**האם תוצאות מבחני השוואה סטנדרטיים מוטות כתוצאה מהבדלי מאמץ תלויי תרבות?**

יובל אופק-שני

# תקציר

**תוצאות תלמידי ישראל במבחני השוואה בינלאומיים מחדדות בשנים האחרונות שני נתונים מרכזיים – הראשון, פערים גדולים בקרב תלמידי ישראל – בין דוברי ערבית ועברית ובין תלמידים מרקע סוציו-אקונומי שונה. השני – פערים גדולים בין תלמידי ישראל ובין תלמידים במדינות המפותחות אליהן היינו רוצים להשתייך. לאור נתונים אלה, ועם העלייה בחשיבות שזוכים לה מבחני השוואה סטנדרטיים כמו המיצ"ב בישראל ומבחני TIMSS ו-PISA בעולם, יש חשיבות גדלה והולכת לתיקוף ממצאי המבחנים ובחינה של הביקורת כלפי מהימנותם. מחקרים קודמים הראו, כי קבוצות אוכלוסיה שונות מתנהגות באופן שונה במעבר ממבחנים עם השלכה על הנבחן (High-Stakes) למבחנים חסרי השלכה. ההבדלים בין גברים לנשים זכו לתשומת לב מחקרית רבה, וגם השוואה בין קבוצות אתניות שונות בארצות הברית בוצעה. מאמר זה ייפתח בהצגת נתונים אודות ביצועי תלמידי ישראל במבחני פיזה 2015 כפי שהופקו במערכת "פיזה פתוח" של המרכז לחקר האינטרנט באוניברסיטת חיפה וקרן טראמפ. לאחר מכן אציג סקירת ספרות שמתייחסת לאפשרות שחלק מנתונים אלה נובעים לא מהבדלים ביכולות אלא מהבדלים במאמץ שמשקיעים התלמידים במבחנים.**

# תלמידי ישראל במבחני פיזה 2015[[1]](#footnote-1)

מחקר פיזה נערך על-ידי ארגון ה- OECD (Organization for Economic Co-operation and Development) ומשתתפות בו מדינות רבות מכל רחבי העולם. המחקר בודק יכולות שונות של תלמידים בני 15 בשלושה תחומים עיקריים: קריאה, מתמטיקה ומדעים. המחקר בוחן באיזו מידה תלמידים הקרובים לסוף חינוך חובה (במרבית המדינות) רכשו כלי חשיבה כלליים והבנה של הנושאים הנבדקים באופן המאפשר התמודדות טובה ויעילה עם סביבתם, ולאו דווקא באיזו מידה רכשו ידע ותכנים ספציפיים המצופים עפ"י תכנית לימודים זו או אחרת. משום כך השאלות הנכללות במחקר בוחנות ידע בגישה מעשית, ידע החיוני ל"עולם המבוגרים", כישורי חיים ויכולת לפתור בעיות מורכבות המצריכות שילוב בין תחומים שונים, תוך דגש על מיומנויות. המחקר נערך במחזוריות של שלוש שנים. אחת לשלוש שנים נבדקים שלושת תחומי האוריינות, אך בכל פעם מושם דגש מיוחד על אחד משלושת התחומים (קריאה, מתמטיקה ומדעים).

מחקר פיזה 2015 נערך בישראל במהלך חודש מרץ 2015. לצורך ביצוע המחקר נדגמו באופן אקראי כ-200 בתי ספר. בכל בית-ספר נדגמו באופן אקראי כ-46 תלמידים ילידי 1999, ללא הגדרה של כיתה או שכבה בבית הספר. התלמידים שנדגמו, השיבו על מבחן ושאלון רקע, ומנהל בית הספר שנדגם השיב על שאלון נוסף. מחקר פיזה הוא מחקר ממוחשב במלואו - הן המבחנים והן שאלוני הרקע הועברו באופן ממוחשב.

המחקר בשנת 2015 בדק את שלושת תחומי האוריינות (מדעים, קריאה ומתמטיקה) עם דגש על תחום אוריינות המדעים. בתחום זה שולבו שאלות עם הנפשות והדמיות (סימולציות) של ניסויים מדעיים.

על פי סיכום נתוני המחקר שבוצע על ידי ראמ"ה (הרשות הארצית למדידה והערכה בחינוך), הישגי התלמידים הישראלים במחקר זה היו דומים להישגיהם במחקר הקודם שהתקיים בשנת 2012. גם בשנת 2015 היו הישגי התלמידים הישראלי נמוכים מממוצע המדינות המפותחות ב-3 תחומי המחקר וההישג הממוצע עמד על דירוג 40 מתוך 70 מדינות. התלמידים הישראלים מתבלטים במיוחד בפערים גדולים ביותר בין תלמידים חזקים לחלשים. בהשוואה של ציון האחוזון ה-95 לציון האחוזון ה-5 נמצא כי פער הציונים של ישראל בין 3 הגבוהים ביותר בכל אחד מהמבחנים. גם בפילוח אתני ותרבותי נמצאו פערים גדולים. בין דוברי עברית וערבית נמצא פער ממוצע של כ-100 נקודות ובין תלמידים מרקע תרבותי-חברתי-כלכלי שונה נמצאו פערים של עד 90 נקודות.

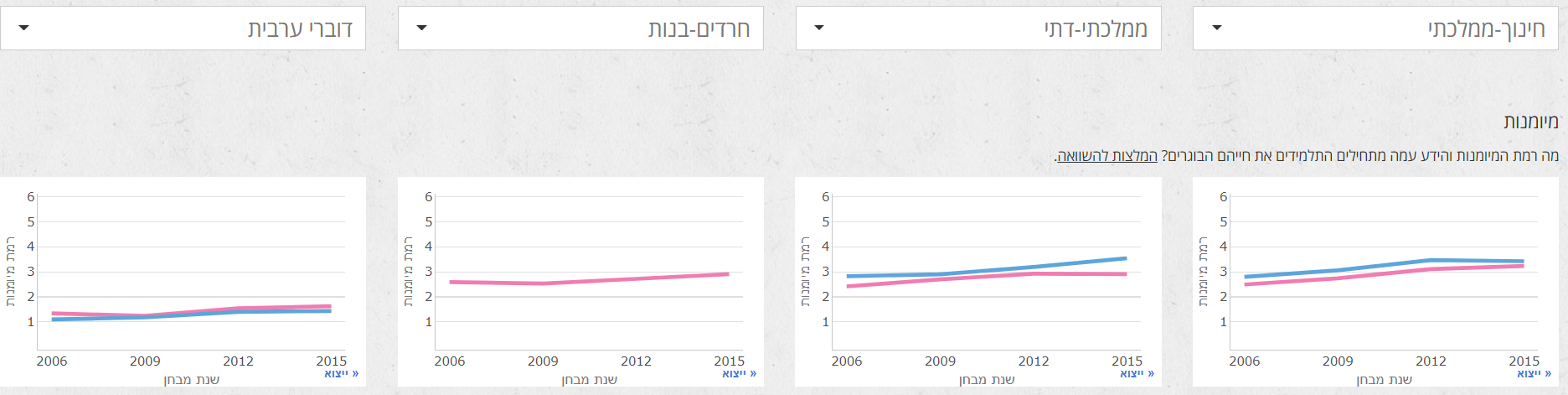
נתונים אלה היו נקודת הפתיחה לסקירת ממצאי המבחנים המוצגת כאן מתוך אתר "פיזה פתוח":

