

הערכת הישגיה של ישראל במבחני פיז"ה (PISA) באמצעות מדידת השינוי לאורך השנים

דוד בורג, יוסי יפה, אור יצקן

תקציר

אחת המתכונות הבולטות ביותר של מבחנים בין לאומיים הוא מבחן פיז"ה (PISA), המשמש להערכת ידע של תלמידים בהבנת הנקרא, במתמטיקה ובמדעים. מבחנים השוואתיים בין לאומיים החלו להיות מונהגים בעולם החל משנות ה-60 של המאה הקודמת, ותוצאותיהם זוכות לעניין ולפופולריות גוברים. המחזור הראשון שבו השתתפה ישראל במבחני פיז"ה הבין לאומיים היה בשנת 2000, ומאז היא יורדת בהדרגה בדירוג העולמי בכל מחזור, עד כי בשנת 2015 היו תלמידיה ממוקמים במקומות ה-37-40 מתוך 70 מדינות בדירוג הבין לאומי, בשלושת המקצועות. במאמר הנוכחי ביקשנו להציע דרך לבחינה השוואתית של ההישגים הבין לאומיים במבחנים אלה, באמצעות נתונים של ישראל לאורך כל מחזורי הפיז"ה, המבוססת על מדדים של שינוי עצמי. בניגוד למגמת הירידה בדירוג העולמי, מדדים אלה מצביעים על שיפור עקבי בהישגי ישראל בכל המקצועות מאז שנת 2000 ועד היום, שמגמתו עדיפה מזו של עמיתותיה האירופיות והאמריקניות שנכללו בבדיקה. מדדי השינוי של ישראל משקפים עלייה של 6-9 נקודות בממוצע בכל מחזור של פיז"ה בכל המקצועות לאורך השנים, שמהווה שיפור כללי של כ-23 נקודות בממוצע בכל מחזור בהשוואה להישגיה בשנת 2000. לפיכך, ממצאי המחקר מאירים באור אחר את הסברה הרווחת על רקע ההישגים הדלים במבחנים הבין לאומיים, כי מצבו של התלמיד הישראלי נמצא במגמת הידרדרות עקבית זה כמה עשורים. הסתירה בין המגמות מוסברת בין היתר על רקע התווספותן של מדינות רבות למבחני פיז"ה במחזורים האחרונים.

מילות מפתח: פיז"ה, PISA, מבחנים בין לאומיים, חינוך, הישגים, מדדים של שינוי

מבוא

מהו מבחן פיז"ה (PISA)?

מקובל להבחין בין שני סוגים עיקריים של הערכת לומדים: הערכה רחבת היקף והערכה כיתתית. הערכות רחבות היקף מעידות על מספר רב של תלמידים, והן אמצעי להשיג מידע בנוגע לקשר שבין ממצאי הישגים שונים לגורמים העשויים להסביר אותם,¹ כמו תוכניות לימוד, פדגוגיות שונות ובאופן כללי – איכותן ותפקודן של מערכות חינוך שונות. הערכת הישגים רחבת היקף מתבצעת בישראל באמצעות שני מנגנונים עיקריים: מערכת של בחינות בגרות ממלכתיות ומערכת של מדדי יעילות וצמיחה בית ספרית (מיצ"ב).² המיצ"ב הוא מערכת מדדים לימודיים-חברתיים-חינוכיים, ומטרתו לספק תמונת מצב עדכנית של מערכת החינוך על רמותיה השונות. נושאי המיצ"ב ביחס להישגים מתמקדים בתחומי שפת האם, מתמטיקה, מדעים ואנגלית. ברמה הבין לאומית מתבצעות הערכות רחבות היקף באמצעות מבחני הישג השוואתיים, דוגמת מבחן פיז"ה.

¹ הוועדה למדידה והערכה בחינוך (2005). הצעה למסגרת לתוכניות לימודים ולפיתוח מקצועי - מדידה והערכה בחינוך. היזמה למחקר יישומי בחינוך.
² שם.

מבחנים השוואתיים בין לאומיים מונהגים בעולם החל משנות ה־60 של המאה הקודמת. מאז שנות ה־90 של אותה המאה התרחב מאוד השימוש במבחנים אלה בתמיכת ארגונים בין לאומיים, כאונסק"ו, הבנק העולמי והארגון לשיתוף פעולה ולפיתוח כלכלי (OECD). תוצאות המבחנים הללו זוכות להדים תקשורתיים וחברתיים רחבים ומושכות תשומת לב ציבורית רבה, ומרבית המדינות המשתתפות בהם אינן מרוצות ממיקומן הבין לאומי, גם כאשר הן ממוקמות בראש טבלת ההישגים (Atkin & Black, 1997; Blass, 2016).

אחת המתכונות הבולטות ביותר של המבחנים ההשוואתיים מאז תחילת שנות ה־2000 היא מבחן פיז"ה (PISA (Programme for International Student Assessment) – המשמש להערכת ידע בהבנת הנקרא, במתמטיקה ובמדעים בקרב שכבת הגיל של בני 15. בהתאמה, המבחן מעריך את היכולת של התלמידים להשתמש בכישורים האורייניים שלהם כדי להבין ולפרש מגוון חומרים כתובים שהם יפגשו בסבירות גבוהה במרוצת חייהם; את מידת יכולתם ליישם את הידע המתמטי שברשותם כדי להתמודד עם בעיות שונות מבוססות מתמטיקה; ואת מידת יכולתם להשתמש בידע המדעי שלהם על מנת להבין ולפרש מגוון מצבים ואתגרים מדעיים שונים (Turner & Adams, 2006).

לפי הבנק העולמי והארגון לשיתוף פעולה ולפיתוח כלכלי,³ מבחן פיז"ה "פותח במקור במטרה לאמוד את היכולת של תלמידים ליישם ידע ומיומנויות במקצועות מפתח, לחשוב בהגיון, לבצע אנליזה ולהעביר זאת באופן יעיל בעת פתרון של בעיות במגוון מצבים". פיז"ה גם מצביע על יכולת למידה של תלמידים לאורך החיים (life-long learning) מתוך דיווח על רמת המוטיבציה, על אמונתם לגבי עצמם ועל אסטרטגיות הלמידה שלהם. לבסוף, פיז"ה מהווה מדד לשלוש יכולות עיקריות אצל סטודנטים צעירים: 1. מוכנות לאתגרים עתידיים; 2. היכולת לנתח, להסביר ולתקשר בצורה אפקטיבית; 3. היכולת להמשך לימודים במהלך החיים (Sadler & Zeidler, 2009).

מבחני פיז"ה נבדלים מבחינת המיקוד הדיסציפלינרי שלהם בכל מחזור: בשנת 2000 הודגש נושא הבנת הנקרא; ב־2003 – מתמטיקה; ב־2006 – נושא המדעים (אוריינות מדעית); ב־2009 – הבנת הנקרא; וב־2012 – שוב מתמטיקה (Blass, 2016). כל מחזור של פיז"ה דוגם לפחות 150 בתי ספר בכל מדינה, ומשתתפים בו לפחות 5,250 תלמידים, הנדגמים מתוכם באופן רנדומלי. פרט למבחני הידע הדיסציפלינריים במקצועות האמורים, התלמידים מתבקשים להשיב גם על שאלון הכולל נתוני רקע דמוגרפיים, פרטים לגבי העדפות לימודיות ומקצועיות, מאפיינים אישיים וכן מאפיינים של סביבת בית הספר והמשפחה שלהם (Turner & Adams, 2006).

אומדני הביצוע במבחני פיז"ה תוקננו כך שממוצע ה־OECD הכללי הוא 500 נקודות, עם סטיית תקן של 100 נקודות. בין השנים 2000 עד 2015 מי שהובילו את המקומות הראשונים ב־OECD, המדינות בעלות ההישגים הגבוהים ביותר במבחנים האלה, היו לסירוגין: פינלנד, קנדה, דרום קוריה, יפן, סין (טאיפיי), סין (שנגחאי), סינגפור, אוסטרליה וניו זילנד (לא בהכרח לפי סדר המקומות), ולעיתים גם מדינות אירופה כמו אסטוניה, הולנד וליכטנשטיין. טווח הציונים במבחני פיז"ה ברמות האלה במהלך השנים נע ברובו סביב 530 עד 555 נקודות בקירוב (בין שליש לחצי סטיית תקן מעל הממוצע), ובשנים מסוימות אף למעלה מזה – לדוגמה

OECD (Organization for Economic Co-Operation and Cultural Development) (2016). PISA 2015 Results in ³ Focus. Retrieved from <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>, עמ' 15.

580 נקודות של סין (שנגחאי) באוריינות מדעית בשנת 2012. להוציא הבדלים פרטניים כאלה ואחרים, באופן כללי נראה שישנה קורלציה בהישגי המדינות המובילות בין שלושת תחומי האוריינות הדיסציפלינריים של פיז"ה.

עשרות מדינות החלו להעריך את יכולות תלמידיהן על פי הישגיהם במבחנים אלה, בהתבסס על ההנחה שדירוג גבוה בהם משקף את היכולת הלאומית העתידית להתחרות בשוק הגלובלי. אי לכך, הפופולריות של הבחינות הללו החלה גואה בעשורים האחרונים, ותוצאותיהן זוכות לסיקור תקשורתי נרחב, המלווה כל העת בניסיון לנתח ולהסביר את ההישגים היחסיים בכל מדינה (Feniger, Livneh, & Yogev, 2012). מרבית ההסברים שהוצעו על ידי החוקרים והמומחים ברחבי העולם ביחס לביצועי התלמידים במבחני הפיז"ה יוחסו במישרין לתפקוד מערכת החינוך ולאיכותה במדינות השונות מבחינת המבנה והמדיניות (Feniger & Lefstein, 2014).

