מבחני הבגרות והפסיכומטרי במתמטיקה ובאנגלית והקשר ביניהם לאורך זמן[1]

שלמה יצחקי, טאינה פודלוב-טרחטנברג ואביאל קרנצלר

מזה שנים רבות מתקיימות בארץ הן בחינות הבגרות והן הבחינות הפסיכומטריות. הצורך להיבחן בשתי בחינות מהווה נטל הן על התלמידים הנדרשים להתכונן לשני מבחנים שמתכונותיהם שונות, והן על המשק הישראלי. מחקר זה מהווה מחקר מקדים לבחינת השאלה של עלות-תועלת בקיום שתי הבחינות – הבגרות והפסיכומטרי – באותם תחומים, ובודק באיזו מידה המבחנים בודקים את אותו תחום יכולות.

החידוש המתודולוגי בעבודה זו הוא בשיטת המחקר, שהיא שיטה אי פרמטרית ולא תלוית התפלגות, ומבוססת על העובדה שבתחום מדידה בחינוך אין לנו יכולת מדידה ישירה כמו במשתנה כמותי רגיל אלא אנו שואלים שאלות ובודקים מי ענה עליהן. כתוצאה מכך הציונים תלויים בהתפלגות הקושי של השאלות.

שיטת המחקר שלנו מתבססת על הטענה הבאה: אנו בודקים את המתאם המתקבל בין הציונים של התלמידים במבחן הבגרות לבין הציונים במבחן הפסיכומטרי, ואם התוצאה המתקבלת היא שהמתאם הוא מונוטוני לכל אורך התחום של הציונים, אז נסיק שאחד המבחנים מיותר. אם לעומת זאת המתאם משנה את סימנו לאורך תחום הציונים אזי המשמעות היא שהבחינות בודקות תחומי ידע שונים או שהתכונות הנדרשות להצלחה בבחינה אחת שונות מהתכונות הנדרשות להצלחה בבחינה האחרת.

התשובה המתקבלת ממחקר זה היא שהמבחנים בודקים את אותו תחום יכולות, אולם קיימת בהם טעות אקראית גדולה. בדיקת התרומות של שתי הבחינות לציון הכולל של כל תלמיד מחזקות את ההשערה שהתרומה של כפל הבחינות אינה רבה. המסקנה דלעיל מתקיימת ביתר שאת לגבי מקצוע האנגלית שבו המתאם בין הציון בבגרות לבין הציון בפסיכומטרי גבוה יותר מאשר במתמטיקה.

כדאי להדגיש שאין במחקר כדי להכריע על איזה מבחן כדאי לוותר, אלא רק שמבחני בגרות ופסיכומטרי במקצועות המקבילים (אנגלית ומתמטיקה) אינם בודקים תחומים שונים שיש לנו יכולת להבחין ביניהם.

מבוא

מזה שנים רבות מתקיימות בארץ הן בחינות הבגרות והן הבחינות הפסיכומטריות. לבחינות יש מטרות שונות ועל כן מתכונתן שונה. בשנים האחרונות שתי הבחינות משמשות גם לצורכי קבלה ללימודים אקדמיים באוניברסיטאות. מערכת בחינות כפולה הנערכות במתכונת שונה משמעותה

1 אנחנו מודים לדמיטרי רומנוב ולשלושה קוראים אלמונים על הערותיהם לגרסות קודמות של המאמר שסייעו רבות בשיפור ההצגה של המאמר.

עומס כפול על האוכלוסייה הנבחנת. על כן, כדאי לבדוק את המידה שבה איחוד של שתי הבחינות למתכונת אחת יפחית את הנטל המוטל על האוכלוסייה הנבחנת. כלומר, מטרתנו בעבודה זו היא לבצע את השלב המקדים בבדיקת עלות לעומת תועלת לבחינה כדי לבדוק האם יש בכפילות זו משום נטל עודף על המשק.

כדי לבצע ניתוח עלות-תועלת יש לבדוק גם את התועלות מקיום המערכת הכפולה וגם את העלויות למשק. מדידת העלויות למשק קלה יחסית, בהנחה שקיימת מערכת נתונים על המועסקים בתחום והערכות של הזמן המושקע בהתכוננות לבחינות. קשה יותר למדוד את ערך האינפורמציה הטמונה בקיומה של מערכת מדידה כפולה,[2] זאת מאחר שאין דרך קלה להבחין בתרומה של קיום אינפורמציה בתחום אבסטרקטי שנקרא ידע.

במאמר זה אנו מתרכזים במדידת תוספת האינפורמציה על יכולות התלמידים הנבחנים במקצועות השונים שמתקבלת ממערכת הבחינות הכפולה, שהיא כאמור השלב המקדים והקשה יותר בכל ניסיון לביצוע מבחני עלות-תועלת ברמת המשק.

שיטת המחקר שלנו מתבססת על הטענה הבאה: אנו בודקים את המתאם המתקבל בין הציונים של התלמידים במבחן הבגרות לבין הציונים שלהם במבחן הפסיכומטרי, ואם התוצאה המתקבלת היא שהמתאם הוא מונוטוני לכל אורך התחום של הציונים נסיק שאחד המבחנים מיותר. אם לעומת זאת מדרג הציונים הוא שונה בתכלית המשמעות היא שהבחינות בודקות תחומי ידע שונים, או שהתכונות הנדרשות להצלחה בבחינה אחת שונות מהתכונות הנדרשות להצלחה בבחינה האחרת.

בדיקת הקשר בין ציונים של שתי בחינות נתקלת בקושי שאינו קיים בבדיקת המתאם בין משתנים כמותיים, כמו למשל המתאם בין המשקל של האדם לגובהו. זאת מאחר שאין אפשרות למדוד ידע בצורה כמותית אלא רק בצורה עקיפה, כגון מניית מספר השאלות שהנבחן הצליח לענות עליהן תשובה נכונה, או לחילופין קריאה והערכה של טיב התשובות שסיפק הנבחן. מספר השאלות שהנבחן הצליח לענות עליהן תלוי גם בהתפלגות הקושי של השאלות בבחינות וגם בעיבודים שנעשים לציונים על ידי הגופים המנהלים את המבחנים. כדי להתגבר על קושי זה אנו נגביל עצמנו לבדיקת המידה שבה המתאם בין הציונים הוא מונוטוני. במונח "מתאם מונוטוני" הכוונה היא שלאורך כל תחום הציונים של הבגרות (או הפסיכומטרי) עלייה בציון הבגרות (או הפסיכומטרי) מביאה לעלייה בממוצע בציון הפסיכומטרי (או הבגרות) בבחינה המקבילה. לצורך זה אנחנו משתמשים בכלי סטטיסטי חדש יחסית המיועד לבדוק קשרים מונוטוניים בתחומים שבהם אין לנו מדידה ישירה של המשתנה.[3]

מבנה העבודה הוא כדלהלן: הסעיף הראשון מכיל תיאור קצר של שאלת המחקר. הסעיף השני מתייחס למשמעויות שיש לייחס למבחן יכולות, וזאת מאחר שאין להתייחס לציון הניתן כפי שמתייחסים למשתנה כמותי כמו משקל או גובה. הסעיף השלישי עוסק בהצגת שיטת המחקר הבאה לבדוק את מונוטוניות הקשר בתחומי האנגלית והמתמטיקה. הסעיף הרביעי מציג את הנתונים המשמשים ואת שיטת הניתוח ואילו הסעיף החמישי מציג את הממצאים האמפיריים. התוצאות של הממצאים האמפיריים הן שקיים קשר מונוטוני בין הצלחה בבגרות לבין הצלחה בפסיכומטרי, וקשר זה חזק יותר באנגלית מאשר במתמטיקה. קשר מונוטוני זה מתחזק במשך השנים. במקביל, נמצא שקיים "רעש" אקראי בתוצאות המבחנים השונים, רעש הגורם להקטנת אמינות המבחנים כמנבאים

2 במחקר שנערך בשנת 2007 על ידי המרכז הארצי לבחינות והערכה (https://www.nite.org.il/files/reports/342.pdf) נמצא שהמערכת הכפולה משפרת את תוקף הניבוי להצלחה בשנה א באוניברסיטה. בסעיף 6 נתייחס לנקודה זו.

3 כל השיטות המשמשות במאמר זה נדונות ב-Yitzhaki and Schechtman (2013).

יכולות בתחום. במקרה כזה יכולה לעלות הטענה שקיום שני מבחנים מטרתו להקטין את הרעש האקראי, ושיפור זה מחייב קיומם של שני מבחנים בכל תחום. סעיף ו בא לענות על שאלה זו שהתעוררה תוך כדי המחקר, והוא בודק בעזרת שימוש במדד ג'יני את השיפור המתקבל משימוש בשני ציונים לעומת שימוש בציון אחד. כלומר הבדיקה נעשית על תרומת השימוש בשני מבחנים לשיפור העמידות של דירוג הציונים. המסקנה שמתקבלת מסעיף זה היא שתוספת האינפורמציה משימוש בשני מבחנים קטנה יותר באנגלית מאשר במתמטיקה.

א. תיאור קצר של שאלת המחקר

בחינות הבגרות נערכות על ידי משרד החינוך במתכונת ארצית, במקצועות ליבה שונים, ומיועדות לדרג תלמידים על פי רמת ההתמחות והיכולות בתחומים השונים. לעומת זאת, על פי הגדרת המרכז הארצי לבחינות והערכה – "הבחינה הפסיכומטרית היא כלי לחיזוי סיכויי ההצלחה בלימודים במוסדות להשכלה גבוהה.[4] היא משמשת את מוסדות הלימוד למיון מועמדים לחוגים השונים. הבחינה מאפשרת לדרג את כל המועמדים על סולם הערכה אחיד. הבחינה מורכבת משלושה תחומים: חשיבה מילולית, חשיבה כמותית ואנגלית. בנוסף, מדווח הציון הכללי. בכל אחד מן התחומים סולם הציונים נע בין 50–150 נקודות. הציון הכללי בבחינה מדווח על סולם הנע בין 200–800 נקודות. הציון הכללי מחושב על סמך ציוני הנבחן בשלושת התחומים המרכיבים את הבחינה ומתבסס על ממוצעים משוקללים שבהם הציונים בתחומים השונים מקבלים משקלות שונים.