## **הישגים במבחנים לפי מגדר:**

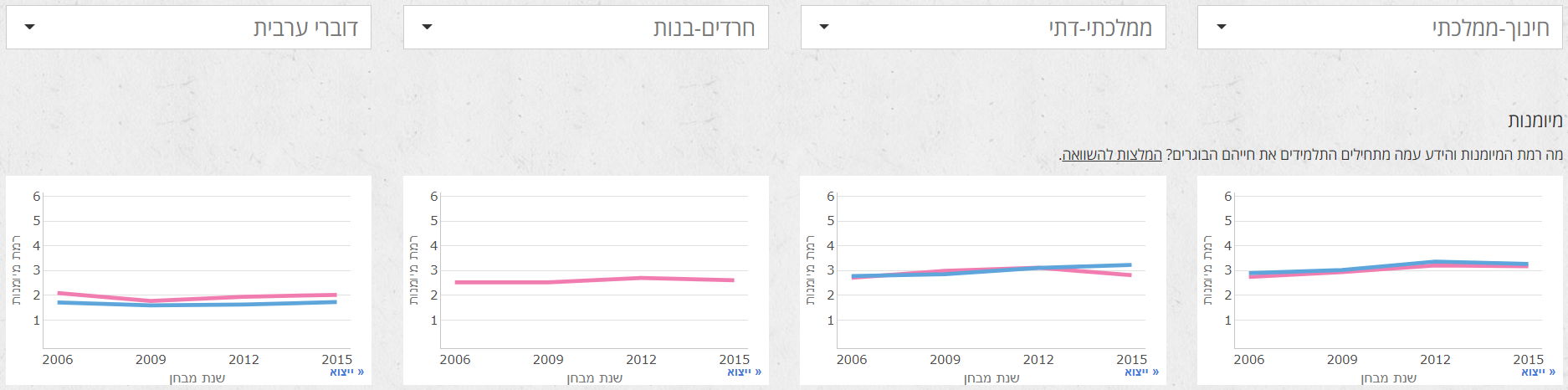
ב-3 התרשימים הבאים מוצגים הישגי התלמידים בישראל במחקרי פיזה בשנים 2006, 2009, 2012, 2015. ההישגים מוצגים בפילוח לתלמידי החינוך הממלכתי (דוברי עברית), ממלכתי דתי, החינוך החרדי (בנות בלבד) ודוברי ערבית תוך השוואה בין בנים (בכחול) לבנות (בורוד).

ניתן לראות כי התלמידים בחינוך הממלכתי מקבלים ציונים גבוהים מיתר זרמי החינוך בכל המבחנים וכי התלמידים דוברי הערבית מקבלים ציונים נמוכים משמעותית מיתר הזרמים.

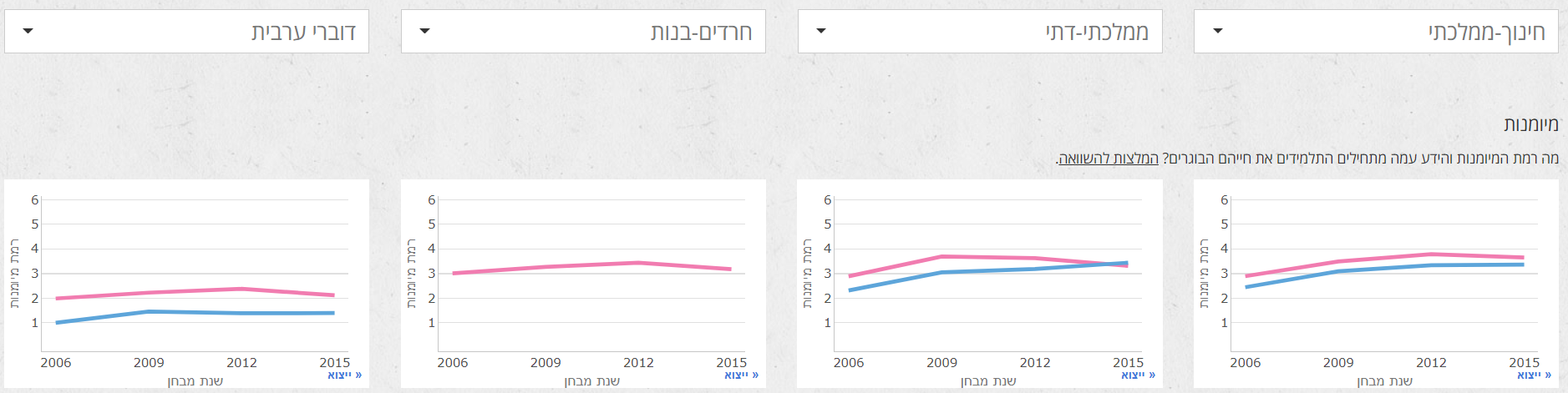
**מתמטיקה:**



**מדעים:**



**קריאה:**



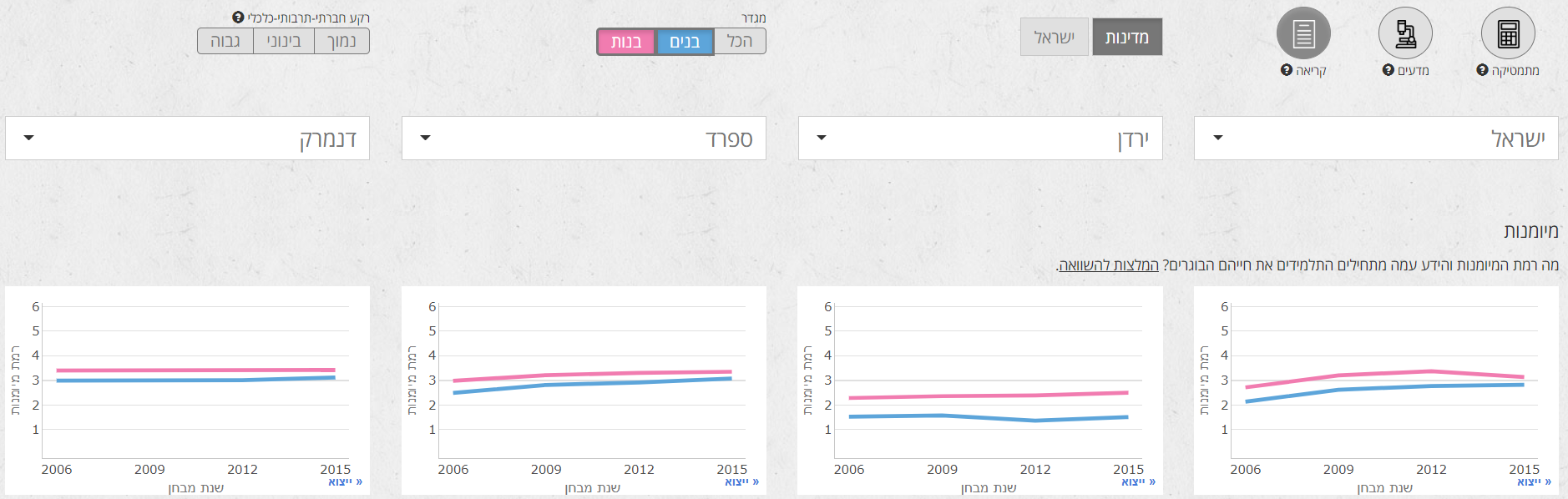
בעוד במתמטיקה ובמדעים לא ניכר הבדל משמעותי בין בנים ובנות, במבחן המיומנות בקריאה, ניכרת מגמה מעניינת. בקרב דוברי הערבית, יכולות הקריאה של בנות גבוהות משמעותית מאלה של הבנים. בשנת 2015 היה הניקוד של בנות ערביות גבוה ב-52 נקודות מזה של בנים. בנות דוברות עברית השיגו ניקוד גבוה ב-14 נקודות מהבנים ובחינוך הממלכתי דתי.

בהתאמה, שיעור המתקשים בקריאה בקרב הבנים דוברי הערבית (בעלי רמת מיומנות 2 ומטה) מגיע ל-32%.

