי בתמ"ג לנפש (אחוזים)						משתנה מוסבר
1999-1980		1989-1980		1999-1990		תקופה מוקדמת
ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	סוג המודל
						ציון מסביר
						קריאה
0.455	0.313	0.32	0.235	0.478	0.26	פיזה 2000
0.455	0.313	0.32	0.235	0.478	0.26	פיזה 2003
0.455	0.313	0.32	0.235	0.478	0.26	פיזה 2003-2000
						מתמטיקה
0.455	0.313	0.32	0.235	0.478	0.26	טימס 1999
0.414	0.346	0.379	0.23	0.468	0.364	פיזה 2000
0.414	0.346	0.379	0.23	0.468	0.364	טימס 2003
0.414	0.346	0.379	0.23	0.468	0.364	פיזה 2003
0.414	0.346	0.379	0.23	0.468	0.364	מתמטיקה משולב
						מדעים
0.454	0.35	0.347	0.268	0.457	0.297	טימס 1999
0.226	0.251	0.053	0.142	0.214	0.235	פיזה 2000
0.226	0.251	0.053	0.142	0.214	0.235	טימס 2003
0.226	0.251	0.053	0.142	0.214	0.235	פיזה 2003
0.232	0.241	-0.051	0.082	0.228	0.234	מדעים משולב
						ציונים משולבים
0.232	0.241	-0.051	0.082	0.228	0.234	טימס 1999
0.232	0.241	-0.051	0.082	0.228	0.234	פיזה 2000
0.261	0.259	0.08	0.161	0.252	0.246	טימס 2003
0.284	0.225	0.125	0.139	0.314	0.205	פיזה 2003
0.52	0.408	0.377	0.278	0.515	0.352	טימס 2003-1999
0.28	0.219	0.116	0.138	0.309	0.197	פיזה 2003-2000
0.281	0.219	0.12	0.138	0.31	0.197	פיזה/טימס 2003-1999

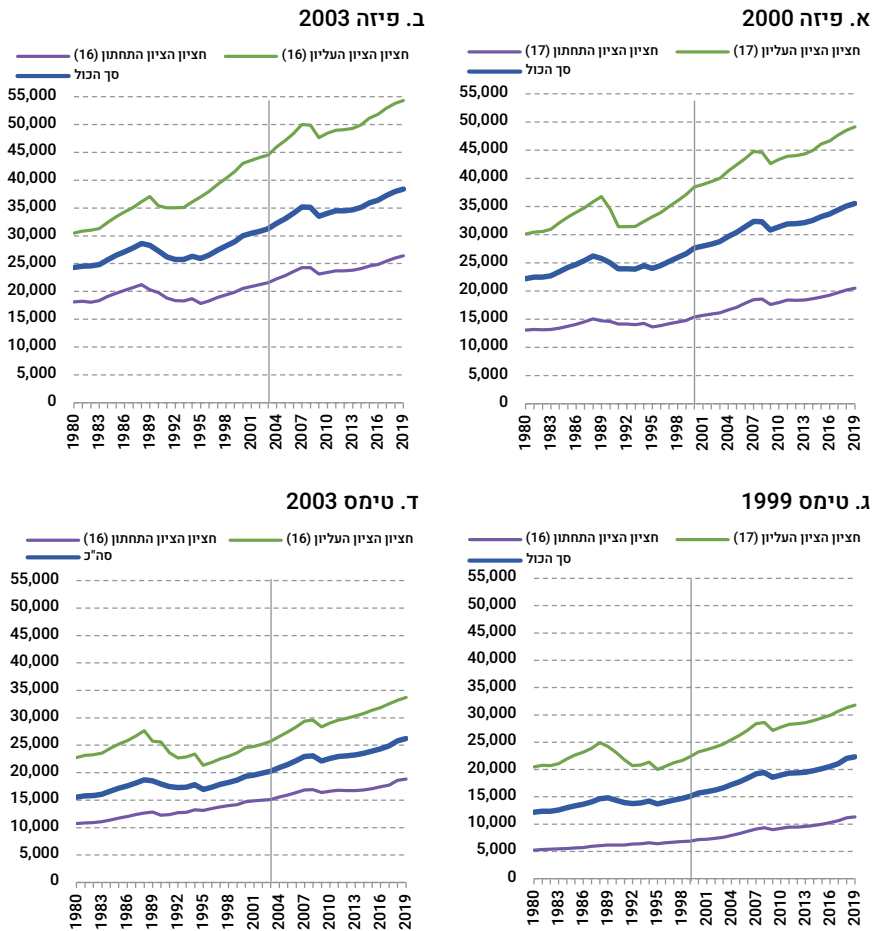
לוח נ'5ב. הפער ב-Adjusted R² בין מודלים עם ציוני מבחנים למודלים ללא ציוני מבחנים