הצורך להיבחן בשתי בחינות מהווה נטל הן על התלמידים הנדרשים להתכונן לשני מבחנים שמתכונותיהם שונות והן על המשק הישראלי, זאת מאחר שקיימת תעשייה לא קטנה שהייתה נעלמת לו נעשתה רק בחינה אחת. מאחר שבחינת הבגרות קדמה למבחן הפסיכומטרי, השאלה הנשאלת היא האם המבחן הפסיכומטרי בודק תחום יכולות אחר השונה במהותו ממבחני הבגרות.

הטענה להצדקת קיום המבחן הפסיכומטרי היא שכושר החיזוי של הבחינה לגבי ההצלחה בלימודים אקדמיים הוא טוב, והצירוף של ציון הבגרות יחד עם ציון הפסיכומטרי הוא בעל כושר ניבוי טוב יותר מאשר כל ציון בנפרד. הטענות על כושר ניבוי אינן נבדקות במאמר זה, זאת מאחר שהתלמידים מתקבלים ללימודים בתחומי ידע שונים, אולם המתודולוגיה המפותחת בו תוכל לשמש בעתיד לבדיקת הטענה בדבר כושר ניבוי.

מטרת מאמר זה היא לבחון טענה צנועה יותר והיא באיזו מידה המבחן הפסיכומטרי בודק תחום יכולות אחר מזה שבוחנת בחינת הבגרות באותו התחום. כלומר, לא נעסוק במאמר זה בשאלת עלות-תועלת כתוצאה מהכפילות למשק וגם לא נעסוק בשאלת כושר הניבוי, אלא נתרכז בשאלה האם שני המבחנים, הבגרות והפסיכומטרי, מהווים מבחן לאותו תחום יכולות. לדוגמה, אם מבחן הבגרות במתמטיקה בודק בדיוק את היכולת המתמטית כפי שהיא נבדקת בחלק הכמותי בבחינה הפסיכומטרית אזי נטען ששני המבחנים בודקים את אותו תחום יכולות. אם לעומת זאת תוצאות הבגרות במתמטיקה שונות באופן מובהק מהתוצאות של המבחן הפסיכומטרי המקביל לו, נאמר ששני המבחנים בודקים תחומי יכולת שונים.

החידוש המתודולוגי בעבודה זו הוא בשיטת המחקר: באמצעות שיטה זו (המוצגת במאמר) אפשר לקבוע האם שני מבחנים בודקים את אותו תחום יכולות. במילים אחרות – נוכל לקבוע שהידע

4 https://www.nite.org.il/index.php/he/tests/psychometric/psychometric-about.htm

והיכולות שמאפשרים הצלחה במבחן האחד זהים לידע וליכולות הנדרשים להצלחה במבחן האחר. בסעיף הבא נבהיר את המשמעות המתמטית שיש לייחס לתוצאות של מבחן, ואילו בסעיף שלאחריו נתאר את הכלי שישמש אותנו.

ב. המשמעות שיש לייחס למבחן יכולות[5]

לצורך הדיון נניח מספר הנחות מפשטות שנסיר אותן לאחר מכן. בשלב הראשון נניח שאנו מסוגלים להגדיר תחומי יכולות של אנשים, כלומר שהיכולת באנגלית שונה ודורשת כישורים ומיומנויות אחרים מאשר היכולת במתמטיקה. נסמן את המשתנה יכולת בתחום שאנו רוצים לבחון באות a .

הנחה שנייה היא שיכולת (או מיומנות) היא משתנה רציף, כך שאם נילי טובה יותר מגיא אזי ההסתברות שהיא תענה נכון על שאלה בתחום המבחן גבוהה יותר מההסתברות שגיא יענה נכון. המשמעות של הנחה זו היא שאם יש שתי יכולות a_1 ו a_2 וידוע ש $a_2 > a_1$, כלומר שהיכולת של פרט 2 גדולה מהיכולת של פרט 1, ומוצגת שאלה בעלת דרגת קושי q אזי ההסתברות של בעל היכולת a_2 לענות נכון על השאלה גבוהה או שווה להסתברות של פרט עם יכולת a_1 לענות על השאלה.[6]

הנחה שלישית היא שאין אקראיות, כלומר לכל שאלה יש רמת יכולת קריטית, כך שלכל הנבחנים שיכולתם גבוהה מהיכולת הקריטית הנדרשת למתן תשובה נכונה לשאלה יתקיים שההסתברות שיענו נכון על השאלה היא אחת, וכל אלו שיכולתם נמוכה מהיכולת הקריטית ההסתברות לענות נכון על השאלה הנדונה היא אפס.[7]

בעזרת שלוש הנחות אלו אנחנו יכולים להגדיר מהו מבחן:

הגדרה: מבחן הוא אוסף של שאלות בתחום ידע מסוים, בעלות דרגות קושי שונות, אשר מיועדות לגלות את היכולות של הנבחנים.

נעבור עתה להבחין בין מדידה של גובה של קבוצת אנשים לבין מדידת הידע שלהם תחת ההנחות דלעיל. במדידת גובה יש משמעות להבדלים בסנטימטרים בין אנשים. במדידת ידע ההבדלים בתוצאות המבחן תלויים בהתפלגות הקושי של השאלות במבחן. למעשה ניתן לטעון שמבחן לבדיקת ידע בתחום מסוים כמוהו כמדידת גובה כאשר הנבחנים עומדים מאחורי מסך והבוחן שואל שאלות כגון למי יש הסנטימטר ה-150, למי יש הסנטימטר ה-160, וכל אלו שגבוהים מ-150 ס"מ או מ-160 ס"מ עונים בהתאם לגובהם, באם יש להם או אין להם הסנטימטר שבו מדובר. מכאן שברור שאין הבדל בין הציונים שמקבלים מי שהגובה שלהם הוא 151 ו-159 ויש הבדל בין הציונים שמקבלים מי שהגובה שלו הוא 159 ומי שגבהו הוא 161. מכאן שניתן לקבוע שתוצאות מבחן הינן כמו מדידת גובה כאשר האנשים ניצבים מאחורי מסך, ועל כן קיומן של שאלות בתחום ידע מסוים או אי קיומן ישפיע על ההבדלים בציונים שמקבלים הנבחנים. כדאי לציין שאין ביכולת המבחן לשנות את הסדר של הנבחנים אלא רק להגדיל או להקטין את המרחקים בין ציוני הנבחנים.

המשמעות המתמטית היא שהפעולה של המבחן כמוה כיצירת פונקציה מתמטית כך שהציון נקבע על ידי היכולת של הנבחן. בונה המבחן, על ידי קביעת התפלגות הקושי של השאלות משפיע על צורת הפונקציה המקשרת בין יכולת לציון. מכאן שניתן לראות במבחן הפעלת טרנספורמציה מונוטונית

---

5 להרחבה בנושא זה ראה Yitzhaki and Eisenstaedt (2003) ו-Schechtman and Yitzhaki (2009).

6 הנחה זו מקובלת במבדקים הבודקים האם השאלה שייכת לנושא המבחן – ראה Lord and Novick (1968).

7 הנחה זו אינה קיימת ב-IRT ותוסר על ידנו לאחר מכן.

לא יורדת על המשתנה יכולת, לצורך תרגומו לציון. רמות יכולת שבהן יש הרבה שאלות מבחינות גורמות לכך שהבדל קטן ביכולת יתורגם להבדל גדול בציון, ואילו רמות יכולת שאין בהן שאלות מבחינות יתורגמו להבדלים אפסיים בציון.

כדאי להעיר שעד עתה המעטנו בהבדל שבין מדידת ידע לבין מדידת גובה כאשר הנמדדים מתייצבים מאחורי פרגוד. זאת מאחר שבמדידת גובה מאחורי פרגוד ידוע לנו המרחק בסנטימטרים בין השאלות השונות ואילו במדידת ידע אין לנו מושג על המרחק ביחידות ידע בין השאלות, אלא ידוע לנו רק מספר המשיבים לכל שאלה. מכאן שאין שום אפשרות לאמוד את פונקציית ההתפלגות של ידע (יכולת) וזאת כי חסרה לנו יחידת מדידה שהיא חיונית להגדרת התפלגות. כל מה שאנו יכולים לקבל ממבחן הוא שהנבחנים מסודרים בסדר לא יורד של יכולת.

כמעט כל השיטות שבהן משתמשים להערכת תוצאות מבחנים, כולל אלו הנקראות אי פרמטריות, אינן מתחשבות בהבדל זה שבין מדידה של יכולת למדידת גובה, והן מתייחסות לתוצאות מבחנים כאילו יש להן משמעויות כמותיות שהיו מתקבלות במדידה ישירה של גובה. כך למשל, מי שמחשב ציון ממוצע או מי שמריץ רגרסיה שבה הציון מופיע כמשתנה תלוי או בלתי תלוי, מתייחס לציון כאילו הוא משתנה כמותי שכמוהו כמדידה ישירה של גובה. בסעיף הבא נצביע על הכלי הסטטיסטי שיאפשר לנו להתגבר על בעיה זו. אולם בטרם נמשיך, נסיר חלק מההנחות המפשטות שהנחנו בתחילת הסעיף:

ההנחה הראשונה שנסיר היא ההנחה השלישית והיא שקיימת ודאות מלאה במעבר מיכולת לציון. נניח שההסתברות לענות נכון על שאלה בדרגת קושי מסוים הינה פונקציה לא יורדת של יכולת הנתונה להפרעה אקראית. במקרה כזה, נטען שבממוצע התשובות של נבחנים בעלי יכולת גבוהה תהיינה עם ציון גבוה יותר מאשר התשובות של תלמידים עם יכולת נמוכה. כלומר נניח שהסיכויים לענות נכון על שאלה היא משתנה מקרי שהתוחלת שלו עולה עם היכולת של הנבחן.