**השוואה בינלאומית – הפערים בין בנים לבנות בקריאה מתמידים**

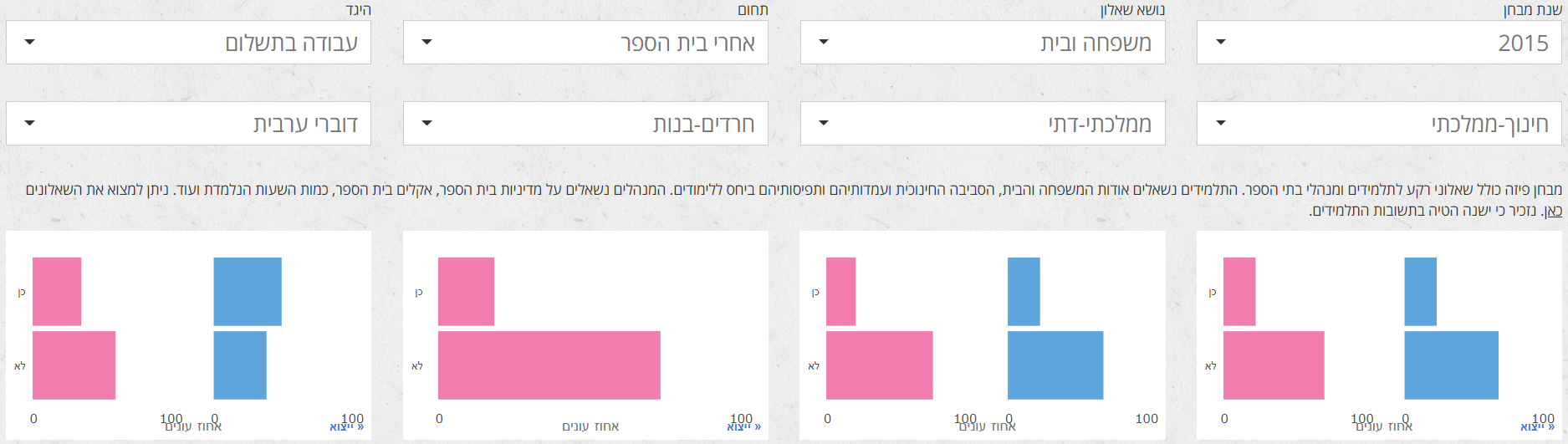
בתרשים המצ"ב מוצגות ישראל ו-3 מדינות נוספות להשוואה. ירדן – בשל השכנות הגיאוגרפית והדמיון לאוכלוסיה דוברת הערבית בישראל. ספרד – בשל הדימיון בתוצר לנפש. דנמרק – בשל הדימיון בגודל המדינה והאוכלוסיה.

מגמת הפערים לטובת הבנות במבחני הקריאה מתמידה גם בבחינה בינלאומית של הנתונים. מעניין לציין שגם כאן, במדינה עם ההישגים הנמוכים ביותר (ירדן) היו הפערים הגדולים ביותר לבין הבנות לבנים. (444 לעומת 372 בהתאמה).

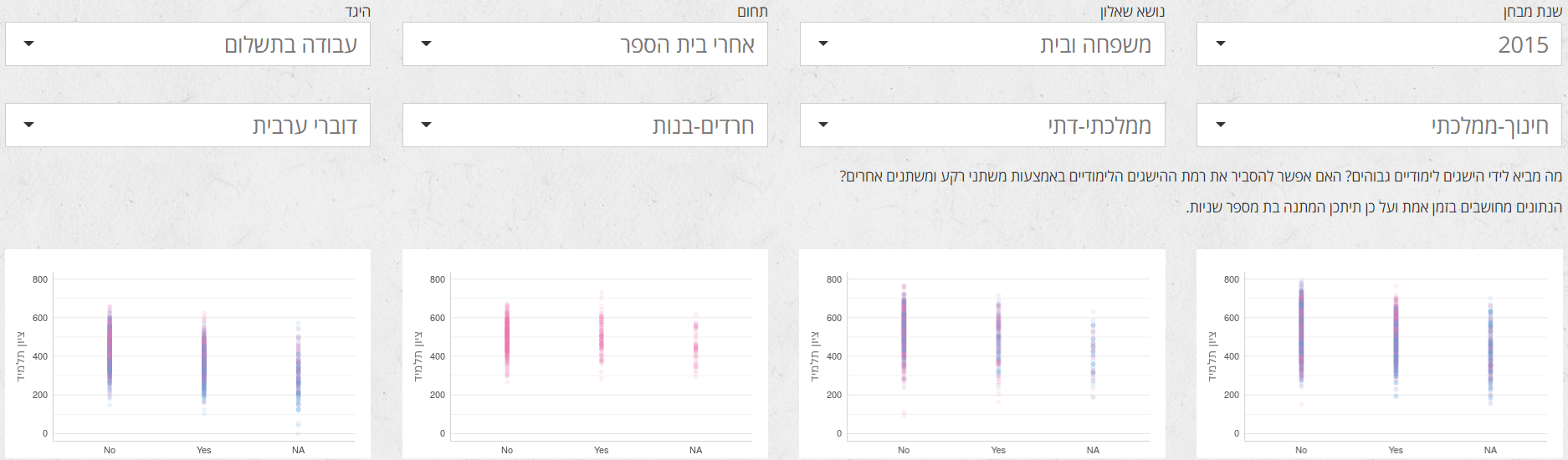


## **הסברים אפשריים לפערים**

הסבר אפשרי לחלק מהפער בהישגים בין דוברי ערבית ודוברי עברית, ניתן למצוא בתרשים הבא. מבחינה של דיווחי התלמידים עולה כי בקרב דוברי הערבית 34% מהבנות ו-48% מהבנים ילידי 1999 עבדו בשנת 2015. שיעור העובדים כנראה אף גבוה יותר כי לא כל הנבחנים ענו על השאלה. (וכך, רק 37% מהבנים ענו שלא עבדו). בקרב יתר המגזרים, רק כ-20% מהבנים והבנות ענו כי עבדו. במפתיע, לא ניכר שוני בתשובה בפילוח סוציואקונומי.

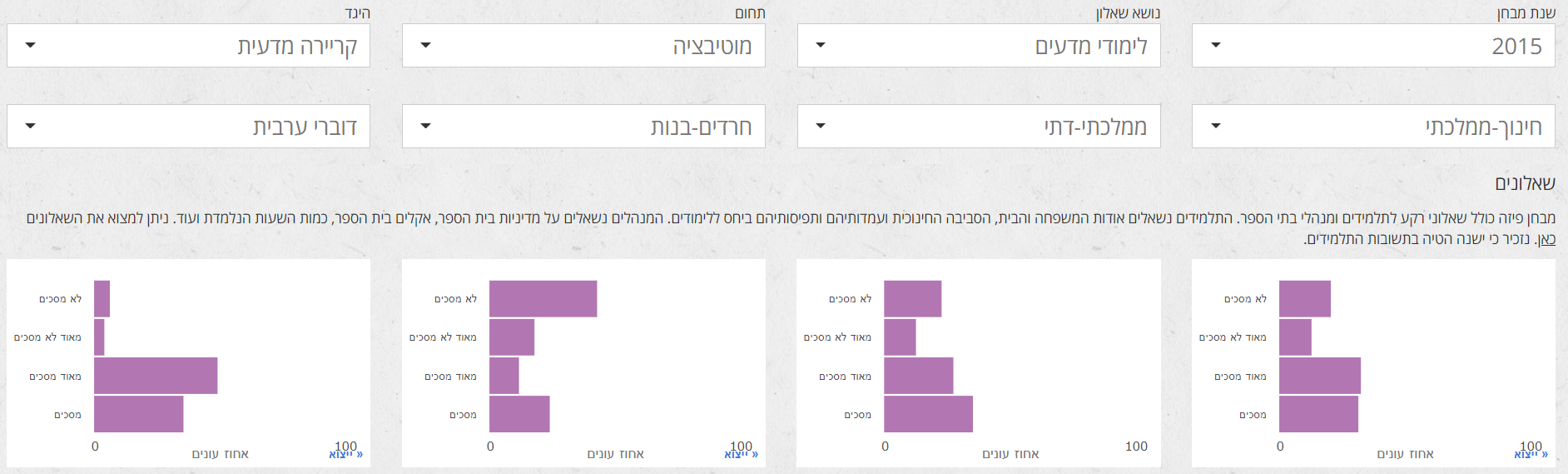


ואכן, בהמשך לכך ניתן לראות בתרשים הבא כי לתלמידים שענו כי **לא** עבדו אחרי בית הספר, ציון גבוה יותר מזה של אלה שענו כי עבדו. כדי להצביע על גודל הפערים נדרש ניתוח מדוייק יותר שלא התאפשר באתר "פיזה פתוח". (התמונה דומה ב-3 מקצועות הבחינה).