בגרמניה, למשל, הוחלט לערוך רפורמות מקיפות במערכת החינוך הכוללות הקמת מנגנון פיקוח על סטנדרטים חינוכיים, לאחר שתלמידיה דורגו מתחת למוצע ה-OECD במחזור המבחנים של שנת 2000. בישראל בשנות ה-90, כמו בארצות הברית, הונהגה מדיניות של קביעת סטנדרטים ביחס להישגים לימודיים והערכתם במערכת החינוך, שאחד מסימני ההיכר הבולטים שלה היה הנהגת מבחני המיצ"ב (יוגב, ליבנה ופניגר, 2009).

אלא שממצאי מחקרים שפורסמו בשנים האחרונות מציעים דרך אחרת לבחון את השונות הבין לאומית בהישגי התלמידים במבחני פיז"ה ולהבינה. שביט ופניגר (Shavit & Feniger, 2011) הראו, למשל, כי שיעור הילודה היחסי הגבוה בישראל מסביר את רוב הפערים בין הישגי התלמידים בארץ למוצע הבין לאומי של ציוני הפיז"ה בשנים 2000 ו-2006. שיעור ילודה גבוה זה בא לידי ביטוי בהיווצרותן של משפחות גדולות יותר ושל כיתות צפופות יותר, ואלה, בתורן, משליכות על הישגיהם החינוכיים של הילדים.

אולם מעבר להפחתה היחסית בהשקעה הלאומית בתלמיד כשלעצמה, מחקרים מראים כי שיעור ילודה גבוה מתקשר עם ירידה בהתפתחות הקוגניטיבית ובהישגים חינוכיים, בראש ובראשונה דרך גודל המשפחה. זאת היות שבמשפחות גדולות יותר מבחינת מספר האחאים, כל אחד מן הילדים זוכה לנתח קטן יותר מן המשאבים המשפחתיים – ובכללם גם תשומת הלב וההשקעה ההורית בחינוך הילד (Blake, 1989).

זאת ועוד, ניתוח נוסף של ציוני הפיז"ה משנת 2000 (שנערך בפועל בישראל ב-2002) מלמד כי הישגי התלמידים בישראל ניתנים לניבוי כמעט מלא (כ-80% מהשונות המוסברת בציוני התלמידים בין המדינות) באמצעות שני משתנים מקרו-כלכליים (תמ"ג – תוצר מקומי גולמי) ודמוגרפיים (שיעור האוכלוסייה הצעירה) (Feniger et al., 2012). החוקרים הצביעו על כך שמצבן הכלכלי והדמוגרפי של המדינות מכתוב במידה רבה את הצלחתן במבחני הפיז"ה, ולא דווקא ההשקעה בחינוך או איכות מערכת החינוך כשלעצמן. בהקשר זה הסיקו החוקרים כי:

המיקום הנמוך של תלמידי ישראל בין שאר המדינות הנבחנות איננו מבטא הישגים כה נמוכים של מערכת החינוך, המצדיקים תגובה ציבורית של הלם והיסטריה. כלל לא ברור אם ממצאים אלה אמורים להוביל למסקנה שיש ירידה בהישגים הלימודיים בישראל. ייתכן שיש ירידה כזו, וייתכן שאין. מכל מקום, ממצאינו מורים שאין בתוצאות המבחנים הבין-לאומיים עילה מספקת לטענה בדבר ירידת הישגים (יוגב ועמיתיו, 2009, עמ' 349).

באופן דומה, מדדים כלכליים ודמוגרפיים מצליחים לנבא ציונים במבחן פיז"ה בקרב שאר מדינות העולם (Feniger et al., 2012), מה שמעלה סימן שאלה משמעותי בנוגע לתקפותם כמדדים של היכולות הלימודיות של לומדים במערכות החינוך ברחבי העולם, ומכאן גם של איכותן של אותן מערכות חינוך. כך, למשל, דוחות פיז"ה משנת 2000 חשפו קשר הדוק בקרב מדינות העולם בין השכלת האם לבין הישגי התלמידים בשלושת תחומי היסוד הדיסציפלינריים, עם פער של יותר מחצי סטיית תקן ביחס לממוצע הכללי בקרב נבחנים שאימותיהם לא סיימו תיכון (Anderson, Lin, Treagust, Ross, & Yore, 2007).

הערכת הישגים של לומדים באמצעות מדידת השינוי

דרך סטנדרטית להעריך ביצועים מבוססת על השוואה חד ערכית ברמת הביצוע של אינדיבידואלים או של קבוצות שונות, ודירוגם בהתאם לרמת הביצוע. גישה זו להערכת הישגים היא שכיחה ומקובלת לצורכי השוואת הישגים בין אינדיבידואלים שונים בתחומי הספורט, החינוך ועוד. בתחום החינוך גישה זו משמשת מקבלי החלטות כדי למיין ולדרג לומדים וקבוצות לומדים לפי הישגיהם על סולם ביצועים אקווילונטי לכאורה. בקנה מידה רחב היקף (לאומי או בין לאומי), הערכות אלה משמשות לשם השוואה בין הישגי תלמידים רבים במטרה לזהות דפוסי חוזק וחולשה בשיטות ההוראה ובמערכות החינוך, כמו גם לשם הערכה של השגת המטרות החינוכיות.⁴

חרף זאת, דרכי הערכת הישגים המבוססות על השוואה בין לומדים בנקודת זמן נתונה, אינן רגישות דיין לדינמיקת השינוי שחלה אצל כל לומד ביחס למצבו הקודם (למשל, בעת הערכת יעילות של תוכניות לימודים חדשות) ואין בהן בהכרח כדי להעריך מגמות של שיפור. דיוויס (Davis, 1964 בתוך Willett, 1989) כתב בהקשר הזה כי "במובנים מסוימים מדידת השינוי האינדיבידואלי לאורך זמן היא הנושא החשוב ביותר במדידה חינוכית, שכן מטרתה העיקרית של הוראה היא לייצר למידה (כלומר, שינוי), והכמות והאופן של אותו שינוי יכול להימדד רק תוך השוואתם של אינדיבידואל ו/או קבוצה למצבם לפני ואחרי תהליך הלמידה". התוספת או השינוי בהישגים שמספק בית הספר (או מערכת חינוך ברמת המאקרו), אומר קולמן, הוא זה שצריך להוות מדד לאיכותו החינוכית (Coleman, 1975).

מדידה מסורתית של שינוי עוסקת בשתי גישות עיקריות. האחת מתמקדת בשינויים אינדיבידואליים לאורך זמן ומתייחסת לשאלות – האם חל שיפור בביצועים? האם השיפור יציב או הדרגתי? והאם הוא מתיישר בתום תקופה מסוימת? הגישה השנייה עוסקת בשאלת ההבדלים הבין אישיים באופי ובקצב השינוי בין קבוצות, כלומר במידה שקבוצה אחת משתפרת באופן מהיר יותר מהקבוצה השנייה או להפך.

השיטה הראשונה מקדימה בדרך כלל את השיטה השנייה וכרוכה בה, שכן הכרחי לדעת תחילה באיזה אופן משתנה כל פרט, קודם שאפשר לשאול אם דפוס השינוי שלו נבדל באופן סיסטמטי מזה של פרט אחר (Willet, 1994). בארט ואלכסנדר (Barrett & Alexander, 1989) מציעים שלוש הגדרות אופרטיביות למדדים או לקריטריונים של שינוי: 1. שינויים לאורך זמן ברמה הממוצעת של ביצועים אינדיבידואליים או קבוצתיים; 2. שינויים במתאמים של מדדי תוקף

⁴ National research council, (2003). Assessment in support of instruction and learning: Bridging the gap between large-scale and classroom assessment-Workshop report. National Academies Press

לאורך זמן; 3. שינויים לאורך זמן בדירוגים של פרטים בתוך קבוצה. מדידה חינוכית של שינוי אינדיבידואלי עוברת בעשורים האחרונים מאוריינטציה של מדידת שינוי חד פעמי (מדידה בשני גלים: לפני-אחרי) למדידה של צמיחה מתמשכת ורב פעמית (Bock, 1976) (Willet, 1994).

הערכת הישגים בין לאומיים במבחני פיז"ה במונחי שינוי עצמי עיון בנתוני הישגים של מבחני פיז"ה במדינות השונות בדיסציפלינות הנבחרות במרוצת כל השנים מאז היווסדם מלמד על הדרך המסורתית והמקובלת להשוות בין רמתם של תלמידי המדינות השונות ולהשליך מכך על איכות מערכות החינוך בכל מדינה ומדינה. כך, השוואת ציוני הגלם הממוצעים בכל תחום ותחום משמשת את דעת הקהל ואת קובעי המדיניות החינוכית בעולם כדי לדרג את מערכת החינוך המקומית בהשוואה לשאר מדינות ה-OECD ובהשוואה לרמתן הכללית בכל תחום דיסציפלינרי (כלומר, המרחק של כל מדינה מן הממוצע הגלובלי). לצד הדרכים המסורתיות הללו, שבאמצעותן מקובל לבחון את מצבן החינוכי של מדינות שונות, קיימת גם זווית התבוננות מעניינת וחשובה לא פחות להערכת הישגים במבחני פיז"ה – שזכתה לעניין ולתהודה פחותים בקרב המומחים. הדברים אמורים במגמת השינוי – שיפור או הידרדרות – שחלה בהישגיהם של תלמידי כל מדינות העולם לאורך העשור וחצי האחרונים, מעבר להישגים עצמם בכל מחזור של מבחני פיז"ה.