אפשרות נוספת היא שתוצאות המבחן אינן בוחנות תחום יכולות אחד אלא בוחנות בערבוביה מספר תחומי יכולות, למשל שהצלחה במבחן תלויה ביכולת מולדת ובשקדנות, ושתי תכונות אלו מפוזרות בצורות שונות באוכלוסייה. במקרה כזה אין זה מחייב שהציון יהיה טרנספורמציה מונוטונית לא יורדת של היכולת המסוימת שנבדקת בו, אלא הוא יהיה תלוי בהתפלגות התכונות, יכולת מולדת ושקדנות באוכלוסייה והמתאם ביניהן.[8] מאחר שאין לנו אפשרות למדוד יכולת והתצפיות שלנו מורכבות רק מהציונים, אנחנו נבדוק את המידה שבה ציון הבגרות מהווה טרנספורמציה מונוטונית לא יורדת של תוצאות הפסיכומטרי וגם את המצב ההפוך: כלומר האם הציון הפסיכומטרי מהווה טרנספורמציה מונוטונית לא יורדת של ציון הבגרות.[9] טענתנו היא שאם הציון בפסיכומטרי הוא פונקציה מונוטונית עולה של הציון בבגרות וגם ציון בבגרות עולה כפונקציה של הציון בפסיכומטרי, נסיק שבגלל המגבלות שקיימות במדידה של ידע הרי שמבחינתנו המבחנים בודקים את אותו תחום יכולות, זאת מאחר שאין לנו יכולת להבדיל ביניהם. מאחר שיכולה להיות הפרעה במונוטוניות של הקשר בין שני הציונים, בגלל ההפרעה האקראית, אנחנו נתעלם מהפרעות קטנות בקשר שבין הציונים ונתייחס רק להפרעות גדולות שהסבירות שהן תוצאה של הפרעה אקראית היא נמוכה.

8 הקורא המעוניין מופנה ל-Yitzhaki, Itzhaki and Pudalov (2012).

9 באופן כללי, ממוצע השיפוע של עקום הרגרסיה של משתנה Y על משתנה X אינו מחויב להיות עם אותו סימן כמו השיפוע הממוצע של עקום הרגרסיה של X על משתנה Y. סימנו של ממוצע השיפועים זהה רק במקרים שבהם הגדרת השונות המשותפת בין המשתנים היא סימטרית במשתנים: רק במקרה זה מתקיימת האמרה "מה שרואים מכאן הוא מה שרואים משם".

ג. בדיקת מונוטוניות של קשר

שיטת המחקר הינה שיטה אי פרמטרית ולא תלוית התפלגות, ומבוססת על האפשרות שטרנספורמציה מונוטונית לא יורדת של אחד המשתנים יכולה להחליף את סימנו של מקדם המתאם ביניהם. החלפת סימנו של מקדם המתאם כמוה כהחלפתו של סימן מקדם הרגרסיה שביניהם. הכלי שנציג ושיאפשר לנו לאתר את ההשפעות השונות בעזרת העין נקרא עקומת LMA, שהיא קיצור השם Line of independence Minus the Absolute concentration curve.

עקומת LMA מאפשרת לזהות האם סימנו של מקדם המתאם בין שני משתנים אינו מתחלף לאורך התחום של אחד המשתנים. אם המסקנה המתקבלת היא שסימנו של מקדם המתאם אינו משתנה לאורך התחום, אזי המסקנה היא שאין טרנספורמציה מונוטונית לא יורדת היכולה לשנות את סימנו של מקדם המתאם, ואם כך נסיק שאין בנמצא כלי שיכול להבחין בין שני תחומי הידע, ועל כן מבחינתנו המבחנים בודקים את אותו תחום יכולות.[10] אם לעומת זאת סימנו של מקדם המתאם משתנה, אזי המשמעות היא שקיימת טרנספורמציה מונוטונית לא יורדת היכולה לשנות את סימנו של מקדם המתאם. בתחום המדידה של מבחנים, המשמעות של קיומה של טרנספורמציה מונוטונית לא יורדת היכולה לשנות את סימנו של מקדם המתאם היא שניתן למצוא מבחן לגיטימי אחר בתחום שיהפוך את התלמיד הטוב במבחן אחד לתלמיד רע במבחן אחר, ולהפך, זאת על ידי שינוי דרגת הקושי של השאלות. המשמעות שיש לייחס לממצא כזה היא שהמבחנים בודקים תחומי יכולת שונים שעל ידי בניית מבחנים מתאימים ניתן יהיה להבחין ביניהם.

המאמר יסביר את הכלים המוצעים לניתוח ויבצע הדגמה על ידי השוואת תוצאות מבחנים פסיכומטריים לעומת מבחני בגרות. מאחר שהשימוש בכלי הוא פשוט יחסית להוכחות הנדרשות כדי לבדוק את נכונות התכונות שלו, ומאחר שההוכחות התפרסמו במספר מאמרים ובספר, הרי שלא נלאה את הקורא בתרגום ההוכחות ונפנה את הקורא המתעניין למאמרים שבהם מוצגות ההוכחות.[11]

ד. שיטת ניתוח הנתונים

אוכלוסיית המחקר היא אלה שנבחנו בבגרויות במתמטיקה ובאנגלית בשנים 2009–2010 והשתתפו במבחן פסיכומטרי. הגבלה נוספת היא שהנבחנים עמדו בדרישות הקבלה לאוניברסיטאות, כלומר, בין היתר, נבחנו ברמה של לפחות 4 יחידות באנגלית וב-3 יחידות ומעלה במתמטיקה. במקביל, על מנת לבדוק את רגישות הממצאים, חזרנו על הבדיקה לשנים 1999–2000. השוואת הממצאים בין השנים מאפשרת ללמוד על הסתגלות האוכלוסייה לקיומם של שני מבחנים.

קיימים מספר קשיים בהשוואת ציונים משני המבחנים: האחד הוא שהנבחן בבחינת הבגרות יכול לבחור את רמת ההעמקה בתחום, וזאת על ידי בחירת מספר היחידות בבחינה, ואילו המבחן הפסיכומטרי הוא אחיד. מכאן שלמעשה בבגרות קיימים מספר מבחנים בהתאם למספר היחידות בבחינה ואילו בפסיכומטרי קיים מבחן אחד. הקושי השני בהשוואה הוא שתחום הציונים (הסקלה) של המבחנים שונה, והשאלה היא כיצד ניתן להביא אותם למכנה משותף הניתן להשוואה.

---

10 במקרה שמדובר בתחומי יכולת שונים, כל אשר אנו יכולים לומר הוא שהמבחנים אינם מסוגלים להבחין ביניהם. זאת משום שאפשר לבחון יכולת רק דרך תוצאות המבחן.

11 ראה Yitzhaki and Schechtman (2013).

על הקושי הראשון נתגבר על ידי כך שנבדוק בנפרד כל רמה של מבחן בגרות מול המבחן הפסיכומטרי, וכן נצרף את כל מבחני הבגרות בהתאם למשקל הניתן על ידי האוניברסיטאות לרמות השונות של הבחינה ועל ידי כך נוכל להשוות את כל התלמידים. אשר לקושי השני הרי שקיימות מספר דרכים לנרמול, כאשר השיטה הנפוצה היא ההנחה של התפלגות נורמלית, ועל כן נהוג לנרמל בעזרת סטיית התקן. אולם כפי שנראה בהמשך, ההתפלגויות של הציונים אינן נורמליות ועל כן שיטת הנרמול שנקטנו בה היא על פי התחום:[12]

השארנו את ציוני הבגרות כמו שהם; נרמלנו את הציון במבחן הפסיכומטרי כך שתחום הציונים בו יהיה זהה לתחום ציוני הבגרות: יהי X ציון פסיכומטרי מקורי, ויהיו A, B ציוני המינימום והמקסימום של ציוני הבגרות ואילו C, D ציוני המינימום והמקסימום של הציון הפסיכומטרי המקורי. יהי Y הציון הפסיכומטרי המנורמל.

נוסחת המעבר הינה:

$$Y = A + \beta(X - C) \text{ כאשר } \beta = \frac{B-A}{D-C} \text{ ו } \alpha = A.$$

התוצאה של הנרמול הינה שתחום ציוני הבגרות ותחום ציוני הפסיכומטרי זהים.

הבדל נוסף הינו בכך שבעוד שהמבחן הפסיכומטרי הוא אחיד לכל הניגשים הרי שמבחני הבגרות במתמטיקה נחלקים לשלוש רמות בהתאם למספר היחידות שהתלמיד נבחן בהם. מכאן שקיים מיון עצמי של התלמיד במבחני הבגרות בהתאם למספר היחידות הנבחר על ידי התלמיד. בשלב זה אנו מגבילים את עצמנו לתלמידים שנבחנו בבגרות ב-5 יחידות. לאחר מכן נצרף הקבצות של תלמידים לקבוצה אחת.

לצורך בחינת המונוטוניות אנחנו נעזרים בשני ציורים של עקומות ריכוז (Absolute Concentration Curves) המאפשרת בחינת המונוטוניות של הקשר.

הציור של העקומה הראשונה הוא בהנחה שתוצאות שני המבחנים בלתי תלויות סטטיסטית. על הציר האופקי מוצגת ההתפלגות המצטברת של הציונים שהתקבלו במבחן האחד. על הציר האנכי מצויר הערך המצטבר של ציוני המבחן השני שקיבלו הנבחנים במבחן זה, בהנחה שציוני המבחנים בלתי תלויים סטטיסטית. הגרף במקרה זה היה קו ישר שמתחיל בראשית הצירים ומסתיים בנקודה (1, ממוצע הציונים).

הציור של העקומה השנייה הוא הערך המצטבר שקיבלו הנבחנים במבחן השני. עקום זה יכול לקבל צורות שונות, וכל מה שנוכל לקבוע בוודאות הוא נקודת ההתחלה ונקודת הסיום. העקום מתחיל בראשית הצירים ומסתיים בנקודה (1, ממוצע הציונים), כלומר באותה נקודה כמו העקום הראשון.[13]

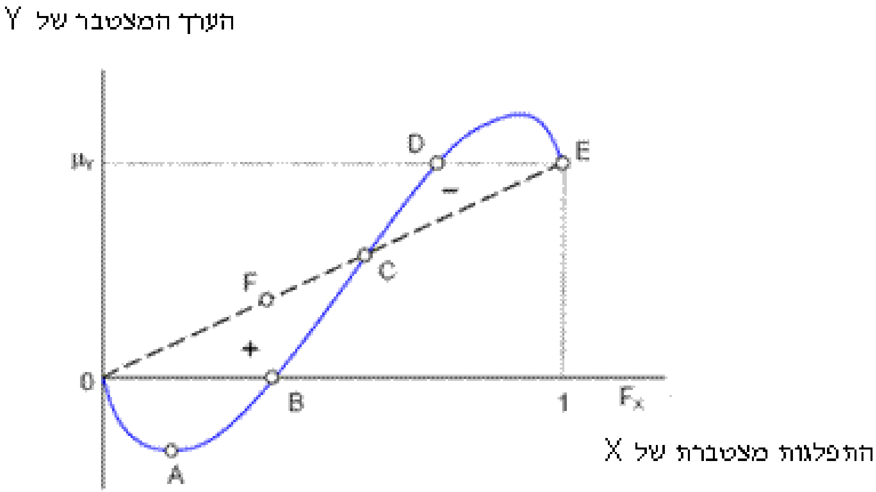
ההפרש האנכי בין שני הגרפים מהווה את עקומת LMA.