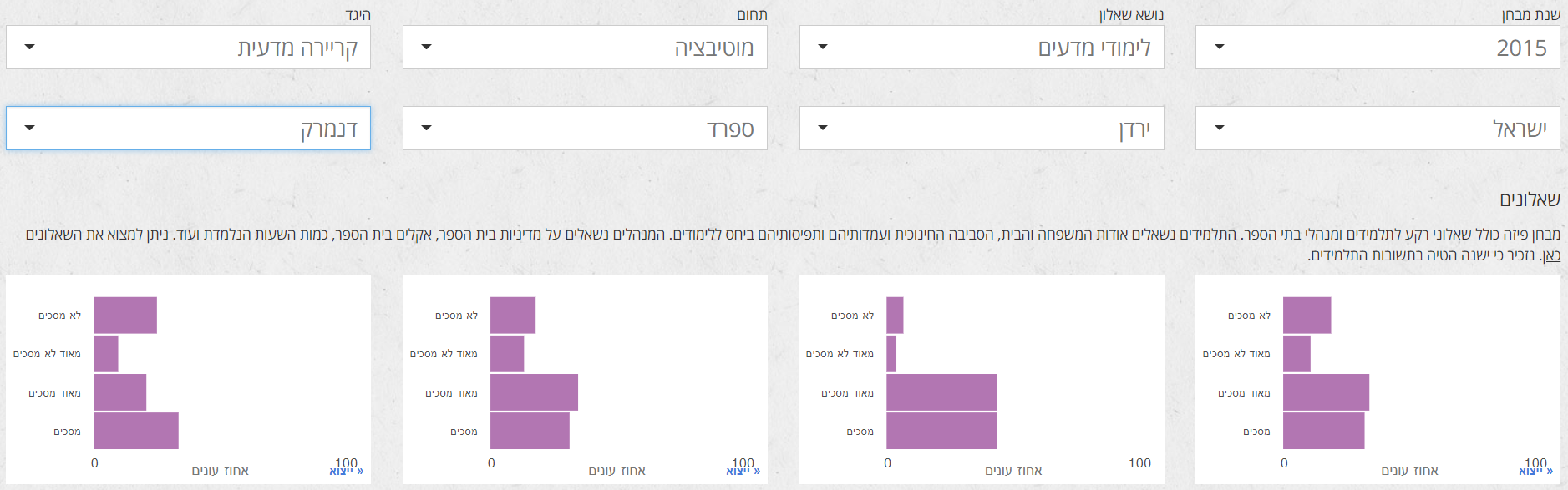


**לימודי מדעים יקדמו קריירה מדעית**?

כ-90% מדוברי הערבית ענו כי הם מעוניינים ללמוד מדעים כדי לקדם קריירה מדעית. לעומת 60%-70% בקרב דוברי העברית בחינוך הממלכתי והממלכתי-דתי ו-35% בקרב התלמידות החרדיות. לא נראה מתאם בין תשובה זו לבין ההישגים האישיים של התלמידים.

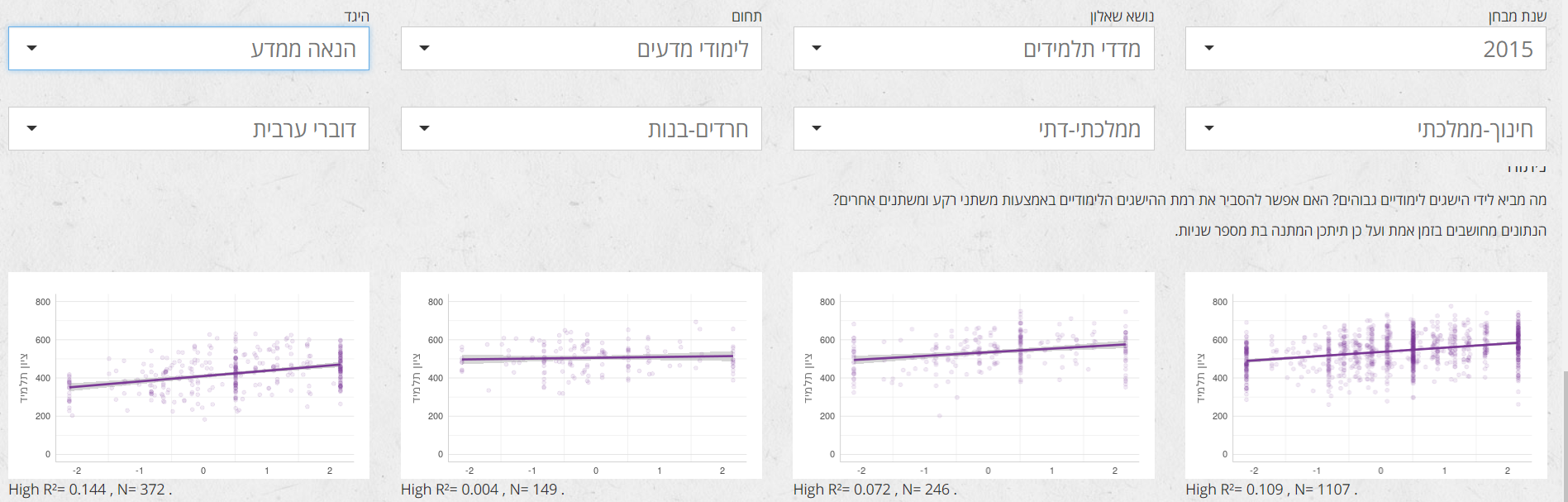


בהשוואה בינלאומית ניתן לראות כי במדגם (הלא ממש מייצג) שנבחר ככל שהציון הממוצע של המדינה יורד, המוטיבציה ללימודי מדעים ככלי לקריירה מדעית עולה. כך בירדן (380 במתמטיקה, 86% מסכימים שרוצים לקדם קריירה מדעית), ישראל (470, 75%), ספרד (486, 65%) ודנמרק (511, 53%).

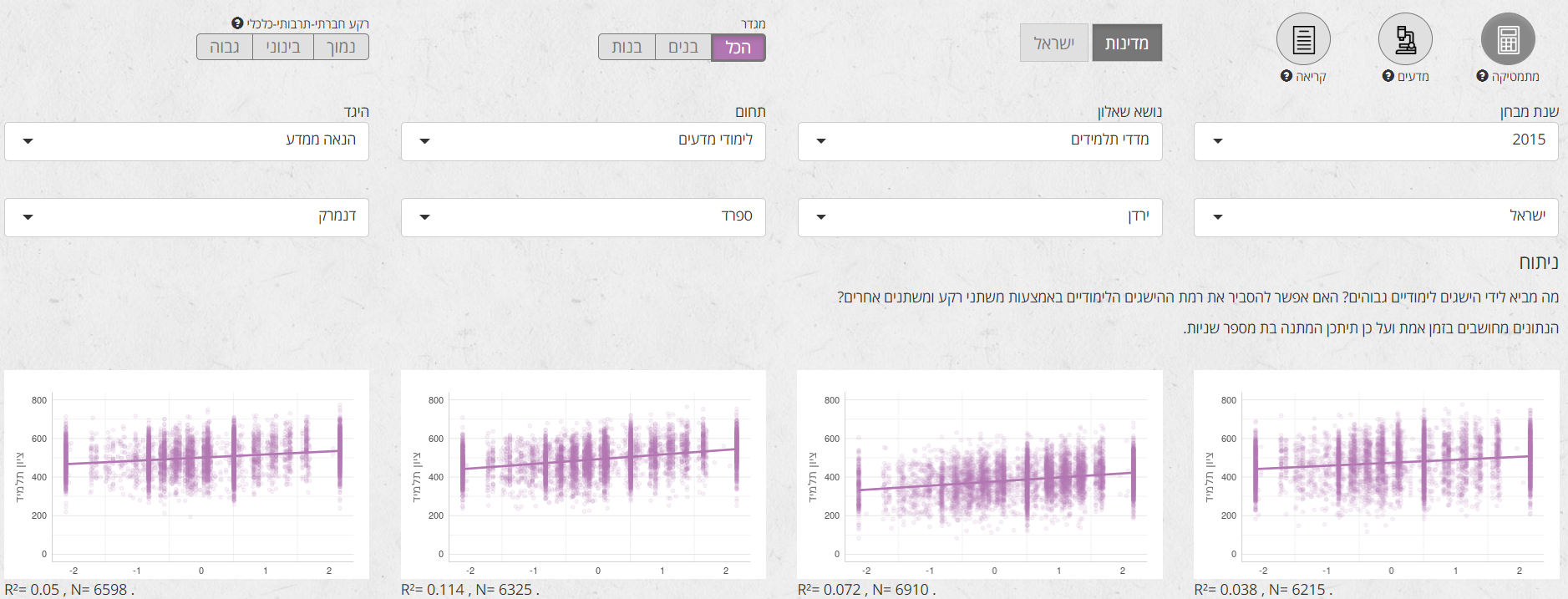


**יותר נהנים והציונים עולים**

בשאלה שמוצגת ב-2 התרשימים הבאים התבקשו התלמידים לדרג את מידת הנאתם ממדעים. ניתן לראות כי למעט התלמידות החרדיות, ההנאה המדווחת מתואמת חיובית ומשמעותית עם ממוצע הציונים (נכון ל-3 המקצועות).