שיעור השינוי בתמ"ג לנפש (אחוזים)						משתנה מוסבר
1999-1980		1989-1980		1999-1990		תקופה מוקדמת
ריבועי	ליניארי	ריבועי	ליניארי	ריבועי	ליניארי	סוג המודל
						ציון מסביר
						קריאה
0.070	-0.021	0.079	-0.011	0.081	-0.020	פיזה 2000
0.059	-0.015	0.034	-0.022	0.076	-0.011	פיזה 2003
0.086	-0.018	0.144	0.006	0.083	-0.021	פיזה 2003-2000
						מתמטיקה
0.045	-0.020	0.039	-0.013	0.061	-0.017	טימס 1999
-0.006	-0.014	-0.001	-0.035	-0.002	-0.015	פיזה 2000
-0.001	-0.005	0.006	-0.036	0.008	-0.005	טימס 2003
-0.023	-0.022	-0.020	-0.034	-0.024	-0.023	פיזה 2003
-0.007	-0.015	-0.002	-0.035	-0.002	-0.015	מתמטיקה משולב
						מדעים
0.044	-0.018	0.057	-0.011	0.046	-0.018	טימס 1999
-0.004	0.032	-0.099	-0.034	-0.013	0.038	פיזה 2000
-0.003	0.031	-0.105	-0.038	-0.013	0.040	טימס 2003
-0.005	0.030	-0.089	-0.031	-0.014	0.034	פיזה 2003
0.034	0.061	-0.080	-0.032	0.025	0.072	מדעים משולב
						ציונים משולבים
0.058	0.068	-0.054	-0.026	0.050	0.083	טימס 1999
0.012	0.049	-0.104	-0.037	0.002	0.057	פיזה 2000
0.014	0.023	-0.059	-0.033	0.006	0.028	טימס 2003
0.007	0.012	-0.024	-0.021	-0.003	0.017	פיזה 2003
0.043	-0.015	0.087	-0.001	0.039	-0.016	טימס 2003-1999
0.003	0.006	-0.033	-0.022	-0.008	0.009	פיזה 2003-2000
0.004	0.006	-0.029	-0.022	-0.007	0.009	פיזה/טימס 2003-1999

הערה: התאים בלוח נ'5א מציגים את ה-Adjusted R² של כל מודל כאשר הציונים מושמטים (המודלים זהים עבור תתי-ציונים שונים של אותו מבחן כיוון שההבדל היחיד בין "מבחנים" הוא מדגם המדינות). התאים בלוח נ'5ב מציגים את הפער בין ה-Adjusted R² של המודלים המוצגים בגוף המאמר לאלו המופיעים בלוח נ'5א.

מקור: נחום בלס, מיכאל דבאוי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי

תרשים נ'1. גובה התמ"ג לנפש בשנים 1980–2019, לפי חציון הציון במבחני הערכה בין-לאומיים בדולרים של 2010



הערה: כל תרשים מציג את גובה התוצר לנפש בשנים 1980–2019 בשתי קבוצות המדינות (מעל ומתחת לציון החציוני בכל מבחן) במבחני הערכה בין-לאומיים לבני 14–15 שנערכו בשנים 1999–2003. הקו האנכי במרכז התרשים מייצג את השנה שבה נערך המבחן. מספר המדינות בכל קבוצה מופיע בסוגריים במקרא. מקור: נחום בלס, מיכאל דבאוי ואלכס וינרב, מרכז טאוב | נתונים: OECD, IEA, הבנק העולמי