לאחר בניית עקומות ה-LMA של מבחני בגרות ופסיכומטרי בתחום מסוים (שנסמנם ב-X ו-Y) נתרכז בשני דברים שחשוב לבדוק כאשר אנו מסתכלים על העקומה:

12 כדאי להעיר שאין בשיטת הנרמול כדי להשפיע על הממצאים. השפעתה היא על הפרשנות שיש לתת לנקודה בציון של מבחן אחד יחסית לנקודה במבחן השני.

13 הערך המצטבר הינו סכום ערכי המשתנה עד לנקודה מסוימת. X מסודר בצורה עולה, אולם אין זה מחייב את הערך של Y בתנאי X.

ציור 4.1: עקומת הריכוז וקו אי תלות



א. **האם העקומה חותכת את הציר האופקי**. במידה שכן אזי הקשר בין המשתנה X ל-Y אינו מונוטוני כאשר X צריך להסביר את Y. כלומר, קיים תחום שבו השונות המשותפת היא שלילית, התחום מתחת לציר האופקי, וקיים תחום שבו השונות המשותפת היא חיובית, התחום שבו העקומה היא מעל הציר האופקי. במקרה כזה טרנספורמציה מונוטונית לא יורדת שמופעלת על הציר האופקי יכולה לשנות את סימנו של מקדם המתאם של פירסון בין שני המשתנים. זאת מאחר שטרנספורמציה מונוטונית של X תוכל להקטין את התחום של המתאם הלא רצוי ולהגדיל את התחום של המתאם הרצוי. המשמעות מבחינתנו היא שהמבחנים אינם בודקים את אותו תחום יכולות, שהרי תוצאות הבדיקה סותרות את קיום ההנחה של מונוטוניות של מתאם בין המשתנים.

ב. **האם קיים קטע שבו העקומה משנה את קעירותה מקעורה לקמורה, או להפך**. גם במקרה זה המשמעות היא שהקשר אינו מונוטוני. אם הקטע הינו קעור אזי המתאם בתוך הקטע הוא חיובי, ואילו קמירות מעידה על מתאם שלילי בתוך הקטע. המשמעות, בהנחה שאין העקומה חותכת את הציר האופקי, היא שאין טרנספורמציה מונוטונית שיכולה לשנות את סימנו של מקדם המתאם. אולם אם נוציא מהמבחן חלק מהתלמידים בתחום שסימנו אינו מוצא חן בעינינו אזי ניתן לשנות את סימנו של מקדם המתאם. כלומר אם נוציא מהאוכלוסייה הנבחנת חלק מהתלמידים שציוניהם בתחום שאיננו חפצים ביקרו אזי קיים מבחן לגיטימי אחר שיהפוך את דירוג הציונים בין שני המבחנים. גם במקרה זה המשמעות היא שאין המבחנים בודקים את אותו תחום יכולות.

ביישום לבדיקת מבחני בגרות ופסיכומטרי, טענתנו תהיה שכל עוד הקשר בין תוצאות מבחן אחד עולה עם עליית תוצאות של מבחן אחר, הרי שמבחינתנו שני המבחנים בודקים את אותו תחום יכולות. לעומת זאת, אם הקשר אינו מונוטוני, נסיק שהמבחנים בודקים יכולות שונות של תלמידים. סייגנו את הקביעה באמירה כי גם אם קיימות יכולות שונות הנדרשות להצלחה בשני המבחנים אין המבחנים, בגלל המונוטוניות של המתאם וגם משום שאין בידינו סרגל למדידת מרחקים בין רמות ידע שונות, מסוגלים להבחין ביניהן.

השטח הכלוא בין עקומת LMA לציר ה-X שווה ל- cov(Y, F(X)) שהוא המקבילה של השונות

המשותפת כאשר מדד הפיזור המשמש אותנו הוא מדד ג'יני. אם ננרמל את הציר האנכי על ידי חלוקה ב- cov(X, F(X)) ישווה השטח הכלוא בין העקום לציר האופקי למקדם הרגרסיה על פי ג'יני של משתנה Y על משתנה X. לעקומה המנורמלת נקרא NLMA, כאשר ה-N מציין שהיא מנורמלת.[14]

לבסוף נעיר שבניגוד לשונות המשותפת המבוססת על השונות שהיא סימטרית בין המשתנים (נובע מכך ש- $\text{cov}(X,Y) = \text{cov}(Y,X)$) הרי שהשונות המשותפת על פי מדד ג'יני אינה סימטרית, ועל כן יש לבדוק האם מה שרואים מכאן הוא מה שרואים משם, כלומר אין ההשפעה של טרנספורמציה מונוטונית עולה על המשתנה X כדין ההשפעה של אותה טרנספורמציה על Y.

בזה נסתיימה סקירת התכונות הרלוונטיות לתחום מדידה בחינוך.

נעבור עתה לבדיקה אמפירית של הנתונים.

ה. בדיקות אמפיריות

אוכלוסיית המחקר היא זכאי בגרות שנבחנו בבגרויות במתמטיקה ובאנגלית בשנת הלימודים 2009–2010 והשתתפו במבחן פסיכומטרי לאחר מכן.[15] כלומר אנו מגבילים את עצמנו לאוכלוסייה שהמיון לצורך קבלה לאוניברסיטה הוא רלוונטי עבורה. אוכלוסייה שנייה המשמשת אותנו היא הנבחנים בשנת הלימודים 1999–2000. סה"כ העומדים בדרישות הסף של האוניברסיטאות בשנת 2009–2010 הוא 31,392, כאשר הנתון המקביל לשנת 1999–2000 הוא 17,867 (בין היתר, אלה שעברו את המבחן במתמטיקה ברמה של 3 יחידות לפחות ובנוסף לכך עברו את המבחן באנגלית ברמה של 4 יחידות לפחות). הגודל של המדגם (למעשה מפקד של הנבחנים בשנה מסוימת) הוא גדול יחסית, כך שכמעט כל סטייה באומדן של פרמטר תהיה מובהקת.

את הבדיקה הראשונה אנו מבצעים על הנבחנים בבגרות באנגלית ברמה של 5 יחידות.

ה.1. הנבחנים ב–5 יחידות באנגלית

לוחות 5.1.1 ו-5.1.2 מרכזים את הפרמטרים המרכזיים המאפיינים את התפלגויות הציונים והמתאמים ביניהם לשתי השנים. כפי שניתן לראות, חל שינוי בציונים בין המבחן הפסיכומטרי (המנורמל) לבין התפלגות ציוני הבגרות בין השנים: בשנים 2009–2010 הממוצע והחציון של הבגרות גבוהים יותר מאשר בפסיכומטרי, ואילו בשנים 1999–2000 הממוצע והחציון בפסיכומטרי גבוהים יותר מאשר בבגרות, תוצאה שניתן להסבירה בכך שקיימת זחילה למעלה בציוני הבגרות ו/או זחילה למטה בציוני הפסיכומטרי. לעומת זאת, אי השוויון בציונים בין השנים נע בכיוונים הפוכים מאלו של הממוצע והחציון. בעוד שבשנים 2009–2010 סטיית התקן והג'יני של הבגרות באנגלית נמוכים יותר מהפרמטרים המקבילים בפסיכומטרי הרי שבשנת 1999–2000 סטיית התקן והג'יני בפסיכומטרי באנגלית נמוכים יותר מאשר בבגרות. ארבעת הטורים האחרונים מתארים את מדדי המתאם המקובלים: מדד פירסון שהוא מדד המתבסס על קשר ליניארי, מדד מתאם הדרגות (ספירמן)

14 את התכונות של מדד ג'יני ניתן למצוא ב-Yitzhaki (2003) ו-Yitzhaki and Schechtman (2013).

15 לכל נבחן נלקח המבחן הפסיכומטרי עם הציון הכולל הגבוה ביותר.

ושני מקדמי המתאם על פי ג'יני,[16] וניתן לראות מתאם גבוה בין תוצאות המבחנים, וזאת על פי כל מקדמי המתאם המקובלים. מאחר שהמתאם בין הציונים הוא החשוב לענייננו הרי שנסתפק בקביעה שהמתאם עלה על פני הזמן.

לוח 5.1.1: פרמטרים מרכזיים של שני המבחנים באנגלית לנבחנים ב 5 יחידות* – 2010–2009

N=17,835	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
בגרות 5 יח'	87.78	90	55	100	7.00	0.044	0.709	0.714	0.719	0.726
פסיכומטרי	84.53	85.32	55	100	9.15	0.062	(0.000)	(0.000)		

*מדד ג'יני המוצג בלוח הוא מדד ג'יני יחסי. על מנת להשוותו לסטיית התקן יש להכפיל בממוצע. כך לדוגמה באנגלית 5 יחידות המדד הוא 0.044 × 87.78 = 3.86 ובפסיכומטרי 0.062 × 84.53 = 5.24.

לוח 5.1.2: פרמטרים מרכזיים של שני המבחנים באנגלית לנבחנים ב 5 יחידות* – 2000–1999

N=12,100	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
בגרות 5 יח'	83.51	85.00	55	100	9.17	0.062	0.682	0.687	0.693	0.693
פסיכומטרי	87.28	89.00	55	100	8.55	0.055	(0.000)	(0.000)		

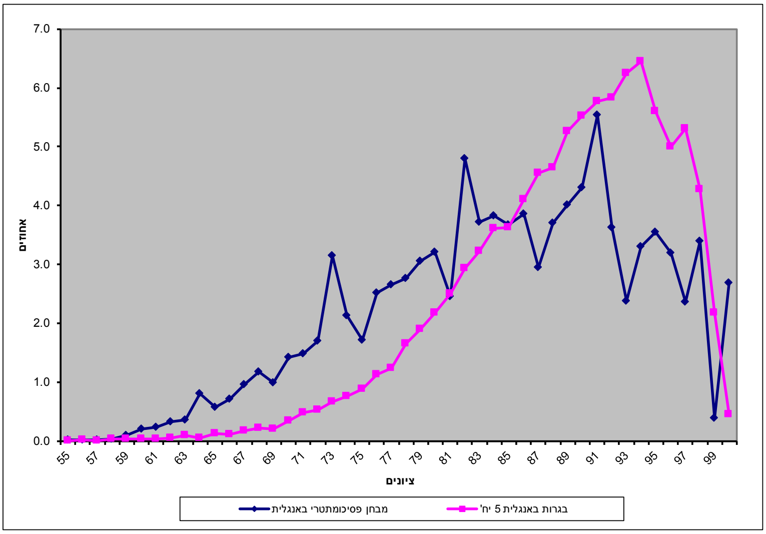
*מדד ג'יני המוצג בלוח הוא מדד ג'יני יחסי. על מנת להשוותו לסטיית התקן יש להכפיל בממוצע. כך לדוגמה באנגלית 5 יחידות המדד הוא 0.062 × 83.51 = 5.18 ובפסיכומטרי 0.055 × 87.28 = 4.80.