גם בהשוואה בינלאומית ניתן לראות קשר חיובי בין ההנאה ממדעים וההישגים במבחנים בכל 4 מדינות המדגם –



## **רקע חברתי-תרבותי-כלכלי:**

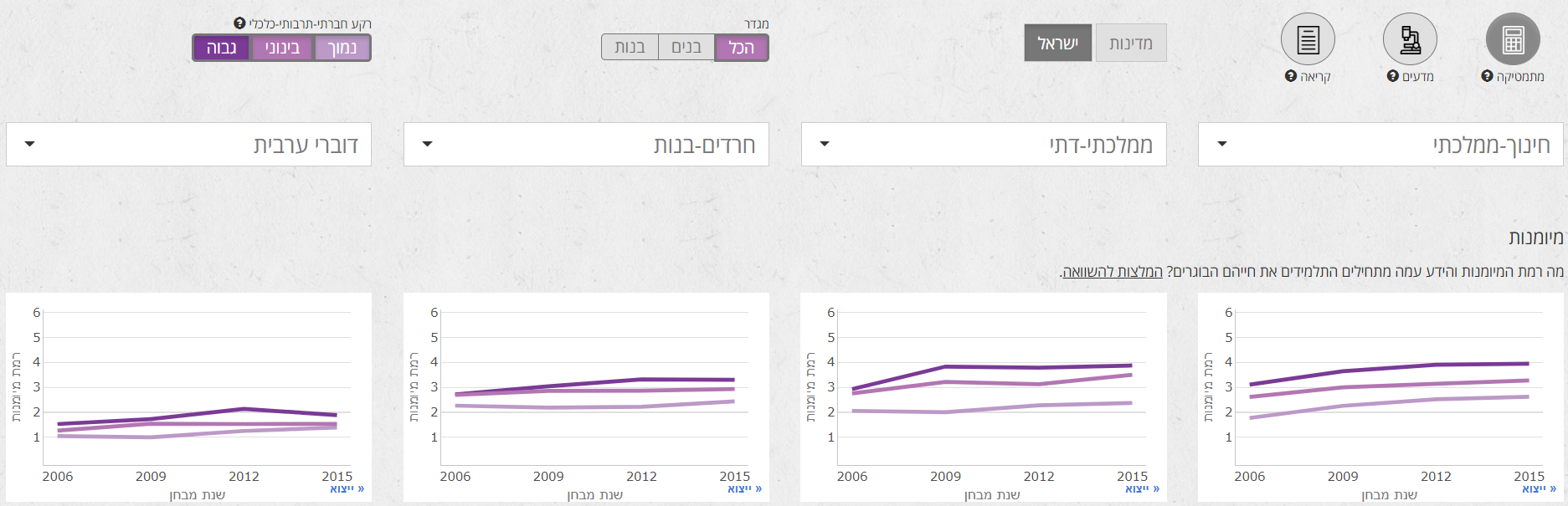
הרקע החברתי-תרבותי-כלכלי נבדק לפי מדד שפיתחו חוקריי פיזה, והוא מחושב על סמך דיווח עצמי בשאלונים לתלמידים. המידע שנאסף הוא בין השאר: עיסוקם של האם והאם, רמת ההשכלה של האב והאם, נגישות של משאבים חינוכיים,תרבותיים וכלכליים בבית וגורמים אחרים המעידים על המצב הכלכלי.

על פי דו"ח ראמ"ה למבחן פיזה 2015, בשלושת תחומי המבחן נמצאו פערים ניכרים בין תלמידים מרמות שונות של רקע חברתי-תרבותי-כלכלי – ככל שהרקע החברתי-תרבותי-כלכלי של התלמידים גבוה יותר כך הישגיהם גבוהים יותר. במחזור 2015 הפערים בהישגים בין תלמידים מרקע גבוה לתלמידים מרקע נמוך עומדים על 91 נקודות במדעים, 89 נקודות בקריאה ו-87 נקודות במתמטיקה.

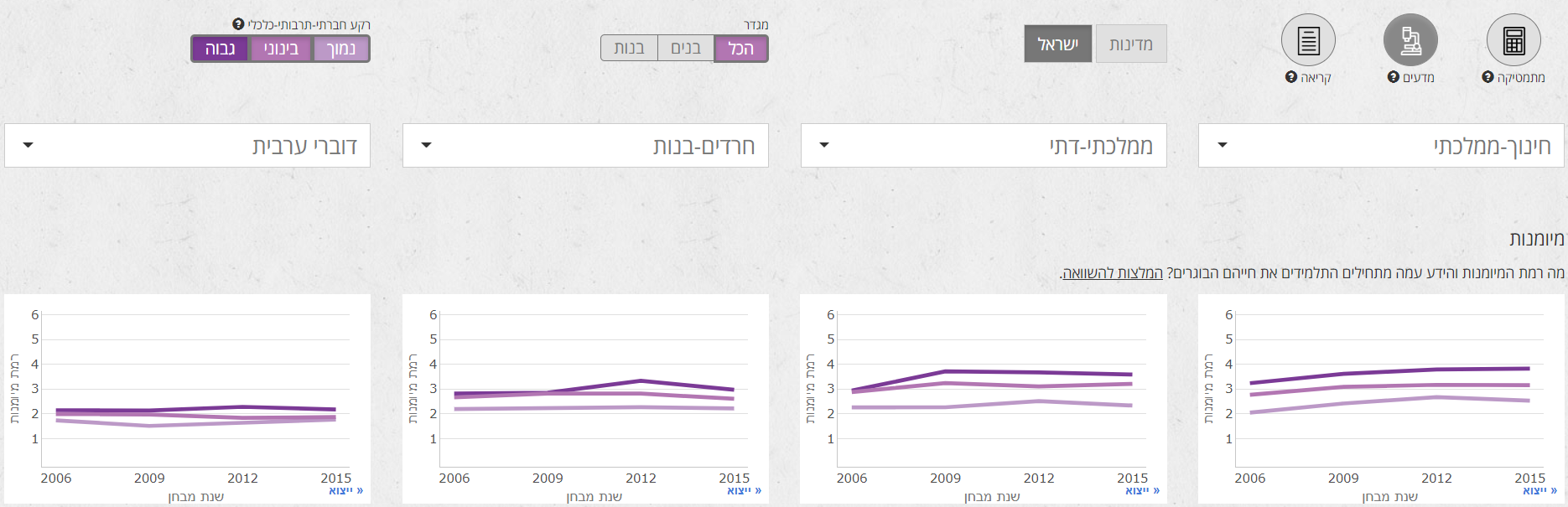
כפי שניתן לראות ב-3 התרשימים המצורפים, מגמה זו מתמידה ואף מתעצמת במידה מסויימת במהלך השנים 2006-2015.

בקרב תלמידים דוברי עברית הפערים בהישגים בין תלמידים מרקע גבוה לתלמידים מרקע נמוך דומים לפערים בקרב כלל האוכלוסייה - 94 נקודות במדעים, 87 נקודות בקריאה ו-84 נקודות במתמטיקה. בקרב תלמידים דוברי ערבית הפערים קטנים יותר ועומדים על 31 נקודות במדעים, 22 נקודות בקריאה ו-32 נקודות במתמטיקה.