בהקשר הזה, עוד לפני הגעתם של מבחני אלה לעולם, התריע ווילאט (Willet, 1989) על נטייתם של מומחים וחוקרים במדידה חינוכית להעריך את השפעתן של פדגוגיות שונות על תפקודים חינוכיים של לומדים באמצעות השוואה בין נבדקים בנקודת זמן בודדת, על פני בחינתן במונחי צמיחה תוך נבדקית לאורך זמן. מגמות של שינוי תוך קבוצתי לאורך זמן (בתוך כל מדינה ומדינה) בציוני הפיז"ה לא רק שעשויים לספק תחזיות להישגי המדינות במבחנים אלה בשנים הבאות, אלא שיש בהם כשלעצמם גם כדי להוות מדד הערכתי חשוב לאיכות התהליכים החינוכיים וליעילות – הן ברמת המיקרו הבית ספרית הן ברמת המקרו החינוכית-לאומית.

תשומת הלב הזעומה שניתנה לנקודת המבט הזו בהערכת הישגים בין לאומיים הצביעה על מגמה מפתיעה ביחס לישראל, שלפיה תלמידיה מדורגים בין עשר המדינות הגבוהות בעולם במדדים שונים של שיפור הישגים במבחני השוואתיים בין לאומיים, לרבות במבחני פיז"ה (Blass, 2016). בניגוד לדעה הרווחת, השינוי לרעה בדירוגה העולמי במקצועות היסוד של פיז"ה קשור יותר בהתווספות מספר המדינות המשתתפות במבחנים האלו, ולא דווקא בהידרדרותם של מערכת החינוך הישראלית ושל תלמידיה.

לנוכח הקשיים המהותיים הכרוכים בהשוואה חד ערכית בין ציוני תלמידים מלאומים ומתרבויות שונים, כפי שהצביעו עליהם מומחים שונים בעבר (Bonnet, 2002; Goldstien, 2004; Romainville, 2002), מתחזקת החשיבות בהערכת רמתן החינוכית של מדינות ה-OECD על בסיס מדד השינוי התוך קבוצתי בהישגים במבחני פיז"ה לאורך ציר הזמן, ולא רק בהסתמך על רמת הישגים כשלעצמה. מהשוואה חינוכית בין לאומית המתיימרת להיות חד ערכית (במובן של מדידת הבדלים ביצועיים בכשרים זהים), מתחייבת בחינת הבדלים בין המדינות במונחים של שינוי עצמי בהישגים. זאת משום שמדדים של שינוי עצמי תוך לאומי בתוצאות מבחני פיז"ה השונים הם סוג של מדדים מתוקננים, שערכיהם מאפשרים השוואה אקוויוולנטית יותר בין ביצועים של תלמידים בני תרבויות ושפות שונות.

התפלגות ההפרשים בין ציוני התלמידים הישראלים לשאר תלמידי מדינות העולם במקצועות מתמטיקה ומדעים במהלך השנים, הובילה למסקנה כי מצבו של התלמיד הישראלי הורע, והוא עגום בתחומים אלו, בשל הירידה לכאורה שחלה בעשורים האחרונים ברמה החינוכית בישראל (ראו למשל: בן דוד, 2003; Ben-David, 2010). אלא שהירידה האמורה פה מתייחסת לירידה של ישראל בדירוג העולמי בהישגים הגולמיים במבחנים הבין לאומיים, וזו אינה משקפת בהכרח ירידה של ממש ביכולות של ישראל בהשוואה לביצועיה הקודמים במבחנים אלה. בקו אחד עם עמדתם של יוגב ועמיתיו (2009) בדבר היעדר הרעה ברמת התלמיד הישראלי, נציע במאמר הנוכחי דרך אחרת לבחינת רמתו של האחרון בהשוואה לשאר תלמידי העולם – וזאת על רקע השינוי העצמי בתוצאות מבחני פיז"ה לאורך השנים, החל משנת 2000 ועד היום. בעוד שפיז"ה מספק ניתוח של מגמת השינוי ההישגי העוקב בין סיבוב אחד לבא אחריו (כלומר, דיווח על השינוי ביחס למחזור האחרון), במאמר זה תנותח מגמת השינוי בציוני פיז"ה על פני כל המחזורים, וכתוצאה מכך יסופק ציון כמותי של השינוי הממוצע, הניתן להשוואה בין מדינות לאורך זמן. בהינתן תשומת הלב הנמוכה יחסית הניתנת להיבט זה של הישגי המדינות במבחנים הבין לאומיים, המחקר הנוכחי שואף לספק ניתוח השוואתי משלים של השינוי העצמי בתוצאות פיז"ה בין מדינות שונות לאורך כל השנים. באמצעות חקירת המקרה של ישראל, זווית ההתבוננות שתוצע פה עשויה להניב תובנות נוספות וחדשות ביחס למצבן החינוכי של מדינות וקבוצות לאומיות שונות בראייה בין לאומית.

מטרת המחקר

מטרת המחקר הינה להראות כי משנת 2000 ועד היום, לא זו בלבד שרמתם של תלמידי ישראל במקצועות היסוד לא ירדה (אף על פי שמיקומם בדירוג העולמי של ה-OECD ירד) – אלא שהיא אף השתפרה באופן עקבי ומשמעותי בהשוואה להישגיהם הקודמים. בהמשך לכך, את הירידה בדירוג הגולמי של ישראל ביחס לשאר מדינות העולם ננסה להבין על רקע הצטרפותן של מדינות נוספות לרשימת המשתתפות במבחנים במהלך השנים. בעבודה הנוכחית נשאף ליישם את האוריינטציות למדידת השינוי העצמי שנידונו לעיל, במטרה להעריך באמצעותן את מגמת השיפור האינדיבידואלי של ישראל במבחני פיז"ה ולהשוותה לזו של כמה מדינות נבחרות. בחינה של הישגי ישראל במדדים של שינוי לאורך השנים תשקף מגמת שיפור עקבית שחלה בכל מקצועות היסוד הנמדדים במבחנים אלה, כזו שעולה במהותה על המגמות שתועדו עבור ארבע מדינות נבחרות נוספות. על רקע תשומת הלב המועטה יחסית שניתנה לנקודת המבט הזו על הישגי המדינות במבחנים הבין לאומיים, עבודה זו תחתור להציג דוח השוואתי מקיף של נתוני השינוי במבחני פיז"ה ביחס למדינות אלו לאורך כל השנים. לצורך ניתוח המקרה של ישראל במונחי מדד השינוי, נבחרו ארבע המדינות הבאות לצורכי השוואה: פינלנד, ארצות הברית, בריטניה וצרפת. פינלנד מהווה בחירה קלאסית להשוואה, כמי שבצוותא עם כמה מדינות אחרות מובילה את העולם בדירוג המבחנים הבין לאומיים בכל המקצועות, וכמי שזה זמן רב מהווה מוקד להתעניינותם של גורמי חינוך בישראל ובעולם. רבים בארץ ובעולם רואים במערכת החינוך הפינית מודל לחיקוי, לכן חשוב לבחון את הישגיה ממגוון נקודות מבט. בריטניה, צרפת וארצות הברית (לפי הסדר) מאיישות כעת את מרכז טבלת הישגים ומעלה של

מבחני פיז"ה, כאשר לפחות בשני מחזורים של המבחן התמקמו מדינות אלה ב-15 המקומות הראשונים. הן נבחרו לשם השוואה עם ישראל בשל הדמיון ביניהן ביחסיות האתנית שבין הרוב הדמוגרפי לבין המיעוטים הגדולים במדינות אלו. כמו כן, הכלכלה שלהן (במונחי תמ"ג) דומה לזו של ישראל (אם כי בארצות הברית התמ"ג גבוה יותר), והן בעלות רמת אבטלה דומה (למעט צרפת). לבסוף, מדינות אלה הן בעלות צורות ממשל דמוקרטיות ומתקדמות. פינלנד נכללה כאן על שום מעמדה כ"מעצמת חינוך", אף שרמת ההשקעה בחינוך גבוהה יחסית גם ויש לה דמוגרפיה הומוגנית. חרף הדמיון התרבותי והחינוכי הרב לזה של ישראל, בריטניה, צרפת וארצות הברית עדיין מצויות גבוה ממנה באופן משמעותי בדירוג הבין לאומי במבחני פיז"ה וזאת למרות המאמץ הישראלי לחקות את המדיניות "אף תלמיד אינו נשאר מאחור" האמריקנית (יוגב ואחרים, 2009). אף על פי כן, שלוש המדינות הללו נמצאות כולן במגמת ירידה מאז שנת 2000, בעוד שישראל נמצאת מאז במגמת שיפור תמידית בכל המקצועות. מטרת המחקר הנוכחי, אפוא, הינה לבטא את ההבדלים בין המדינות הללו במדדים של שינוי עצמי, ולהראות שבמונחים ההשוואתיים הללו – ההיררכיה ביניהן מתהפכת.

שיטת המחקר

המשתתפים

תוצאות מבחני פיז"ה בשלושת תחומי היסוד האורייניים (הבנת הנקרא, מתמטיקה ומדעים) בין השנים 2000-2015 (לפירוט השנים, ראו לוח 1), כפי שדווחו עבור חמש המדינות הבאות: פינלנד, ישראל, ארצות הברית, בריטניה וצרפת.