מאחר שחלו שינויים בפרמטרים השונים כדאי להסתכל על התפלגויות הציונים בכל שנה כדי לראות את ההבדלים בתוך כל שנה ועל פני השנים.

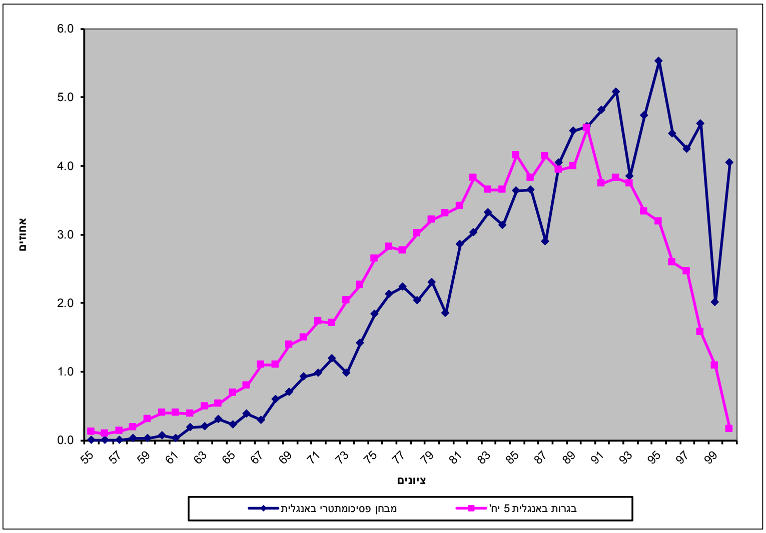
ציורים 5.1.1 ו-5.1.2 מציגים את ההתפלגויות של הציונים בבגרות מול פסיכומטרי מנורמל, הראשון לגבי 2010–2009 והשני לנתוני 2000–1999. מהציורים בולט שבין השנים חל הבדל משמעותי בהתנהגות ציוני הבגרות יחסית לפסיכומטרי. ניכרת הזחילה למעלה של ציוני הבגרות, כך שבשנים 2010–2009 אנו מקבלים התרכזות של ציוני הבגרות בתחום הגבוה של הציונים, וזאת למרות שהמבחן הפסיכומטרי הוא המבחן האחיד כך שבו היינו מצפים שתהיה התרכזות של ציוני נבחנים בחמש יחידות. לעומת זאת, בשנת 2000–1999 הפסיכומטרי מבחין טוב יותר בתחום

16 למדד ג'יני קיימים שני מקדמי מתאם השווים אחד לשני כאשר ההתפלגות המשותפת היא סימטרית בשני המשתנים. חישוב טעות הדגימה של מקדם המתאם של ג'יני מחייב בניית תוכנה מיוחדת. מאחר שכל מקדמי המתאם דומים אחד לרעהו ומאחר שהמדגם גדול, לא חישבנו טעות דגימה.

ציור 5.1.1: התפלגות נבחנים באנגלית 5 יחידות – 2010–2009*



ציור 5.1.2: התפלגות נבחנים באנגלית 5 יחידות – 2000–1999*



* התצפיות חוברו בעזרת קטעים ישרים

היכולות הנמוך ואילו הבגרות מתרכז בתחום היכולות הגבוה, תוצאה שניתן להסביר אותה מחד בהקבצות שקיימות בבגרות ומאידך בכך שהפסיכומטרי הוא מבחן יחיד המיועד לכל. בנוסף, קל לראות שהציונים בשתי השנים ובשני המבחנים אינם מתפלגים נורמלית, וזאת מאחר שההתפלגויות הן א-סימטריות.

השאלה הבאה שנרצה לבחון היא המידה שבה שני המבחנים בודקים את אותו תחום יכולות. לצורך זה אנחנו משתמשים בבדיקת המונוטוניות של הקשר. אם הקשר מתגלה כמונוטוני, כלומר שכאשר עולה הציון במבחן אחד הרי שבממוצע עולה הציון במבחן השני, והקשר אינו משנה את סימנו לאורך ההתפלגות של התוצאות של המבחן, הרי שהמבחנים מודדים את אותו סוג של יכולת.[17] אם לעומת זאת אנחנו רואים שהקשר משנה את סימנו לאורך ההתפלגות הרי שההיסק שלנו יהיה שהמבחנים אינם מודדים את אותה יכולת, זאת מאחר שקשר מונוטוני משמעותו שאיננו יכולים לזהות בין שני התחומים של יכולת.

לצורך בחינת המונוטוניות אנחנו נעזרים בשני ציורים של עקומת NLMA[18] המאפשרת בחינת המונוטוניות של הקשר. על הציר האופקי מוצגת התפלגות מצטברת של הציונים שהתקבלו במבחן האחד. על הציר האנכי מצויר ההפרש בין הערך המצטבר של ציוני המבחן השני, לו היו הציונים בין המבחנים בלתי תלויים סטטיסטית, לבין הערך המצטבר בפועל לאותם נבחנים.

ציורים 5.1.3 ו-5.1.4 בודקים האם הציון הפסיכומטרי יוצר קשר מונוטוני עם ציון הבגרות. תכונות העקומה המצוירת הם אלו: אם העקומה עולה (יורדת) סימן הוא שהציונים בתחום הזה נמוכים (גבוהים) מהממוצע של הציונים במבחן. אם העקומה קעורה (קמורה) סימן שהציונים בתחום זה עולים (יורדים). השטח הכלוא בין העקומה לציר האופקי שווה למקדם המתאם של הג'יני. אם לעומת זאת העקומה חוצה את הציר האופקי סימן הוא שקיים מבחן לגיטימי אחר[19] שיכול להפוך את סימנו של המתאם בין ציוני שני המבחנים.

משני הציורים אנחנו למדים שהעקומה עולה בתחילה ולאחר מכן יורדת – כלומר שבתחום שבו ציוני הבגרות נמוכים גם ציוני הפסיכומטרי נמוכים ואילו בתחום שבו ציוני הבגרות גבוהים גם ציוני הפסיכומטרי הם גבוהים. אשר לקעירות העקומה אנחנו רואים שהעקומה היא קעורה, כך שאנחנו יכולים להסיק שציוני הפסיכומטרי יוצרים קשר מונוטוני לציוני הבגרות.

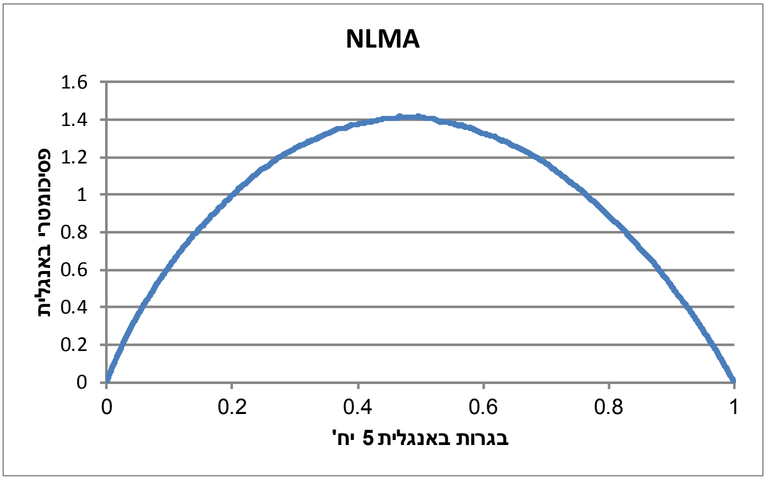
כדי לבדוק האם גם הציון של הבגרות יוצר קשר מונוטוני עם הציון הפסיכומטרי, אנחנו חוזרים על ציורים 5.1.3 ו-5.1.4 בהיפוך הצירים: על הציר האופקי הציון הפסיכומטרי ואילו הציר האנכי מציג את ציון הבגרות. ציור 5.1.5 מציג את עקומת NLMA של ציוני הבגרות כפונקציה של ההישגים במבחן הפסיכומטרי לשנת 2009–2010. העקומה המצטיירת מביאה למסקנה זהה לזו שהתקבלה מההסתכלות בדרך ההפוכה – ציון הבגרות מהווה קשר מונוטוני לציון הפסיכומטרי. מכאן שהקשר הוא סימטרי ומונוטוני ועל כן, לפחות לגבי קבוצה זו, המבחנים מהווים תחליפים (שאינם מושלמים) אחד לשני. אנחנו מציינים שהם אינם מושלמים בגלל רמת המתאם הנמוכה יחסית בין ציוני שני המבחנים. ציור 5.1.6 חוזר על ציור 5.1.5 וזאת לשנת 1999–2000. המסקנה אינה משתנה.

17 ייתכן כמובן שההצלחה במבחנים השונים תלויה בגורמים שונים אולם אין המבחנים מסוגלים להבחין ביניהם.