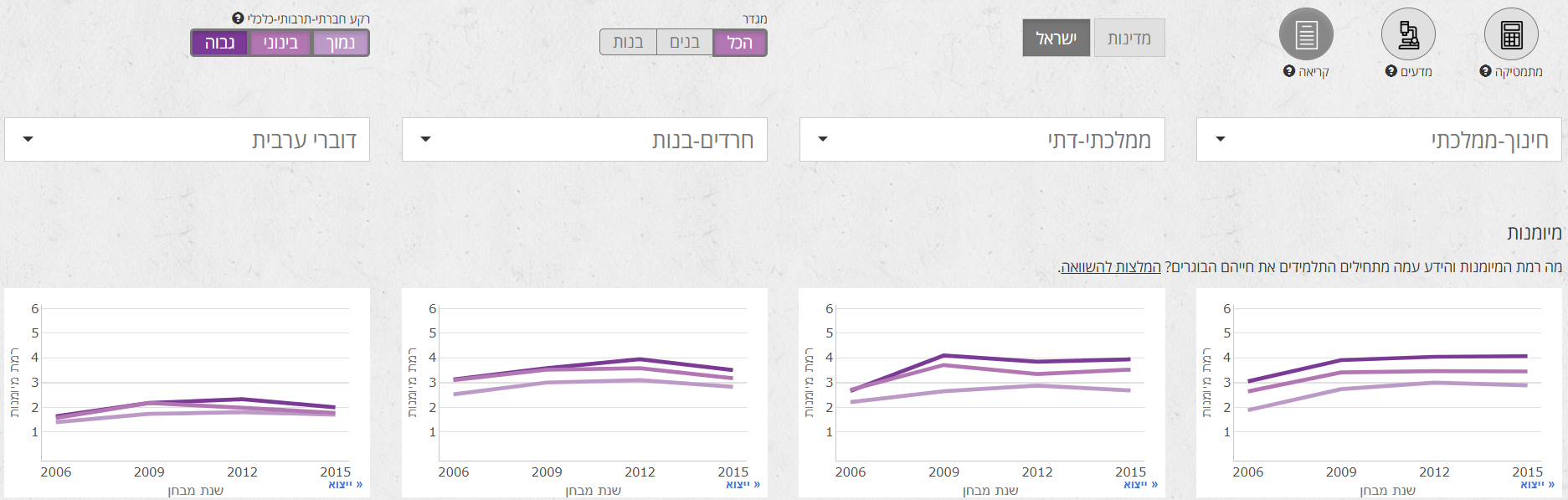
**מתמטיקה:**



**מדעים:**

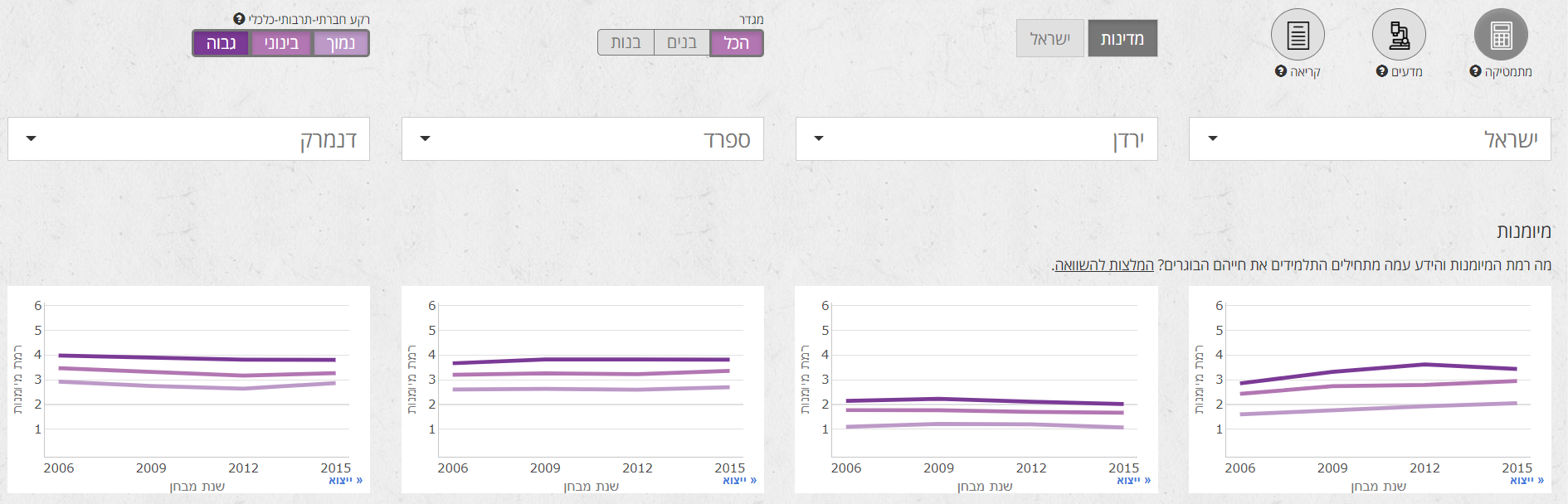


**קריאה:**



**בהשוואה בינלאומית:**

כפי שניתן לראות בתרשים הבא, הפערים בהישגים בין תלמידים מרקע תרבותי-חברתי-כלכלי מתמידים במדינות המדגם השונות אך בישראל הם הבולטים ביותר.



הנתונים המוצגים כאן מראים בצורה עקבית פערים גדולים בין מגזרים שונים בתוך ישראל ובין ישראל למדינות אחרות. מקובל להניח שפערים אלה הם תוצאה של פערים ביכולות התלמידים. בהמשך חיבור זה אטען כי יתכן שלפחות חלק מהשוני בהישגי התלמידים נובע מהיות מבחני פיזה מבחנים "חסרי השלכה" (Low-Stakes) לגבי התלמידים. בהיעדר השלכות, התלמידים עשויים להשקיע פחות מ"מאמץ מיטבי" בפתרון הבחינות. הירידה במאמץ עשוייה לגרום לירידה בהישגים בבחינות פיזה (ואחרות). אם הירידה במאמץ אינה אחידה לכל המגזרים ולכל המדינות, התמונה שמתקבלת מתוצאות מבחני פיזה היא תוצאה מוטה (Biased). בסיום המסמך אציג ניסוי שיכול לבחון במידת מה את ההטייה של מבחני השוואה סטנדרטיים חסרי השלכה, על תוצאות התלמידים.

# סקירת ספרות – מוטיבציה והישגים במבחנים חסרי השלכה

מוטיבציה היא מונח עם הגדרות רבות ושונות. בתחום החינוכי אחת ההגדרות המקובלות היא "תהליך שבו פעילות מוכוונת מטרה מתחילה ומתמידה" (Pintrich & Schunk, 1996). בהקשר של מבחני השוואה, הפעילות היא פתרון מבחן, והיא יכולה להיות מונעת על ידי מטרות שונות.

מכיוון שמוטיבציה היא מושג מופשט, לא ניתן למדוד אותה באופן ישיר, אלא יש לבחון אותה באמצעות דיווחים עצמיים או התנהגות. ניתן למשל לשאול את התלמידים "עד כמה התאמצתם?" או "כמה חשוב היה לכם להצליח במבחן?" וניתן לבחון עד כמה הצליחו התלמידים במבחן ביחס למבחנים אחרים שאנחנו יודעים שהיו חשובים להם. בשנים האחרונות התפתחה שיטה נוספת בפתרון מבחנים ממוחשבים בעזרתה ניתן לראות האם תהליך פתרון המבחן מצביע על מאמץ (Wise & Kong, 2005).

בסיס תיאורטי טוב להבנה של התנהגות התלמידים בפתרון מבחן הוא תיאוריית "ערך וציפייה" של מוטיבציה להישגים (Wigfield & Eccles, 2000; Wigfield, Tonks, & Klauda, 2009). על פי התיאוריה, המוטיבציה של התלמידים להתאמץ נקבעת בהתאם לאמונותיהם לגבי הסיכויים שלהם להצליח במשימה (ציפייה) ובהתאם לערך שהם מייחסים להצלחה במשימה. האמונה של התלמידים לגבי הצלחתם במבחן תושפע מהערכתם את הקושי בפתרון המבחן ואמונתם לגבי יכולותיהם. הערך שהם ייחסו להצלחה במבחן יהיה מורכב מהחשיבות שהם מייחסים לו, ממידת ההנאה שפתרון מבחן מסב להם, מהתועלת שהצלחה במבחן יכולה להעניק להם (כמו ציון או פרס כספי), מהמאמץ שידרש מהם כדי להצליח ומהויתורים שהצלחה במבחן תדרוש.