מקור הנתונים

עבודה זו מתבססת על נתוני הישגים של מדינות נבחרות שהשתתפו במבחני פיז"ה לאורך כל השנים, כפי שהתפרסמו בקובצי מאגרי המחקר של הארגון לשיתוף פעולה ולפיתוח כלכלי – OECD.⁵ ממאגרי המידע הוצאו ונותחו באופן השוואתי נתונין של חמש מדינות: ישראל, ארצות הברית, פינלנד, בריטניה וצרפת. ביחס לכל מדינה נלקח ממוצע הציונים הלאומי בכל מחזור פיז"ה, כפי שהוא מדווח עבור כל מקצוע יסוד בנפרד (מדעים, מתמטיקה, הבנת הנקרא).

⁵ OECD (Organization for Economic Co-Operation and Cultural Development) (2003). PISA 2000. Available from <http://www.oecd.org/edu/school/programmeforinternationalstudentassessmenttpisa/33690591.pdf>

OECD (Organization for Economic Co-Operation and Cultural Development) (2005). PISA 2003 Technical Report. Available from <https://www.oecd.org/edu/school/programmeforinternationalstudentassessmenttpisa/35188570.pdf>

OECD (Organization for Economic Co-Operation and Cultural Development) (2007). PISA 2006 - OEI. Available from <http://www.oei.es/historico/evaluacioneducativa/ResumenEjecutivoFinalingles.pdf>.

OECD (Organization for Economic Co-Operation and Cultural Development) (2010). PISA 2009 Results: What Students Know and Can Do. Retrieved from <http://www.oecd.org/pisa/pisaproducts/48852548.pdf>.
OECD (Organization for Economic Co-Operation and Cultural Development) (2014). PISA 2012 Results in Focus What 15-year-olds know and what they can do with what they know. Retrieved from <https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>.

OECD (Organization for Economic Co-Operation and Cultural Development) (2016). PISA 2015 Results in Focus. Retrieved from <https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf>

המצאים

מדרג ציוני פיז"ה הבין לאומיים והתפלגותם (2000-2015) במבט קונבנציונלי להלן נבחן את ההישגים של חמש המדינות שנכללו במחקר במבחי פיז"ה בשני העשורים האחרונים, תוך סקירת דירוגן הבין לאומי בכל מחזור, כפי שפורסם במהלך השנים. על מנת לבסס מבט השוואתי תמציתי, הבחינה תתמקד בשלוש נקודות זמן מייצגות של המבחנים (השנים 2000, 2009, 2015) ובכמה מדינות המשקפות מגמות בולטות הרלוונטיות לענייננו. כפי שזכר קודם לכן, ציוני מבחני הפיז"ה מתוקננים לפי ממוצע 500 וסטיית תקן של 100.

לוח 1 מציג את נתוני ההישגים של חמש המדינות בשלושת מקצועות היסוד בכל מבחני הפיז"ה משנת 2000 ועד היום. מתוך מבט ממוקד בדירוגי המדינות שבהן מתמקד דיוננו, עולה כי בשנה הראשונה שבה החל מבחן פיז"ה (2000), פינלנד מדורגת כגבוהה ביותר מבין החמש, בהיותה בין חמשת המקומות הראשונים בכל שלושת התחומים יחדיו. לעומת זאת ארצות הברית מדורגת בין המקומות 15-20 בתחומים אלו (בעדיפות לתחום המדעים).

שתי מדינות אירופה הבולטות, אנגליה וצרפת, משיגות ציונים גבוהים יותר מעמיתתן האמריקנית ומתברגות במקומות 5-8 (בריטניה), ו-11-13 (צרפת), בעוד ישראל מדורגת רחוק מאחור, במקומות 31-33. בשנת 2003 לא השתתפה ישראל במבחנים. באופן כללי, אפשר לראות שפרט לפינלנד, שמשמרת את מקומה באזור חמש המדינות הראשונות, שאר המדינות שבהן מתמקד דיוננו יורדות מאז שנת 2000 שלושה עד חמישה מקומות בדירוג העולמי בכל שלושת המקצועות.

הדיון הבא יתעכב על המגמות העיקריות שמופיעות החל משנת 2009, על רקע העלייה החדה בכמות המדינות המשתתפות במבחני פיז"ה הבין לאומיים – 64 מדינות לעומת 41 מדינות בשנת 2000. ובכן, במבחנים אלה שומרת פינלנד על המקומות ה-2-5, ארצות הברית נעה בין המקומות 17-31 (עדיפות לתחום הקריאה), ואילו בריטניה וצרפת נמצאות במקומות ה-16-25 (עדיפות לקריאה) וה-22-27 (עדיפות למדעים ולמתמטיקה), בהתאמה. זאת בעוד שישאל ממוקמת הרחק מאחור, בין המקומות ה-37 ל-42 (עדיפות לקריאה). אפשר לראות, אם כן, שבין היתר על רקע הגידול בכמות המדינות המשתתפות במבחנים, כל המדינות הנידונות פה (פרט לפינלנד, כאמור), מאבדות ממקומן ההתחלתי בדירוג העולמי. חדי העין אף יבחינו בכך שפערי ההישגים בין שלושת התחומים גדלו באופן עקבי אצל כל המדינות.

ננעל את הסקירה הזאת במבט על שנת המבחנים האחרונה עד כה, שנת 2015. ושוב, כאן, היחידה שמשמרת את מיקומה בחמשת המקומות הראשונים, היא פינלנד. ארצות הברית תופסת את המקומות 24-25, מצמצמת פערים בין התחומים, אך פרט למתמטיקה אינה משיגה התקדמות כללית משמעותית. בריטניה (15-21) וצרפת (20-27) משיגות שיפור כללי של מקומות אחדים כל אחת, וישראל נמצאת הרחק מאחור, בין המקומות 37 עד 40 בשלושת המקצועות.

לוח 1

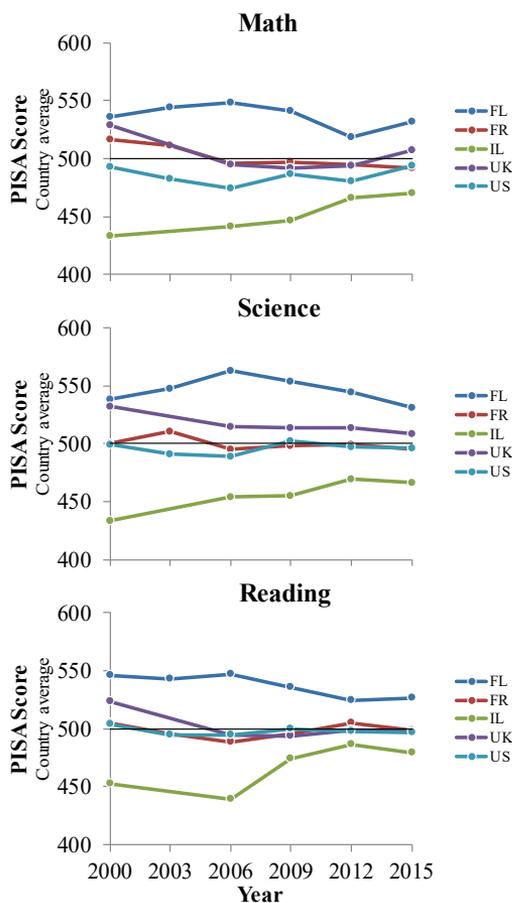
נתוני ההישגים של חמש המדינות הנבחרות במבחני פיז"ה בין השנים 2000-2015

ציון (דירוג)						שנה מספר מדינות משתתפות	מדינה
2015 70	2012 65	2009 64	2006 56	2003 41	2000 41		
(4) 526	(6) 524	(3) 536	(2) 547	(1) 543	(1) 546	קריאה	1. פינלנד
(5) 532	(12) 519	(5) 541	(2) 548	(2) 548	(5) 536	מתמטיקה	
(5) 531	(5) 545	(2) 554	(1) 563	(1) 548	(4) 538	מדעים	
(21) 498	(23) 499	(25) 494	(18) 495	(12) 507	(8) 523	קריאה	2. בריטניה
(15) 507	(26) 494	(28) 492	(25) 495	(17) 510	(9) 529	מתמטיקה	
(15) 509	(21) 514	(16) 514	(14) 515	(12) 518	(5) 532	מדעים	
(20) 499	(21) 505	(22) 496	(23) 488	(18) 496	(15) 505	קריאה	3. צרפת
(27) 492	(25) 495	(22) 497	(24) 496	(13) 519	(11) 517	מתמטיקה	
(27) 495	(26) 499	(27) 498	(25) 495	(14) 511	(13) 500	מדעים	
(24) 497	(24) 498	(17) 500	(18) 495	(19) 495	(16) 504	קריאה	4. ארצות הברית
(25) 494	(36) 481	(31) 487	(36) 474	(30) 477	(20) 493	מתמטיקה	
(25) 496	(28) 497	(23) 502	(29) 489	(24) 491	(15) 499	מדעים	
(37) 479	(34) 486	(37) 474	(40) 439	-	(30) 452	קריאה	5. ישראל
(40) 470	(41) 466	(42) 447	(41) 442	-	(31) 433	מתמטיקה	
(40) 467	(41) 470	(42) 455	(38) 454	-	(33) 434	מדעים	

בינתיים קשה לומר בבירור מה אפשר ללמוד מהמגמות שנסקרו עד כה ביחס להשתנות ההישגים של חמש המדינות הנידונות כאן. במבט כללי נראה כי החל משנת 2003 הן מאיישות טווחי דירוג עקביים יחסית בדירוג העולמי (בכפוף לשינויים שחלו במספר המדינות המשתתפות), עם נטייה עקבית משנת 2000 לאבד כמה מקומות בכל מחזור (פרט לפינלנד) והתייצבות או עלייה במחזור האחרון. לוח 1 מדרג באופן כללי את חמש המדינות לפי הישגיהן הממוצעים במבחנים ומיקומן בדירוג פיז"ה העולמי במרוצת השנים, כך שפינלנד ממוקמת במקום הראשון מבין החמש עם הדירוגים הגבוהים ביותר, אחריה (בכפוף ליוצא מן הכלל) בריטניה וצרפת ולבסוף ארצות הברית וישראל עם הציונים הנמוכים ביותר – רובם מתחת לממוצע הבין לאומי.