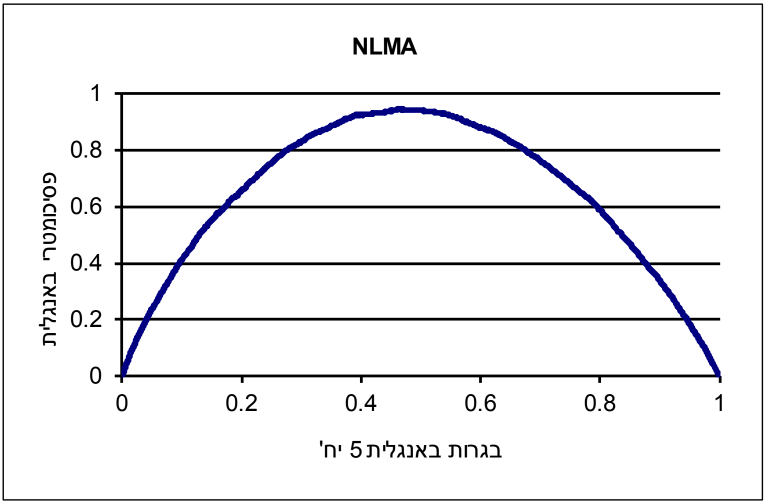
18 על תכונות העקומה והשימוש בתחום החינוך – ראה Yitzhaki, Itzhaki and Pudalov (2012).

19 במונח "מבחן לגיטימי אחר" הכוונה היא למבחן אחר עם התפלגות שונה של קושי השאלות בו.

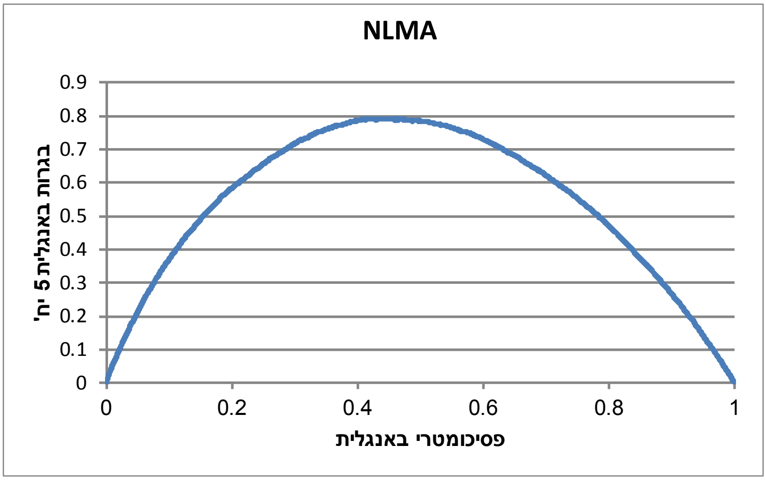
ציור 5.1.3: ציון פסיכומטרי באנגלית כפונקציה של בגרות באנגלית 5 יח' – 2010–2009



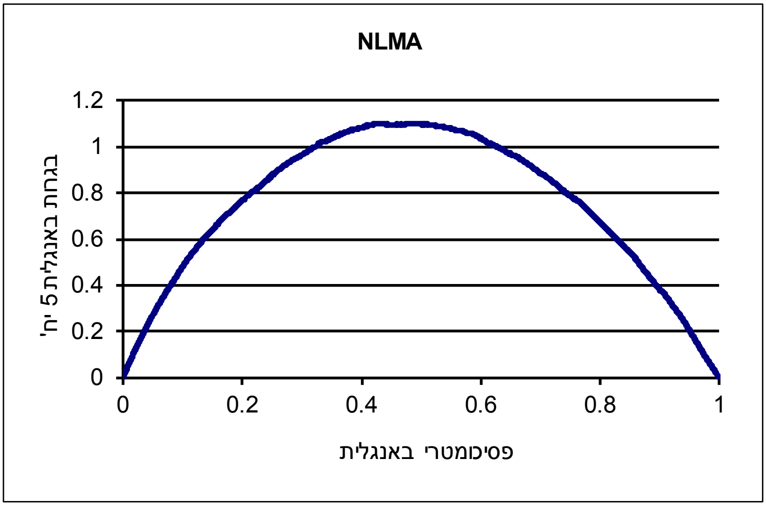
ציור 5.1.4: ציון פסיכומטרי באנגלית כפונקציה של בגרות באנגלית 5 יח' – 2000–1999



ציור 5.1.5: ציון בגרות באנגלית 5 יחידות כפונקציה של ציון פסיכומטרי באנגלית – 2010–2009



ציור 5.1.6: ציון בגרות באנגלית 5 יחידות כפונקציה של ציון פסיכומטרי באנגלית – 2000–1999



כדי לבדוק בצורה אי-פרמטרית את הסימטריה בקשר בין ציוני המבחנים, ליווינו את הציונים בעזרת מטריצת מעבר. לוח 5.1.3 מציג את מטריצת המעבר בין עשירוני הציונים לשנת 2009–2010 ואילו לוח 5.1.4 – לשנת 1999–2000. על הציר האופקי נמצאים עשירוני ציון בבגרות באנגלית 5 יחידות ועל הציר האנכי נתוני הפסיכומטרי מנורמל לאותם תלמידים. האלכסון מציג את אחוז התלמידים שציוניהם נמצאים באותו עשירון בשני המבחנים. כ-50 אחוזים מקבוצת העשירון התחתון נשארים בעשירון התחתון הן בבגרות והן בפסיכומטרי, וקרוב לחמישים אחוזים מהעשירון העליון נשארים בעשירון העליון. רק חמישה אחוזים מהעשירון התחתון בפסיכומטרי נמצאים מעל החציון בבגרות, ואחוז דומה מהעשירון התחתון בבגרות עולה מעל החציון בפסיכומטרי.

המשמעות שניתן לייחס לממצאים ממטריצת המעבר היא שהמבחנים בודקים את אותו נושא, אולם יש הרבה רעש אקראי בתוצאות.

השוואה בין מטריצות המעבר בין השנים מלמדת שהמתאם בין ציוני הבגרות והפסיכומטרי הלך והתחזק על פני העשור, מה שמעיד כנראה על השקעה גדולה יותר אצל התלמידים. מבחינתנו, התוצאות מעידות שהממצאים של מתאם מונוטוני התחזקו על פני העשור. כך למשל עלה אחוז הנשארים לאורך האלכסון מ-24.3 אחוזים מהאוכלוסייה ל-26.1 אחוזים. הקצוות של מטריצת המעבר התרוקנו, כלומר השינויים בדירוג שהיו בשנת 1999–2000 אינם, ואילו הסימטריה נשארה.

קיומו של "רעש" בתוצאות המבחנים צריך להטריד מאחר שהמשמעות היא שיש נבחנים העוברים מהעשירון העליון לעשירון התחתון של הציונים, והמעברים נראים "סימטריים" כך שאין אפשרות לייחס אותם לחוסר תקינות במערכות הציונים של אחד המבחנים. המסקנה המתקבלת היא ששני המבחנים בוחנים את אותו נושא, אולם כמות הטעויות בדירוג של כל מבחן יחסית לרעהו היא גבוהה ועל כן הממוצע של שני המבחנים יכול לשמש כאינדיקטור טוב יותר מאשר הסתמכות על תוצאות מבחן אחד.

כדאי להעיר שללא הערכת עלויות חברתיות אין לנו יכולת להשיב על השאלה באם כפילות זו כדאית לחברה. כל מה שאנו יכולים לקבוע הוא שאין הבדל בין המבחנים מבחינת הנושא שהם בודקים ושסדרי הגודל של ההבדלים, ברמת הפרט בין שני המבחנים, הם גדולים. למי שמעוניין בבחינת עלות-תועלת של קיום שני המבחנים, הרי שבנוסף לאופציה של ביטול אחד המבחנים כדאי יהיה לבדוק קיום חוזר של אותו מבחן וזאת על מנת שלא יהיה צורך לדרוש מהתלמידים להתכונן לשתי מערכות מבחנים שונות.

ה.2. הנבחנים במתמטיקה – 5 יחידות בבגרות

בסעיף זה אנו משווים בין תוצאות מבחני הבגרות במתמטיקה לתוצאות המבחן הפסיכומטרי הכמותי בקרב תלמידים שניגשו למבחן בגרות במתמטיקה של חמש יחידות. מספר התלמידים שיש לנו עבורם ציונים בשתי הבחינות הוא 11,949 בשנת 2009–2010 לעומת 7,059 במהלך 1999–2000. כדאי לשים לב שמספר הנבחנים במתמטיקה 5 יחידות נמוך בכ-40 אחוז מאשר מספר הנבחנים המקביל באנגלית 5 יחידות (כ-7,000 לעומת כ-12,000). המשך הניתוח דומה במבנהו לניתוח שנעשה עבור המבחנים באנגלית, ועל כן אנו מתרכזים בממצאים בלבד.

לוח 5.1.3: מטריצת מעבר: עשירוני ציונים – בגרות באנגלית 5 יח' ופסיכומטרי באנגלית – 2010–2009

		בגרות באנגלית 5 יחידות										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	סה"כ
פסיכומטרי באנגלית	**1**	5.0	2.3	1.3	0.6	0.3	0.2	0.1	0.1			10
	2	2.6	2.5	1.9	1.2	0.9	0.6	0.2	0.2			10
	3	1.3	2.0	1.9	1.9	1.2	0.8	0.4	0.3	0.1		10
	4	0.6	1.3	1.8	2.0	1.6	1.2	0.5	0.7	0.3	0.1	10
	5	0.2	0.9	1.2	1.7	1.8	1.1	1.4	0.9	0.4	0.1	10
	6	0.2	0.4	0.9	1.0	1.3	1.8	1.9	1.5	0.8	0.2	10
	7	0.1	0.3	0.6	0.7	1.1	1.4	2.1	2.0	1.4	0.5	10
	8		0.2	0.3	0.5	0.8	1.4	1.9	1.5	2.7	0.7	10
	9		0.1	0.2	0.3	0.5	1.0	1.3	1.5	2.1	3.0	10
	10				0.1	0.2	0.4	0.4	1.3	2.1	5.4	10
	סה"כ	10	10	10	10	10	10	10	10	10	10	100

לוח 5.1.4: מטריצת מעבר: עשירוני ציונים – בגרות באנגלית 5 יח' ופסיכומטרי באנגלית – 2000–1999