במבחנים שאין להם השלכה ישירה על התלמידים (באנגלית – Low-Stakes Test, להלן "מבחנים חסרי השלכה"), הערך שהם ייחסו להצלחה במבחן עשוי להיות נמוך. במקרה כזה, תיאוריית ה"ערך והציפייה" צופה שהם לא ישקיעו מאמץ מקסימלי בפתרון המבחן. מכך נובע כי יכולים להיות תלמידים שמשקיעים הרבה מאמץ בפתרון המבחן וכאלה שמשקיעים מאמץ קטן בלבד. אם השוני במאמץ לא מתחלק בצורה אקראית בקרב האוכלוסיה הוא יכול לגרום להטייה של תוצאות מבחני השוואה בין בתי ספר ובין מדינות ולפגוע ביכולתם להוות כלי טוב למחקר וקביעת מדיניות.

כדי להתמודד עם מאמץ נמוך של תלמידים במבחנים יש להתחיל במיפוי ובחינה של מאמץ התלמידים במבחנים. לשם כך, מרבית המחקרים עושים שימוש בשאלונים בהם נשאלים התלמידים לאחר המבחן כמה הם השקיעו במבחן. שאלונים כאלה יכולים להיות מורכבים ממספר שאלות מהן ניתן להסיק את רמת המאמץ והמוטיבציה או משאלות ישירות על רמת המוטיבציה והמאמץ (Butler & Adams, 2007; Sundre, 1999; Wolf & Smith, 1995). שיטה נוספת היא בחינה של הזמן אותו מקדישים התלמידים לכל שאלה והסקה מתוך כך האם התלמידים הקדישו מאמץ כן לפתרון השאלות (Wise & Kong, 2005).

## השפעה של רמת מאמץ נמוכה על תוצאות ההערכה של יכולת התלמידים

ההשפעה של מוטיבציה ומאמץ נמוכים על מהימנות תוצאות מבחני הערכה נמצאת במחלוקת. מרבית החוקרים מסכימים כי הם אכן יוצרים הטייה משמעותית בהערכת היכולת. סיכום של מחקרים שהשוו בין קבוצות עם מוטיבציה גבוהה ונמוכה, העלה כי ב-24 מ-25 ניסויים שנעשו, הקבוצה בעלת המוטיבציה הגבוהה יותר (מסיבות שונות) הגיעה להישגים טובים יותר במבחנים. הציון הממוצע של הקבוצה רבת המוטיבציה היה בממוצע גבוה ב-0.59 סטיות תקן מציון הקבוצה השניה (Wise & DeMars, 2005). מספר מחקרים אף מראים כי תוצאות מבחנים בהם הנבחנים דיווחו על מוטיבציה גבוהה יותר, היו מהימנות יותר בהערכת יכולות הנבחנים (Wise & DeMars, 2005; Zerpa & van Barneveld, 2015).

מצד שני, ישנם חוקרים שטוענים כי ברבים מהתרחישים בהם מתבצעת הערכת תלמידים, הבדלי המוטיבציה אינם בעלי משמעות. למשל, ישנם תלמידים, בעיקר בכיתות הצעירות, שמדווחים על מאמץ רב גם במבחנים חסרי השלכה (Eklöf, 2010). בנוסף עלתה הטענה כי על פי דיווחי מאמץ של תלמידים במבחני פיזה 2003, לא ניכר שוני רב בדיווחים בין המדינות השונות ולפיכך אין מקום לצפות כי יהיה שוני ניכר במאמץ בפועל או בהישגים (Butler & Adams, 2007). לדעתי, הטיעון לא משכנע מספיק.

## השפעת התמריצים על תוצאות המבחנים

מחקרים רבים ניסו להעריך את ההשפעה של תמריצים שונים על המוטיבציה, המאמץ וההישגים של תלמידים בניסיון להבין את התנהגות התלמידים במבחנים חסרי השלכה. תמריצים יכולים להיות כספיים וגם לא כספיים. גם ציון שנחשב לתעודה יכול להיחשב כתמריץ, אבל במקרה כזה לא נקרא למבחן "חסר השלכה" ולסוג זה של מבחנים אתייחס בפרק נפרד.

מרבית הניסויים שנעשו בתחום העלו כי לתמריצים כספיים ולא-כספיים יש בחלק גדול מהמקרים השפעה חיובית על המאמץ והביצועים. יוצאי דופן בתחום התמריצים הכספיים הם תמריצים קטנים, שעלולים דווקא לפגוע במוטיבציה ומקרים בהם התגמול ניתן בפער זמן ניכר מהניסוי עצמו, אז עלולה להתבטל השפעת התמריץ (Gneezy & Rustichini, 2000; Levitt, List, Neckermann, & Sadoff, 2013).

בניסויים שבהם הודיעו לתלמידים רגע לפני המבחן על תמריץ להצליח בו, נמצא כי בתרחיש בו הפרס חולק לתלמידים לפני המבחן בצורה מותנית, כל התלמידים שיפרו את הישגיהם ביחס לקבוצת הביקורת ב-0.12-0.22 סטיות תקן. בתרחיש בו הפרס חולק לתלמידים אחרי המבחן, נמצא שיפור מובהק במרבית אך לא כל הניסויים. גיל, נושא המבחן ומגדר התלמידים נמצאו כגורמים שמשפיעים על מידה ההשפעה של התמריץ. השתייכות אתנית ויכולת (כפי שנמדדה במבחנים אחרים) לא השפיעו (Braun, Kirsch, & Yamamoto, 2011; Levitt et al., 2013).

למרות שמרבית המחקרים מזהים השפעה של תמריצים כספיים, ישנם מחקרים בהם לא נמצאה השפעה כזו, או שהיא נמצאה רק בחלק מקבוצות הגיל (Baumert & Demmrich, 2001; Eisenkopf, 2011; H. F. O'Neil, Abedi, Miyoshi, & Mastergeorge, 2005; J. O'Neil Harold F, Sugrue, & Baker, 1995). הסבר אפשרי לכך הוא שבמחקרים בהם התמריצים לא השפיעו, זה היה בגלל שהם חולקו לתלמידים במרחק זמן רב מהמבחן (Levitt et al., 2013).

במספר מחקרים נבדקה גם השפעתם של תמריצים לא כספיים. תמריצים כאלה יכולים להיות למשל דירוג בכיתה, משוב של המורה, גביע או תשורה סמלית אחרת. מרבית המחקרים מצאו כי תמריצים כאלה יכולים להשפיע על המוטיבציה וההישגים בתרחישים מסויימים. כך למשל, דירוג התלמידים בכיתה וגביע סמלי נמצאו כמשפיעים על המוטיבציה וההישגים, לעתים אף כמו פרס כספי של 10$-20$. לעומת זאת מתן ציון "אובייקטיבי" בין A-F ללא השלכה נוספת על התלמיד, לא השפיע (Jalava, Joensen, & Pellas, 2015; Levitt et al., 2013).

## מבחנים בעלי השלכה (High-stakes) מול מבחנים חסרי השלכה (Low-stakes)

מכיון שהתוצאה במבחני השוואה סטנדרטיים אינה חשובה לתלמידים אנחנו מצפים לראות פער בין המאמץ והתוצאות במבחנים אלה למאמץ ולתוצאות במבחנים בעלי השפעה משמעותית על התלמידים כמו מבחני בגרות, מבחנים פסיכומטריים, מבחני נהיגה ואפילו מבחני סוף שנה בבית הספר. פער מאמץ זה מושפע הן ממאפייני הבחינה והן ממאפייני הנבחנים.