תרשים 1

תיאור ההישגים במבחני פיז"ה של חמש המדינות הנבחרות בין השנים 2000-2015



מבט ויזואלי ממוקד יותר ביחס ההישגים בין חמש המדינות הנידונות פה (תרשים 1), מראה כי מדרג חמש המדינות הכללי ניכר גם ברמה הפרטנית בכל אחד מן המקצועות. בכל התחומים מובילה פינלנד את החמישה, עם ציונים מעל הממוצע בכל השנים, וישראל במקום האחרון עם נתוני הישגים מתחת לממוצע. ביחס למקצוע המתמטיקה, נראה כי בריטניה וצרפת נושקות לממוצע במקום השני והשלישי, ואילו ארצות הברית אחריהן עם ציונים נמוכים מהממוצע הבין לאומי בכל השנים. ביחס למדעים, בריטניה בדרך כלל במקום השני עם הישגים מעל הממוצע בכל השנים, ואילו צרפת וארצות הברית נושקות יחד לממוצע העולמי (כשלצרפת יתרון קל). ביחס לקריאה, לבסוף, אפשר לראות כי על פי רוב, שלוש המדינות האמצעיות משיגות הישגים זהים במקצוע זה בכל השנים, בציונים הנושקים לממוצע הבין לאומי.

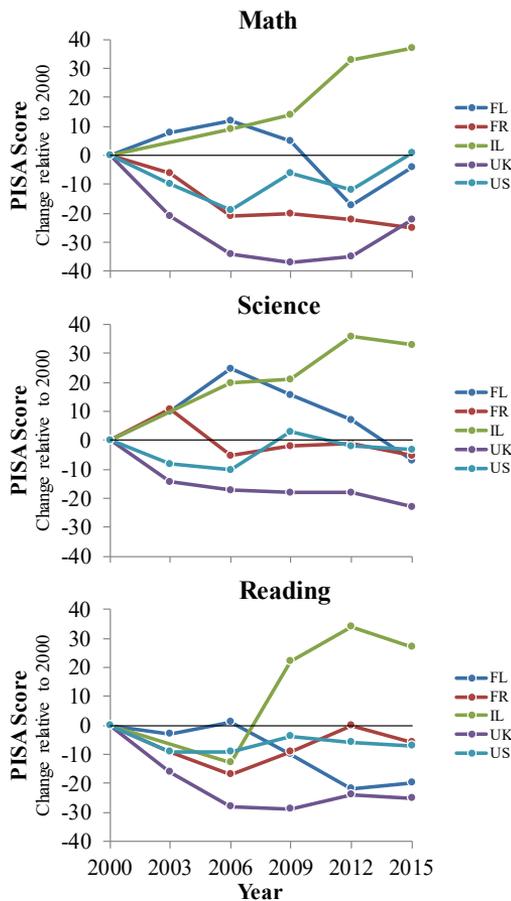
השוואה בין לאומית קונבנציונלית של נתוני ההישגים של חמש המדינות לאורך השנים מציעה אפוא היררכיה ברורה למדי לפחות ביחס לשלוש מדינות מתוך החמש, מבחינת הרמה החינוכית

של תלמידיהן בשלושת מקצועות היסוד. במונחים אלה הישגי פינלנד במבחני פיז"ה משקפים בבירור את הרמה החינוכית הגבוהה ביותר מבין החמש, כשאחריה מהבחינה הזאת נמצאות מדינות אירופה וארצות הברית ולבסוף ישראל.

תוצאות מבחני פיז"ה במדדים תוך קבוצתיים של שינוי (Measurement of change) בחלק זה נבחן את הישגי חמש המדינות האמורות במונחים של טיב השינוי שהושג בכל מחזור של מבחני פיז"ה בהשוואה להישגים הקודמים, זאת במטרה לדרג את המדינות לפי מדד של שינוי כללי – מעבר לשנים, המתאר את המגמה (חיובית/שלילית) ואת הגודל של השינויים שהושגו על ידי תלמידיהן בשני העשורים האחרונים במקצועות היסוד.

תרשים 2

תיאור השינוי בנקודות בהישגי חמש המדינות הנבחרות במבחני פיז"ה בשלושת המקצועות בין השנים 2015-2000



בהתייחס לחמש המדינות הנידונות פה, תרשים 2 מציג את פערי ההישגים יחסית לתוצאות שנת 2000 בשלושת מבחני פיז"ה. שינוי חיובי בהישגים של מדינה נתונה מתואר בנקודות שמיקומן ביחס לנקודת תחילת ביצוע מבחני פיז"ה בשנת 2000. עיון בתרשים מגלה, בראש ובראשונה, כי המדינה היחידה מבין חמש המדינות שנמצאת במגמת עלייה חיובית כמעט מתמדת בתוצאות פיז"ה מאז שנת 2000 – בכל אחד משלושת תחומי היסוד – היא ישראל (פרט לתחום הקריאה בשנת 2006).

לא זו בלבד שמגמת השינוי הישראלית היא על פי רוב חיובית (קרי, השינוי בהישגים בשנה נתונה בהשוואה להישגים בשנת 2000 הוא גדול מ־0), אלא שהתוספת שלה בשנה נתונה בהשוואה להישגי שנת 2000 גדולה על פי רוב מזו של השנה הקודמת (כפי שמשקפת מגמת העלייה של הגרף הישראלי). בעקבות מבחני פיז"ה האחרונים, מגמת השיפור הישראלית צמחה עד כדי 40 נקודות בקירוב, המהווה שיפור של כ־8%, בשלושת המקצועות בהשוואה לנקודת הפתיחה שלה בשנת 2000.

פינלנד היא המדינה היחידה, פרט לישראל, שההישגים שלה במרוצת השנים במתמטיקה ובמדעים (אך לא בקריאה) נמצאים באופן חלקי במגמת עלייה יחסית לשנת 2000, אולם החל משנת 2006 קצב השיפור בהישגים שלה בין שנה לשנה העוקבת נמצאת במגמת ירידה עקבית. מגמת ירידה זו היא עקבית דיה, כך שהישגיה של פינלנד במתמטיקה ובמדעים במבחני פיז"ה האחרונים יורדים אל מתחת להישגיה בשנת 2000.

באשר לשאר המדינות, אפשר לראות בתרשים 2 שרובן ככולן נמצאות במגמת ירידה בהישגים בשלושת התחומים יחסית לשנת 2000 (כפי שמשקף בנקודות מתחת לקו האמצע), אולם מגמה זו נשארת יציבה יחסית ואינה מידרדרת. ארצות הברית, לדוגמה, שומרת לאורך השנים על פער יציב בן נקודות בודדות בהשוואה להישגיה בשנת 2000 בשלושת המקצועות. בריטניה, שהחלה עם הישגים גבוהים למדי בשנת 2000, מראה ירידה משמעותית בהישגיה ב־2006 בכל התחומים, אולם החל מ־2006 ירידה זו מתייצבת ואינה פוחתת עוד באופן משמעותי. מגמה זהה ניכרת גם בהישגי הצרפתים במקצועות הקריאה והמתמטיקה, אולם החל מ־2006 הישגיהם במדעים אינם נופלים באופן משמעותי מהישגיהם בשנת 2000.

במטרה לשקלל את נתוני השינוי של המדינות השונות לכדי כללי של טיב השינוי, חושבו ממוצעי השינויים בהישגי חמש המדינות בכל מקצוע של פיז"ה בהשוואה להישגי שנת 2000 כנקודת ייחוס (כלומר ממוצע ההפרשים בין התוצאות בשנת 2000 והתוצאות בכל שאר המחוזים העוקבים). באופן כללי, ככל שמדד שינוי זה גבוה יותר, כך הוא משקף מגמת שינוי חיובית יותר בכל תחום בהשוואה לנקודת הפתיחה. למדדי השינוי עבור שלושת התחומים חושב הממוצע כללי, כדי לדרג את חמש המדינות לפי טיב השינוי הכללי שלהן מעבר למקצוע. לוח 2 מציג את התוצאות האלה.