		בגרות באנגלית 5 יח'										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	סה"כ
פסיכומטרי באנגלית	**1**	4.8	2.0	1.6	0.7	0.5	0.2	0.1	0.03	0.01		10
	2	2.0	2.2	2.4	1.5	0.8	0.6	0.3	0.1	0.03		10
	3	1.3	1.8	1.9	1.7	1.0	1.1	0.8	0.2	0.2	0.02	10
	4	0.9	1.6	1.4	1.8	1.4	1.3	0.5	0.8	0.3	0.05	10
	5	0.6	1.1	0.9	1.1	1.5	1.7	1.1	1.2	0.7	0.1	10
	6	0.2	0.6	0.8	1.4	1.7	1.6	1.1	1.7	0.8	0.2	10
	7	0.1	0.3	0.5	0.7	1.1	1.2	2.0	2.0	1.5	0.6	10
	8	0.1	0.3	0.3	0.5	1.0	1.0	1.7	1.8	2.1	1.2	10
	9	0.1	0.2	0.3	0.3	0.7	0.8	1.5	1.2	2.0	3.1	10
	10		0.1	0.1	0.2	0.3	0.4	0.9	1.0	2.4	4.7	10
	סה"כ	10	10	10	10	10	10	10	10	10	10	100

לוח 5.2.1: פרמטרים מרכזיים של שני המבחנים במתמטיקה לנבחנים ב-5 יחידות – 2010–2009

N=11,949	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
מתמטיקה 5 יח'	86.4	89.0	55	100	10.55	0.068	506 .0 (0.000)	0.527 (0.000)	0.534	0.529
מתמטיקה פסיכומטרי	89.4	90.6	55	100	6.68	0.042				

לוח 5.2.2: פרמטרים מרכזיים של שני המבחנים במתמטיקה לנבחנים ב-5 יחידות – 2000–1999

N=7,059	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
מתמטיקה 5 יח'	85.9	88.0	53	100	10.79	0.070	0.384 (0.000)	0.391 (0.000)	0.409	0.392
מתמטיקה פסיכומטרי	87.9	89.0	53	100	7.94	0.049				

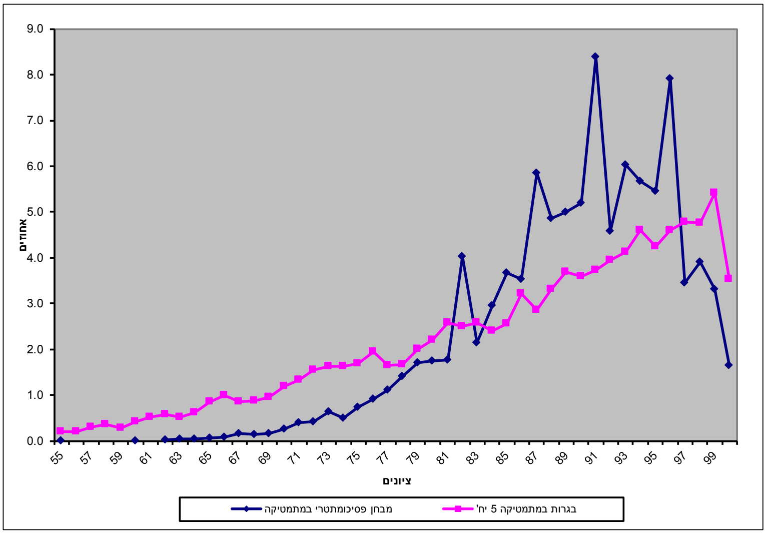
כפי שניתן לראות מלוחות 5.2.1 ו-5.2.2 ממוצע הציון והחציון (המנורמלים) בפסיכומטרי גבוהים יותר מהממוצע והחציון של הציונים בבגרות (5 יחידות). לעומת זאת שני מדדי הפיזור שהשתמשנו בהם, סטיית התקן ומדד ג'יני, גבוהים יותר במבחן הבגרות מאשר במבחן הפסיכומטרי. המשמעות היא שציוני הבגרות הם בעלי פיזור גדול יותר מאשר ציוני הפסיכומטרי, כלומר יחסית לפסיכומטרי הם מרוכזים יותר בקצוות ההתפלגות של הציונים. מצב מסוג זה יכול לנבוע מנרמול הציונים לנבחנים ב-5 יחידות בבגרות, נרמול שנעשה באופן בלתי תלוי לשאר קבוצת הנבחנים במתמטיקה. זאת, מאחר שבפסיכומטרי יש מבחן יחיד והרי טבעי הדבר שמי שניגש ל-5 יחידות במתמטיקה יהיה במקבץ הגבוה של הציונים בפסיכומטרי.[20] הממצא המפתיע כאן הוא המתאם הנמוך יחסית בין שני מבחנים שהועברו לאותם אנשים, שהינם בעלי ההשכלה הרחבה ביותר במתמטיקה. כל מקדמי המתאם לשנת 1999–2000 בין תוצאות שני המבחנים אינם עולים על 0.4, וזאת כאשר באנגלית הם עומדים על 0.7, לעומת זאת המתאמים לשנת 2009–2010 גבוהים יותר והם בסביבות 0.5, מה שמחזק את ההשערה שהאוכלוסייה "למדה" להתאים עצמה לקיומם של שני מבחנים שונים. ממצא מפתיע שני הוא שבדומה לתוצאות המבחן באנגלית, גם כאן מתקבלת סימטריה בהתפלגות המשותפת של הציונים בבגרות ובפסיכומטרי. ממצא מפתיע נוסף הוא הדמיון בין שני מקדמי המתאם של הג'יני,

20 הטענה החבויה מאחורי משפט זה יכולה להיבדק על ידי בדיקת ההשערה שהנבחנים ברמות שונות בבגרות מהווים קבוצות ריבוד במבחן הפסיכומטרי. השערה זו לא נבדקה בעבודה זו.

כך ש "מה שרואים מכאן (ממבחן הבגרות) דומה למה שמרואים משם (מהמבחן הפסיכומטרי)", כלומר ששינוי הבסיס להשוואה אינו משנה את הממצא ביחס למתאם. הסימטריה שבין ההתפלגויות והמתאם הנמוך יותר במבחני המתמטיקה מעידים שכמות הרעש האקראי במתמטיקה גדולה יותר מכמות הרעש האקראי באנגלית.

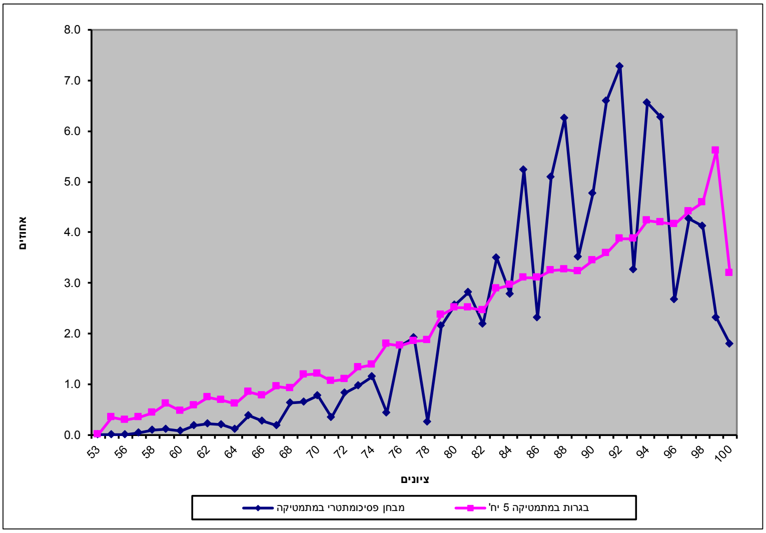
ציורים 5.2.1 ו-5.2.2 מציגים את ציוני הבגרות וציוני הפסיכומטרי המנורמלים לתלמידים שנבחנו במתמטיקה בבגרות ברמה של 5 יחידות. מהציור בולט שהמבחן הפסיכומטרי, יחסית לבגרות 5 יחידות, מתרכז בהבחנה בתחום העליון ובתחום התחתון של היכולות ואילו הבגרות מתרכזת בתחום האמצעי של היכולות. אנחנו רואים זאת מכך שהצפיפות של הנבחנים גבוהה יותר בפסיכומטרי במרכז ואילו בבגרות 5 יחידות הצפיפות גבוהה יותר בשני הקצוות של ההתפלגות.[21] מאחר שקיים רק מבחן פסיכומטרי אחד ואילו בבגרות יש שלוש הקבצות, נראה הגיוני שמבחני הבגרות יהיו בעלי הבחנה נמוכה יותר בתחום היכולות התחתון והעליון. אולם כפי שציינו באנגלית הדבר יכול לנבוע מהליך הנרמול שנערך בבגרות עבור הציונים המתקבלים בין נבחני 5 יחידות בנפרד.

ציור 5.2.1: התפלגות ציוני פסיכומטרי ובגרות במתמטיקה של נבחני מתמטיקה 5 יחידות – 2009–2010



21 כדאי לשים לב שמאחר שההתפלגות של הציונים תלויה בהתפלגות הקושי של השאלות, אין יכולת לאמוד את ההתפלגות האמיתית של האוכלוסייה.

ציור 5.2.2: התפלגות ציוני פסיכומטרי ובגרות במתמטיקה של נבחני מתמטיקה 5 יחידות – 1999–2000*



* בבגרות במתמטיקה 5 יחידות יש סה"כ 10 שאלות: 6 שאלות של 16⅔ נקודות (3 יחידות), ו-4 שאלות של 25 נקודות (השלמה ל-5 יחידות). לאחר מכן משקללים את התוצאה. בפסיכומטרי במתמטיקה יש 50 שאלות (שני חלקים של 25 כל אחד). חוסר השיאים הרבים בתוצאות הבחינה הפסיכומטרית נראים לנו מוזרים, אבל אלו הן התוצאות שקיבלנו.

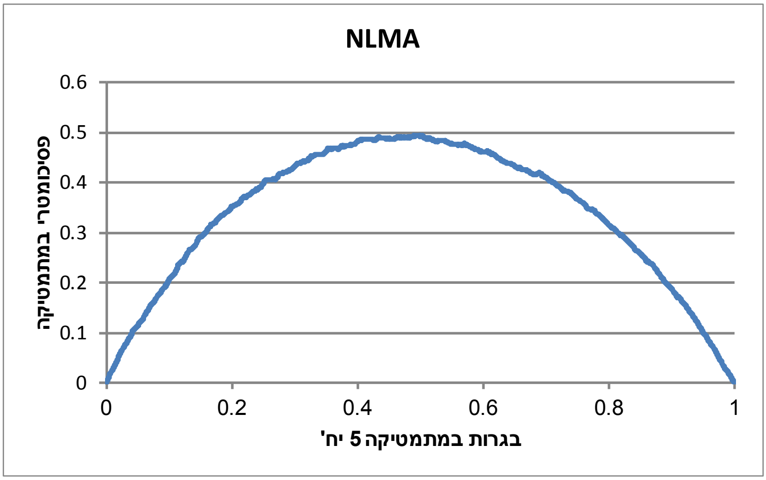
הממצא המעניין מהציורים הוא שאין כל תוספת אינפורמציה בהשוואת הממצאים לגבי שתי התקופות, ומה שניתן להסיק מנתוני תקופה אחת תקף גם לנתוני התקופה השנייה.