מסקירה של מספר מחקרים, ניתן לזהות כי ישנם מספר מאפיינים של הבחינה שמשפיעים לחיוב על המאמץ במבחנים חסרי השלכה ובכך מקטינים את פער המאמץ בין שני סוגי המבחנים וישנים מספר מאפיינים שפועלים בכיון ההפוך ומגדילים את פער המאמץ בין המבחנים. שאלות רב-ברירתיות (אמריקאיות) ושילוב גרפיקה בתוך השאלה, מגבירים את המאמץ וההישגים של התלמידים במבחנים חסרי השלכה. לעומת זאת, שאלות פתוחות, שאלות שכוללות הרבה טקסט ושאלות שמגיעות בשלב מאוחר של המבחן, יזכו לפחות מאמץ באופן יחסי. (DeMars, 2000; Sundre, 1999; Wise, Pastor, & Kong, 2009; Wolf, Smith, & Birnbaum, 1995)

השפעת מאפייני הנבחן על פער המאמץ בין שני סוגי המבחנים קצת יותר קשה למחקר, כי היא דורשת מעקב אחרי המאפיינים היחודיים של כל נבחן (או קבוצת נבחנים) והישגיה בשני סוגי המבחנים. במחקר שנעשה תוך השוואה בין הישגים במבחן GRE ובין מבחן וולנטרי שהוצע לנבחנים למלא אחריו, נמצאו מספר מאפיינים של הנבחנים שהשפיעו על ההישגים. הציון של גברים ירד משמעותית יותר מנשים במעבר ממבחן ה-GRE למבחן הוולנטרי. גם הציון של לבנים ירד משמעותית יותר מהציון של אסיאתיים, שחורים והיספניים (Attali, Neeman, & Schlosser, 2011). ממצאים אלה יכולים להיות מוסברים על ידי מחקרים רבים שהציגו הבדלים בגישה כלפי מבחנים בין גברים ונשים ובין קבוצות אוכלוסיה שונות (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; DeMars, Bashkov, & Socha, 2013; Eklöf, 2007)

## האם תוצאות מבחני השוואה סטנדרטיים מוטות כתוצאה מהבדלי מאמץ תלויי תרבות?

השפעת שייכות אתנית על ירידת ההישגים הוערכה במחקר רחב היקף בארצות הברית ונמצאה משמעותית (Attali et al., 2011). הבנה של השקעת המאמץ של קבוצות שונות בחברה במבחני הערכה סטנדרטיים תסייע לנו לעשות שימוש טוב יותר בתוצאותיהם במחקר וקביעת מדיניות. יתר על כן, ממצאי מבחני הערכה בינלאומיים נמצאים לעתים תחת ביקורת לפיה ממצאיהם מוטים בשל שוני במאמץ הנבחנים במדינות השונות. הבנה של השוני או היעדרו בקרב קבוצות שונות בחברה הישראלית, יכולה לרמוז על תקפות ביקורת זו ברמה הבינלאומית.

העמדות לגבי קיום מתאם בין יכולת התלמידים ובין המאמץ שמושקע על ידיהם במבחנים חסרי השלכות אינן חד משמעיות. מכיון שמאמץ מדווח ומאמץ בפועל נמצאו מתואמים, אם אין מתאם בין מאמץ מדווח ובין יכולת התלמידים, ניתן לשפר את דיוק תוצאות המבחנים על ידי השמטה מהחישוב של התלמידים שדיווחו על מאמץ נמוך במבחן(Wise & DeMars, 2005; Zerpa & van Barneveld, 2015). אם מתאם כזה קיים, השמטה של התלמידים שדיווחו על מאמץ נמוך, תגרום להטייה של התוצאות.

כדי לבחון את שאלות אלה בכוונתי לערוך ניסוי שדה בקרב תלמידי כיתה ח' מבתי ספר בזרמי חינוך שונים בישראל: יהודי ממלכתי (4 בתי ספר), יהודי ממלכתי-דתי (2 בתי ספר), ערבי מוסלמי (2), ערבי נוצרי (2) ודרוזי (2). התלמידים ייבחנו בשני מבחנים דומים. אחד יהיה בעל השלכה על הציון הסופי השנתי שלהם והשני יהיה ללא השלכה. לאחר כל מבחן, התלמידים יימלאו שאלון מאמץ. השוואה בין תוצאות זרמי החינוך השונים תאפשר לנו להבין את התנהגות התלמידים השונים במבחני השוואה סטנדרטיים ולבחון את המתאם בין יכולת ובין מאמץ במבחנים אלה.

# סיכום

מחקרים רבים העוסקים במבחנים חסרי השלכה מראים כי מבחנים אלה מביאים להישגים נמוכים יותר של הנבחנים. ירידה זו בהישגים, במידה והיא אינה אחידה או אינה מפוזרת בצורה אחידה, תביא להטייה של ממצאי מבחני פיזה או מבחני השוואה סטנדרטיים אחרים. יתכן לפיכך שחלק מהפערים שהוצגו בפתיחת מאמר זה ובסיכומים קודמים של מבחני פיזה בישראל, נובעים מפערים במאמץ במבחנים חסרי השלכה ולא מפערים בהישגים.

ניתן לבחון שאלה זו על ידי ניסוי שדה בבתי ספר שאת ראשיתו כבר החלתי במסגרת עבודת הדוקטורט שלי.

## References

Attali, Y., Neeman, Z., & Schlosser, A. (2011). Rise to the challenge or not give a damn: Differential performance in high vs. low stakes tests.

Baumert, J., & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education, 16*(3), 441-462.

Braun, H., Kirsch, I., & Yamamoto, K. (2011). An experimental study of the effects of monetary incentives on performance on the 12th-grade NAEP reading assessment. *Teachers College Record, 113*(11), 2309-2344.

Butler, J., & Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement, 8*(3), 279.

Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology, 82*(2), 300.

DeMars, C. E. (2000). Test stakes and item format interactions. *Applied Measurement in Education, 13*(1), 55-77.

DeMars, C. E., Bashkov, B. M., & Socha, A. B. (2013). The role of gender in test-taking motivation under low-stakes conditions. *Research & Practice in Assessment, 8*

Eisenkopf, G. (2011). Paying for better test scores. *Education Economics, 19*(4), 329-339.

Eklöf, H. (2007). Gender differences in test-taking motivation on low-stakes tests. *The Annual Meeting of the American Educational Research Association (AERA), Chicago, IL, April 2007,*

Eklöf, H. (2010). Skill and will: Test‐taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice, 17*(4), 345-356.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics,* , 791-810.

Jalava, N., Joensen, J. S., & Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. *Journal of Economic Behavior & Organization, 115*, 161-196.

Levitt, S., List, J., Neckermann, S., & Sadoff, S. (2013). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *Framed Field Experiments, 379*

O'Neil, H. F., Abedi, J., Miyoshi, J., & Mastergeorge, A. (2005). Monetary incentives for low-stakes tests. *Educational Assessment, 10*(3), 185-208.

O'Neil, J., Harold F, Sugrue, B., & Baker, E. L. (1995). Effects of motivational interventions on the national assessment of educational progress mathematics performance. *Educational Assessment, 3*(2), 135-157.

Pintrich, P. R., & Schunk, D. H. (1996). *Motivation in education: Theory, research, and applications.* Prentice Hall.

Sundre, D. L. (1999). Does examinee motivation moderate the relationship between test consequences and test performance? *James Madison University,*

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68-81.

Wigfield, A., Tonks, S., & Klauda, S. L. (2009). Expectancy-value theory. *Handbook of Motivation at School,* , 55-75.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1-17.

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education, 18*(2), 163-183.

Wise, S. L., Pastor, D. A., & Kong, X. J. (2009). Correlates of rapid-guessing behavior in low-stakes testing: Implications for test development and measurement practice. *Applied Measurement in Education, 22*(2), 185-205.

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education, 8*(3), 227-242.

Wolf, L. F., Smith, J. K., & Birnbaum, M. E. (1995). Consequence of performance, test, motivation, and mentally taxing items. *Applied Measurement in Education, 8*(4), 341-351.

Zerpa, C., & van Barneveld, C. (2015). The effect of removing examinees with low motivation on item response data calibration. *International Journal of Applied Psychology, 5*(2), 33-44.

1. הנתונים הופקו ממערכת "פיזה פתוח" בכתובת <http://opisa.org/#home> בתאריך 10/6/2018. [↑](#footnote-ref-1)