לוח 2
 תיאור מדדי השינוי במונחי ממוצע הפערים (\bar{D}) בין ציוני פיז"ה בין השנים 2000 לכל השנים
 העוקבות

מדינה	מתמטיקה	מדעים	קריאה	מדד השינוי הכללי
1. ישראל	+23.25	+27.5	+17.5	+22.83
2. פינלנד	+0.8	+10.2	-10.8	+0.06
3. ארצות הברית	-7.0	-10.4	-4.0	-7.13
4. צרפת	-18.8	-0.4	-8.2	-9.13
5. בריטניה	-29.4	-18	-24.4	-23.9

הערה: ציוני מדדי השינוי בכל מקצוע התקבלו באמצעות הנוסחה הבאה:
 $i - i$ אינדקס מבחני פיז"ה החל משנתה השנייה וכלה בשנתה השישית; X - ציון פיז"ה בשנת 2000/i

מלוח 2 עולה כי שתי המדינות היחידות שמדד השינוי הכללי שלהן חיובי, הינן ישראל ופינלנד. מדד זה משקף מגמה כללית של שיפור ביחס לציוני התלמידים במדינות הללו במבחני פיז"ה. ביחס לפינלנד הנתונים מצביעים על ירידה ממוצעת של כ-11 נקודות ברמת הקריאה בהשוואה לנקודת המוצא שלה במבחנים האלה בשנת 2000. אצל ישראל, לעומת זאת, חל שיפור חיובי גדול יותר בהשוואה לפינלנד בכל שלושת המקצועות, עם התווספות ממוצעת של כמעט 23 נקודות בהישגיה במהלך השנים בהשוואה לנקודת הפתיחה של בשנת 2000. בתרשים 2 משתקפת המגמה הזאת בגרף של ישראל, המצוי ברובו מעל קו האמצע בשלושת המקצועות. באשר לשלוש המדינות הנוספות, אפשר לראות שמדד השינוי המצטבר הכללי שלהן הינו שלילי, כאשר זה משקף אצל כולן כאחת ירידה ממוצעת בכל שלושת המקצועות בהשוואה לנקודת המוצא שלהן בשנת 2000. המגמה השלילית הבולטת ביותר בהקשר הזה מצויה אצל בריטניה, שפתחה את מבחני פיז"ה עם הישגים גבוהים יחסית בכל שלושת התחומים (ראו לוח 1), והידרדרה באופן משמעותי מנקודה זו בהמשך השנים. מדד השינוי המבוסס בלוח 2 מייטיב לשקף את התנועות שחלו בהישגים ביחס לנקודת המוצא, אך חסרונו בכך שהוא אינו רגיש לשיפור שעשוי להתרחש בהמשך בין שני מחזורי עוקבים. כך, למשל, מגמות השיפור שחלו בין מחזורי עוקבים אצל בריטניה אחרי שנת 2000 אינן מקבלות ביטוי במדד הזה, היות שהישגיה בשנים הללו נמוכים מאלו שהשיגה בשנת 2000. כדי להתגבר על הבעייתיות הזו חושב מדד שינוי נוסף, המתבסס בכל מקצוע לממוצע הפערים בין שנת פיז"ה אחת לשנה העוקבת לה (להלן: מדד השינוי העוקב). לוח 3 מתאר את תוצאות המדדים בעבור חמש המדינות, ומדרג אותן לפי מדד השינוי העוקב הכללי (המהווה ממוצע השינויים הכללי מעבר לתחום).

לוח 3

תיאור מדדי השינוי במונחי ממוצע הפערים (\bar{D}) בין ציוני פיז"ה בשנה i לשנה העוקבת ($i+1$)

ממד השינוי הכללי	קריאה	מדעים	מתמטיקה	
+8.08	+6.75	+8.25	+9.25	1. ישראל
-0.73	-4.0	+2.6	-0.8	2. פינלנד
-1.26	-1.4	-2.6	+0.2	3. ארצות הברית
-2.4	-1.2	-1.0	-5.0	4. צרפת
-4.6	-5.0	-4.6	-4.4	5. בריטניה

הערה: ציוני מדדי השינוי בכל מקצוע התקבלו באמצעות הנוסחה הבאה:
 $i - i$ אינדקס מבחני פיז"ה החל משנתה הראשונה וכלה בשנתה החמישית; $X - X$ ציון פיז"ה בשנת i

לוח 3 מלמד שאף על פי שבמדד השינוי העוקב מציגה פינלנד מגמת שינוי שלילית, הרי שככלל – דירוג המדינות אינו משתנה בסדרו. גם המדד הכללי המשקף את השינוי העוקב בין מחזורי פיז"ה בתחומי היסוד השונים, מציב את ישראל בראש הטבלה עם מדד שינוי עוקב כללי חיובי. כיוונו וערכו של מדד זה מלמד כי תלמידי ישראל השתפרו ביותר מ-8 נקודות בממוצע בין מחזור למחזור. טווח המדדים הפרטני של תחומי היסוד מלמד שהשיפור של התלמיד הישראלי במתמטיקה ובאוריינות מדעית היה הגדול ביותר לאורך השנים, אך גם בהבנת הנקרא הושג שיפור עקבי המתקרב ל-7 נקודות בממוצע בכל מחזור של פיז"ה.

השינויים החיוביים הנדונים פה ביחס לישראל משתקפים במגמת העלייה של הגרף (תרשים 2) בין הנקודות של השנים בכל מקצוע. לעומתה, שלוש המדינות הסוגרות את הטבלה מציגות כולן מגמות שינוי שליליות של הישגים בין המחזורים העוקבים של פיז"ה, הנעות בטווח ירידה עקבי של 1.3 ל-4.5 נקודות בממוצע. גם על פי המדד הנוכחי, הבולטת ביותר בהן היא בריטניה, שהישגי תלמידיה בכל שלושת התחומים ירדו בממוצע ביותר מ-4 נקודות בין מחזור למחזור. לסיכום, שני מדדים שונים של שינוי עצמי בהישגים במבחני פיז"ה בשלושת מקצועות היסוד מציבים את מגמת השיפור של תלמידי ישראל לאורך השנים במקום הגבוה ביותר מבין חמש המדינות שנבחרו לצורך דיונו. זאת אף על פי שדירוגה במונחי הישגים גרידא הינו הנמוך ביותר מביניהן בכל השנים הללו מאז מחזור המבחנים הראשון. אומנם יש לתת את הדעת לעובדה שישראל מסוגלת להשתפר ממקום נמוך יותר ממדינות ההשוואה שלה, אך אין די בהנחה כי קל יותר להשתפר ממקום נמוך כדי להסביר את יתרונה עליהן בשני מדדי השינוי. יש לזכור כי מאז שנת 2003 מתקברים הישגיהן של שלוש מתוך ארבע מדינות ההשוואה לממוצע הכללי (500 עם סטיית תקן של 100), ולעיתים אף נמוכים ממנו בכל שלושת התחומים של מבחני פיז"ה (ראו טבלה 1).

דיון

במאמר הנוכחי ביקשנו להציע דרך לבחינה השוואתית של ההישגים הבין לאומיים במבחני פיז"ה, המבוססת על מדדים תיאוריים של שינוי. גישה זו, שיושמה במאמר, נסמכת על האוריינטציה של מדידת השינוי התוך אישי של ביצועים אינדיבידואליים במונחים של שיפור וצמיחה לאורך זמן (Barrett & Alexander, 1989; Willet, 1994). בהיותה מדורגת במקום נמוך משמעותית מעמיתותיה בדירוג העולמי (ועל פי רוב, נמוך מהממוצע העולמי), היותה ישראל "חקר מקרה" קלאסי להמחשת התזה של המאמר ביחס לחשיבות של מדידת השינוי בהישגים הבין לאומיים במבחני פיז"ה.

הישגיה של ישראל בשנת 2000 היו מתחת לממוצע הכללי. מאז חל שיפור עקבי בהישגיה בכל המחוזות הבאים (במיוחד החל מ-2009), אך הם נותרו נמוכים מן הממוצע הבין לאומי, ונשארו ברמה דומה גם בשנים 2012 ו-2015. הממצא העיקרי העולה מניתוח הנתונים באמצעות מדדי השינוי שהוצעו פה, הינו כי ישראל השיגה מגמת שיפור חיובית ועקבית בכל המקצועות מאז שנת 2000 ועד היום, וכי זו עדיפה על מגמת השינוי של עמיתותיה האירופיות והאמריקניות, שנכללו בניתוח.

מדדי השינוי ביחס להישגי התלמיד הישראלי שיקפו עלייה של 6-9 נקודות בממוצע בכל מחזור ומחזור של פיז"ה לאורך השנים (בכל המקצועות). תוצאות אלו, יחד עם המדדים הנגזרים מהשוואה הרגילה בין הציונים הממוצעים של המדינות במבחני פיז"ה, ממחישים את החשיבות של שתי הדרכים להערכת הביצועים של מערכות החינוך של המדינות באמצעות מבחנים בין לאומיים. הסיבה לכך היא שאחד מהאינדיקטורים הללו לבדו עלול, בנסיבות מסוימות, להציג תמונה חלקית ומטעה של ביצועי החינוך של המדינות לאורך זמן.

ישראל השיגה שיפור כללי של יותר מ-20 נקודות בהשוואה להישגיה בשנת 2000 בכל שלושת המקצועות של פיז"ה, המהווים שיפור של 6 עד 9 נקודות בכל מחזור. על רקע מדדים אלה, נתוני השינוי בהישגים של כל שאר המדינות, פרט לפינלנד, היה שלילי – מה שמלמד על מגמת הידרדרות עקבית בהישגים בין מחזור למחזור, ובאופן כללי בהשוואה למחזור הראשון של פיז"ה בשנת 2000. במדדי טיב השינוי הנוכחיים, אפוא, ישראל ופינלנד מדורגות במקומות הראשונים (בהתאמה) בין חמש המדינות, ואילו בריטניה נמצאת במקום האחרון. עם זאת, יש לציין כי העדיפות של ישראל במונחים של ציוני השינוי, עשויה לנבוע מהעובדה שמדינות שהישגיהן גבוהים במבחנים מלכתחילה (למשל, פינלנד) מוגבלות במידת יכולתן של תלמידיהם להוסיף ולהשתפר (תופעה הידועה כ"אפקט גג"), ואילו לתלמידים ממדינות בעלות הישגים נמוכים יש הרבה יותר מקום לשיפור.

ובכן, בניגוד לסברה הרווחת בנוגע למצבו המדרדר של החינוך בישראל (בן דוד, 2003; Ben-David, 2010; Blass, 2016), המסקנה המתבקשת מן הממצאים הללו היא שרמתו החינוכית של התלמיד הישראלי (במונחי ההישגים הבין לאומיים במבחני פיז"ה) נמצאת דווקא במגמת שיפור עקבי מאז המחזור הראשון בשנת 2000, שכן ההישגים של תלמידיה גבוהים באופן משמעותי מאז המחזור הראשון של פיז"ה. אף על פי שלא לגמרי ברור מהי משמעותה של מגמת שיפור זו שניכרת בנתונים, ברור לחלוטין מהממצאים כי לא חלה כל ירידה ברמתו של התלמיד הישראלי על פי הישגים אלה, אף שביצועי ישראל נותרו נמוכים מהממוצע במדינות המשתתפות.

מחקר זה, שלראשונה מתייחס להישגים במבחני פיז"ה הבין לאומיים לאורך כל ששת מחזוריהם (וחמישה מחזורים מתוכם שבהם השתתפה ישראל), מחזק את הספק שהעלו חוקרים אחדים באשר לאפשרות שהחינוך הישראלי נמצא בירידה מתמדת (יוגב ואחרים, Feniger et al., 2009; 2012). ממחקר זה עולה כי לא זו בלבד שהתלמיד הישראלי לא איבד מידיעותיו האורייניות בתחומי המדעים והשפה, אלא שאלו אף השתפרו משמעותית לאורך השנים. העבודה הנוכחית התמקדה במכוון באפיון מגמת השינוי הזאת במונחים כמותיים תיאוריים, בעוד ניתוח משמעותה החינוכית מהווה נושא לבדיקה נפרדת בשל מורכבותו.

להערכת מחברי מאמר זה, קשה לצפות לשיפור שעולה בהרבה על 10 נקודות (העלייה הממוצעת של ישראל במקצוע המתמטיקה בין מחזורי פיז"ה) במשך שלוש השנים המפרידות בין כל מחזור בחינות של פיז"ה, כך שבמבט מצטבר (כ־40 נקודות שיפור מאז שנת 2000, המהווה קצת פחות מחצי סטיית תקן) – נראה כי מדובר במגמת שינוי שקצבה משמעותית. ואומנם, מגמת השיפור הישראלית במבחנים הבין לאומיים הוכרה זה מכבר כאחת המובילות בין מדינות העולם (Blass, 2016), אם כי ככל הידוע לנו עד כה היא טרם נותחה לאורך כל המחזורים.

בהנחה שהשיפור הישראלי המדווח פה ביחס לצינוי פיז"ה בשלושת תחומי האוריינות הוא אכן משמעותי, אנו מציעים כי אפשר שהדבר קשור במערכת של מדדי היעילות והצמיחה של בתי הספר (מיצ"ב) שהונהגה במערכת החינוך בישראל ב־2002 (מיד אחרי השתתפותה הראשונה של ישראל במבחני פיז"ה הבין לאומיים בשנת 2000). המיצ"ב הישראלי הוא מסגרת הערכת לומדים לאומית רחבת היקף, אשר בדומה למבחני פיז"ה בוחנת גם ביצועים של תלמידים בשלושת התחומים הבסיסיים (כלומר, שפת אם, מתמטיקה ומדעים וטכנולוגיה). המיצ"ב הוצג כחלק ממגמה עולמית הולכת וגוברת של עידוד הערכת הלמידה והגברת האחירות של בתי הספר ושל מערכת החינוך בכללה להוראה וללמידה.

כיצד אם כך אפשר ליישב את המגמה החיובית הזאת בהישגי התלמיד הישראלי עם העובדה שמיקומו בדירוג העולמי בשלושת המקצועות ירד באופן משמעותי לאורך כל התקופה הזאת? כך, למשל, בין 2000 ל־2006 השיגה ישראל עלייה של 9 נקודות במקצוע המתמטיקה, אך ירדה 10 מקומות בדירוג העולמי. בשנת 2009 הממוצע הישראלי בתחום זה צמח בעוד 5 נקודות בהשוואה למחזור הקודם, אך מקומה של ישראל המשיך וירד מקום אחד בדירוג העולמי. בדומה לכך, במדעים שיפר התלמיד הישראלי את הישגיו ב־20 נק' בממוצע בין 2000 ל־2006, אך ירד במיקום העולמי ב־5 מקומות, למקום ה־38.

ולבסוף, ב־2015 מאיישת ישראל את המקומות 37-40 בדירוגים, לאחר עלייה משמעותית בכל המקצועות מאז 2006. ברי כי הרקע המרכזי לסתירה הזאת בין שתי המגמות הוא העובדה שבמרוצת השנים הללו חלה תוספת משמעותית במספר המדינות המשתתפות במבחני פיז"ה (ביחס לדוגמה שנוכרת לעיל, נציין שהתווספו 15 מדינות ב־2006 ועוד 8 ב־2009). על רקע התווספות קבועה שחלה בכמות המדינות המשתתפות במבחני פיז"ה עד מחזור 2005 (נתונים מדויקים בהקשר הזה מובאים בלוח 1), הרי שמגמת השיפור היציבה של התלמיד הישראלי בכל המקצועות אינה מקבלת ביטוי מותאם בדירוג העולמי.

מיותר לציין כי התוצאות שאליהן העפיל התלמיד הישראלי במחזור פיז"ה האחרון בשלושת המקצועות, היו מדרגות אותו במקומות גבוהים יותר במתכונת של 41 מדינות (זו שהתקיימה

בשנת 2000). דומה הדבר למקרהו של תלמיד חיובי בבית הספר, שמצליח להשתפר בלימודיו באופן עקבי יחסית, ועדיין אינו זוכה להתבלט בכיתתו, משום שמדי שנה מתווספים אליה תלמידים שציוניהם גבוהים משלו.

מטרת ממצאינו איננה להציג תמונה ורודה של מצב החינוך בישראל, שהרי אין להתעלם מן העובדה שמגמת השיפור העקבית של התלמיד הישראלי מאז שנת 2000 עדיין מציבה אותו מתחת לממוצע העולמי בכל המקצועות הנבדקים במבחני פיז"ה. השתמעות הממצאים הללו הינה גלובלית, ואינה מצטמצמת למקרה הפרטי של ישראל. כפי שהודגם במאמר זה, מגמות של שיפור בביצועים אקדמיים שמשתקפות במדדים של שינוי עצמי עלולות שלא לתאום את רמת הביצוע היחסי עצמו.

בכו אחד עם עמדתם של חוקרים אחרים בישראל ובעולם (Bonnet, 2002; Feniger et al., 2012; Goldstien, 2004; Husen, 1987; Romainville, 2002), ממצאים אלו, מטרתם להזהיר מפני הבעייתיות הטמונה בהשוואת ביצועים אקדמיים של תלמידים על רקע הבדלי תרבויות ומשתנים סוציו-אקונומיים אחרים. התמקדות יתר בהישגי המדינות הגולמיים כמדד להשוואה חינוכית בין לאומית, עלולה לייצר רושם מוטעה ביחס לרמה החינוכית הלאומית, לזרוע תבהלה ציבורית ואף להשפיע לשלילה על מדיניות החינוך באותה מדינה.

לפיכך, יש צורך להכיר במגבלות של ההשוואות הבין-לאומיות. ווטק (Wuttke, 2007) מציף בהקשר הזה חלק מן המכשולים הסטטיסטיים, חוסר הוודאויות וחוסר השקיפות הטמונים במבחני פיז"ה. בראש ובראשונה מתועדות בעיות חמורות ביותר בכל הקשור לתרגום הפריטים לשפות השונות במדינות המשתתפות. נוסף על כך חלק מהמדינות אינן נדגמות באופן מייצג, משום ששיטת הדגימה במקרים מסוימים אינה מאפשרת ייצוגיות (ראו ערך המקרה של סינגפור). כך, למשל, חוסר השתתפות של תלמידים או של מגזרים חלשים עשוי להקפיץ את הציון של המדינה בכ-9 נקודות.

לבסוף, ישנה גם השוונות התרבותית בין מדינות ותרבויות, המתבטאת ביחס שונה של מערכת החינוך, של ההורים ושל התלמידים כלפי מבחנים בכלל וכלפי מבחני פיז"ה בפרט. יש להניח, בהקשר הזה, כי לתלמידים ממדינות שונות ישנה גישה נבדלת לגבי מבחנים בעלי השפעה אפסית על עתידם (Sjoberg, 2007). בסינגפור למשל "be best – teach to the test" היא מנטרה, וחוברות הכנה לפיז"ה מוצגות לראווה במכולות ובסופרמרקטים במחיר סמלי.

לעומת זאת במערב אומנם מתייחסים הפוליטיקאים לתוצאות פיז"ה באופן רציני ביותר, אך סביר מאוד שהתלמידים אינם שותפים לכמיהה שלהם, ולכן אינם משקיעים את המאמצים הנדרשים להצליח במטלה ארוכה וקשה ללא תועלת. על כן, הניסיון למדוד מסוגלות לימודית, הכרוכה במוטיבציה של התלמידים לסיים את המשימה בהצלחה, עלול שלא להצליח. נוכל להציע בזהירות את הסברה שהיחס התרבותי בישראל כלפי בחינות באופן כללי וכלפי חינוך פורמלי אינו מאפשר השוואה תקפה בינה ובין מדינות אחרות בעלות מערכת ערכים שונה.

מכאן שהשוואה בין לאומית ורב תרבותית של הישגים חינוכיים מחייבת מתן תשומת לב רבה יותר לבחינת ההישגים במונחי טיב השינוי העצמי לאורך זמן. זאת משום שמדדים של שינוי המשקפים את ההתקדמות של אינדיבידואל בביצועים אקדמיים (תלמיד או מערכת חינוך לאומית) ביחס לעצמו, הם לבדם יכולים להוות בסיס להשוואה חד ערכית בין לאומית בהישגים.

לא בטוח שנכון יהיה להשוות את ישראל למדינות בעלות באופי דמוגרפי ותרבותי שונה באופן מהותי (דוגמת פינלנד וסין), ואף לנסות לחקות אותן.

מעניין לציין שבמשך העשור האחרון, שבמהלכו פורסם מידע רב בדבר המגבלות הבין לאומיות של מבחני פיז"ה, אין התייחסות זהירה של פרשני הדירוג הבין לאומי ושל מקבלי החלטות ביחס לממצאיה. לכן, ההצעה לפרש את תוצאות מבחני פיז"ה במדדי שינוי של מערכת חינוכית תוך לאומית עשויה להניב תובנות מוצקות יותר. גישה זו מניחה שההבדלים בין מחזורי בחינה בכל מדינה הינם שוליים, ושההטיות דומות, לכן יש בה כדי להתגבר על השונות הבין לאומית שמקשה על השוואה חד ערכית בין הישגי המדינות.

יגוב ועמיתיו (2009) כתבו בהקשר הזה, כי "הדגשת הציון של תלמידי המדינות השונות במסגרת מדרג המייחס חשיבות לריבוד הבין לאומי של מדינות לפי הישגי תלמידיהן, היא בעוכרי המבחנים הבין לאומיים והמשמעות הציבורית והחינוכית הניתנת להם... אנו מתריעים על השימוש הפוליטי המוגזם במבחנים הבין לאומיים המוביל לעתים להתמקדות בסוגיות החינוכיות הפחות מהותיות, לאמץ התייחסות ביקורתית יותר באשר להשפעתם של המבחנים על המדיניות במדינות שונות" (שם, עמ' 350). בהזדמנות זו השכילו אותם חוקרים להצביע גם על הפערים ההישגיים העצומים בין מגזרים שונים בחברה הישראלית בצינוי פיז"ה, מה שיכול להסביר במידה רבה את העובדה שחרף השיפור העקבי – ישראל עדיין מצויה מתחת לממוצע הבין לאומי (משום שהישגי מגזרים מסוימים מושכים את הממוצע הישראלי הכללי למטה). על רקע זאת, הרי שמדדים של שינוי עצמי כבסיס להערכת הישגים בין לאומיים מקבלים משנה חשיבות.

קובעי מדיניות שיבחנו ברצינות את מגמת השיפור שהשיג התלמיד הישראלי במבחנים הבין לאומיים לאורך שני העשורים האחרונים, יידרשו לשאלה, האם על רקע התנאים הדמוגרפיים והחינוכיים הנוכחיים אפשר וצריך לצפות ליותר מכך בשנים הבאות? חוקרים בישראל ציינו כי הציפייה הלאומית לנסיקה דרסטית בדירוג העולמי במבחנים הבין לאומיים איננה ריאלית, על רקע משתנים כלכליים, דמוגרפיים ותרבותיים ישירים ועקיפים שמנבאים הישגים אלה (Feniger et al., 2012). הממצאים שבידינו ביחס לנתוני השינוי העצמי בהישגי ישראל במבחנים הבין לאומיים תומכים בעמדתם זאת.

אף על פי שהם מעידים על כך שהתלמיד הישראלי מצוי במגמת עלייה עקבית בביצועים החינוכיים הנמדדים במבחני פיז"ה, ואף כי יש במגמה זו כדי להדביק את הממוצע העולמי בעוד כעשור, הרי שהקצב החיובי הזה אינו מצדיק את ציפיית ועדת דברת משנת 2004⁶ להתברג בתוך פרק זמן של שנים ספורות מאז, בין מדינות העולם המובילות את ה-OECD. אי לכך, אנו מציעים להקדיש תשומת לב ציבורית וחינוכית רבה יותר בישראל (ואף במדינות אחרות בעולם) למגמות השינוי התוך לאומיות בהישגים במבחנים בין לאומיים, ולמקד את המדיניות בהיבטים אלה. בהקשר הזה, חידוד המדדים להערכת מגמות השינוי הללו בהישגים, תוך הצבת יעדי שיפור ריאליים וברורים להשגה בעשור הקרוב (ברמה הלאומית ככלל והמגזרית בפרט), עשויים למנוע את שחיקת המגמה החיובית הקיימת ככל שממוצע ההישגים הישראלי שואף לממוצע הבין לאומי ומתקרב אליו.

⁶ מדינת ישראל, כוח המשימה הלאומי לקידום החינוך בישראל (ועדת דברת) (2004). התוכנית הלאומית לחינוך. חלק א': עיקרי ההמלצות. ירושלים: המחבר.

מקורות מידע

בן דוד, ד' (2003). מערכת החינוך של ישראל בראי כלכלי-חברתי בעידן הגלובליזציה. הרבעון לכלכלה, 50(1), 47-71.

יוגב, א', ליבנה, ע', פניגר, י' (2009). סינגפור במקום כרכור? מבחני הישג בין לאומיים והגלובליזציה של יעדי החינוך. מגמות, מו(3), 337-355.

Anderson, J. O., Lin, H. S., Treagust, D. F., Ross, S. P., & Yore, L. D. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, 5(4), 591-614.

Atkin, J. M., & Black, P. (1997). Policy perils of international comparisons: The TIMSS case. *Phi Delta Kappan*, 79(1), 22.

Barrett, G. V., & Alexander, R. A. (1989). Rejoinder to Austin, Humphreys, and Hulin: Critical reanalysis of Barrett, Caldwell, and Alexander. *Personnel Psychology*, 42(3), 597-612.

Ben-David, D. (2010). Israel's education system - An international perspective and recommendations for reform, In D. Ben-David (Ed.), *State of the nation report: Society, economy and policy in Israel 2009* (pp. 115-156). Taub Center for Social Policy Studies in Israel.

Blake, J. (1989). Number of siblings and educational attainment. *Science*, 245, 32-36.

Blass, N. (2016). International exams and their importance to Israel's education system. Jerusalem: Taub center.

Bonnet, G. (2002). Reflections in a critical eye: on the pitfalls of international assessment. *Assessment in Education: Principles, Policy & Practice*, 9(3), 387-399.

Coleman, J. S. (1975). Methods and results in the IEA studies of effect of school on learning. *Review of Educational Research*, 45, 335-386.

Feniger, Y., & Lefstein, A. (2014). How not to reason with PISA data: An ironic investigation. *Journal of Education Policy*, 29(6), 845-855.

Feniger, Y., Livneh, I., & Yogev, A. (2012). Globalization and the politics of international tests: the case of Israel. *Comparative Education*, 48(3), 323-335.

Feniger, Y., & Shavit, Y. (2011). Fertility and educational achievement: Israel in comparative perspective. *Research Paper* (2010.05).

- Goldstein, H. (2004). International comparisons of student attainment: some issues arising from the PISA study. *Assessment in Education: principles, policy & practice*, 11(3), 319-330.
- Husén, T. (1987). Policy impact of IEA research. *Comparative Education Review*, 31(1), 29-46.
- Romainville, M. (2002). Du bon usage de PISA. *Revue nouvelle*, 115(3-4), 86-99.
- Sadler, T. D., & Zeidler, D. L. (2009). Scientific literacy, PISA, and socioscientific discourse: Assessment for progressive aims of science education. *Journal of Research in Science Teaching*, 46(8), 909-921.
- Sjøberg, S. (2007). PISA and “real life challenges”: Mission impossible? In S. T. Hopman, G. Brinek, & M. Retzl (Eds.), *PISA according to PISA — does PISA keep what it promises?* (pp. 203-225). Berlin: Lit Verlag.
- Turner, R., & Adams, R. J. (2007). The programme for international student assessment: An overview. *Journal of applied measurement*, 8(3), 237.
- Willett, J. B. (1989). Questions and answers in the measurement of change. In Z. Ernest Rothkopf (Ed.), *Review of Research in Education*, Vol 15 (pp. 345-422). Washington, D.C.: American Education Research Association.
- Willett, J. B. (1994). Measurement of Change. In T. Husen and T. N. Postlethwaite (Eds.), *International Encyclopedia of Education*, 2nd. ed. (pp. 671-678). Oxford, UK: Elsevier Science Press.
- Wuttke, J. (2007). Uncertainties and bias in PISA. In S. Th. Hopmann, G. Brinek & M. Retzl (Eds.), *PISA zufolge PISA - PISA according to PISA* (pp. 241-263). Wien: LIT.