השאלה הבאה שנרצה לבחון היא המידה שבה שני המבחנים בודקים את אותו תחום יכולות. לצורך זה אנחנו משתמשים בבדיקת המונוטוניות של הקשר. אם הקשר מתגלה כמונוטוני, כלומר כאשר הציון במבחן אחד עולה הרי שבממוצע גם הציון במבחן השני עולה, והקשר המונוטוני אינו משנה את סימנו לאורך התפלגות התוצאות של המבחן, הרי שהמבחנים מודדים את אותו תחום יכולות. אם לעומת זאת אנחנו רואים שהקשר משנה את סימנו לאורך ההתפלגות הרי שההיסק שלנו יהיה שהמבחנים אינם מודדים את אותו תחום יכולות.

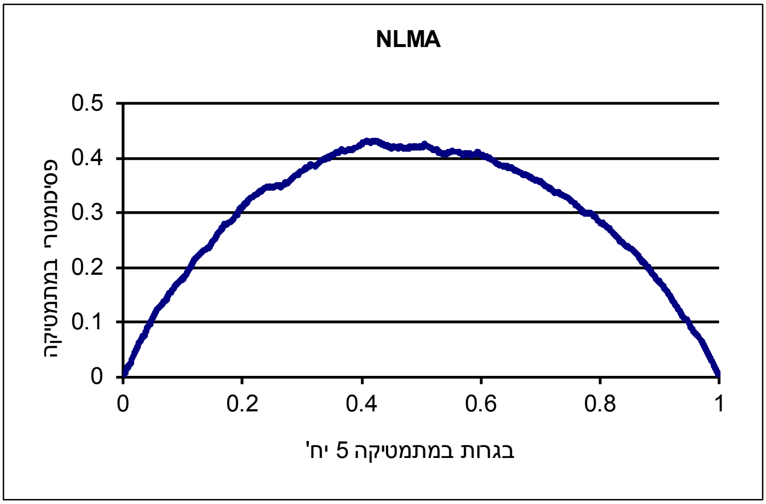
ציורים 5.2.3 ו-5.2.4 מציגים את עקומות ה-NLMA לשתי השנים. נראה שגם כאן אין הבדל בשתי השנים. הקשר הוא מונוטוני, וזאת למרות שהמתאם במבחני המתמטיקה קטן יותר מהמתאם באנגלית. עקום ה-NLMA הוא פחות חלק מאשר באנגלית, דבר היכול לנבוע גם מכך שמספר הנבחנים במתמטיקה קטן יותר מאשר באנגלית, שם מספר נבחנים גדול יכול לגרום להחלקה של ההתפלגות ולהיעלמות של תנודות אקראיות.

ציורים 5.2.5 ו-5.2.6 מציגים את עקומות NLMA של ציוני הבגרות כפונקציה של ההישגים במבחן הפסיכומטרי. בדומה לאנגלית, בתחום המתמטיקה העקומה המצטיירת מביאה למסקנה זהה לזו שהתקבלה מההסתכלות בדרך ההפוכה. ציון הבגרות מהווה קשר מונוטוני לציון הפסיכומטרי, להוציא תנודות אקראיות לאורך תחומים קטנים.

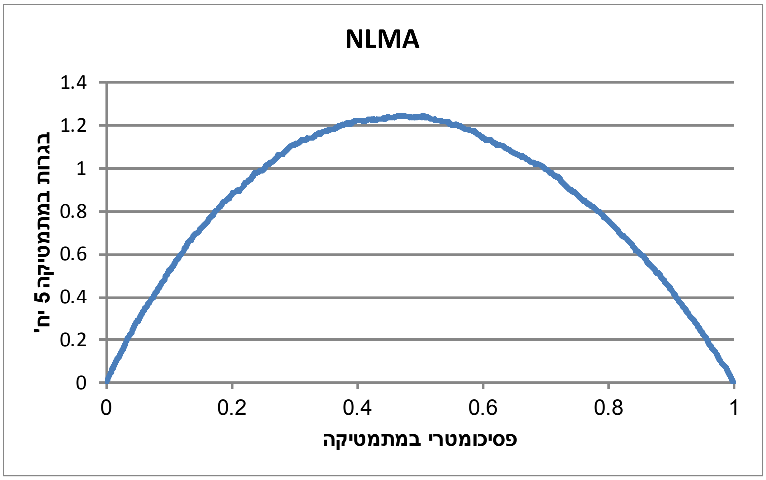
ציור 5.2.3: ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה 5 יח' – 2009–2010



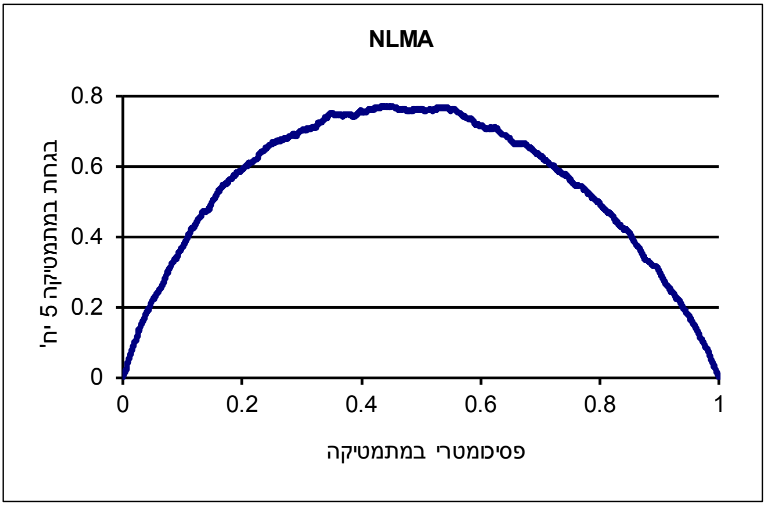
ציור 5.2.4: ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה 5 יח' – 1999–2000



ציור 5.2.5: ציון בגרות במתמטיקה 5 יח' כפונקציה של ציון פסיכומטרי במתמטיקה – 2009–2010



ציור 5.2.6: ציון בגרות במתמטיקה 5 יח' כפונקציה של ציון פסיכומטרי במתמטיקה – 1999–2000



מבחינה זו, הממצא שלמרות שהקשרים הם מונוטוניים מקדמי המתאם נמוכים מטריד במקצת. כדי להבין את משמעותו נציג בלוחות 5.2.3 ו-5.2.4 מטריצות מעבר בין עשירוני ציונים. על הציר האופקי נמצאים עשירוני ציון בבגרות במתמטיקה 5 יחידות ועל הציר האנכי נתוני הפסיכומטרי. האלכסון מציג את אחוז התלמידים שנשארו באותו עשירון. אם נבדוק את העשירון התחתון בשני המבחנים נראה שרק 3.4 אחוזים נשארו בעשירון התחתון בשני המבחנים ב-2009–2010 ואילו הנתון המקביל ל-1999–2000 הוא ש-2.7 אחוז מהתלמידים הם גם בעשירון התחתון של הבגרות וגם בעשירון התחתון של הפסיכומטרי. כלומר, לגבי העשירון התחתון יש הסכמה בין המבחנים רק לגבי כ-3 אחוזים מהתלמידים (שהם כ-30 אחוזים מהקבוצה). המפתיע הוא שכ-14 אחוזים מהקבוצה של העשירון התחתון בפסיכומטרי בשנת 2009–2010 (כ-18 אחוזים בשנת 1999–2000) נמצאים מעל החציון בבגרות (0.1+0.2+0.2+0.4+0.5) וכ-15 אחוז מהעשירון התחתון של הבגרות נמצאים מעל החציון בפסיכומטרי (כ-25 אחוז לשנת 1999–2000).

ממצא דומה ניתן לקבל אם מתרכזים בעשירון העליון שרק 43 אחוז (29 לגבי שנת 1999–2000) מתוכו נשארים בעשירון העליון של שני המבחנים. התרכזנו בעשירונים העליונים כי בהם הסיכוי להישאר באותו עשירון בשני המבחנים הוא הגבוה ביותר, זאת מאחר שהצפיפות של ההתפלגות גבוהה יותר במרכז מאשר בשוליים, מה שגורם שהסיכוי להישאר באותו עשירון במרכז נמוך יותר מאשר בשוליים. מטריצות המעבר נראות כמעט סימטריות, מה שמעיד על פיזור שהוא רעש לבן.

המסקנה המתבקשת היא שכושר הניבוי של מבחן יחיד לגבי ההצלחה במבחן האחר הוא נמוך ומסתכם במקדם מתאם של כ-40 אחוז. המסקנה המתקבלת היא שהמבחן במתמטיקה יש בו רעשים אקראיים המעיבים על יכולת הניבוי של הצלחה במבחן אחד לעומת ההצלחה במבחן השני. לא בדקנו את כושר הניבוי להצלחה בלימודים אקדמיים, אולם המתאם הנמוך בין הבגרות והפסיכומטרי אינו מנבא טובות לכושר הניבוי לגבי מבחנים מסוג נוסף.

לוח 5.2.3: מטריצת מעבר: עשירוני ציונים – בגרות במתמטיקה 5 יח' ופסיכומטרי במתמטיקה 2009–2010

		בגרות במתמטיקה 5 יחידות										
		1	2	3	4	5	6	7	8	9	10	סה"כ
פסיכומטרי במתמטיקה	1	3.4	2.1	1.4	1	0.8	0.4	0.5	0.2	0.2	0.1	10
	2	2.1	2.1	1.3	1.3	1.1	0.6	0.7	0.3	0.2	0.1	10
	3	1.4	1.8	1.6	1.3	1.3	0.7	0.9	0.5	0.3	0.2	10
	4	0.9	1.3	1.4	1.2	1.4	0.9	1.5	0.6	0.5	0.3	10
	5	0.8	0.9	1.1	1.2	1.4	1.4	1.5	1	0.7	0.4	10
	6	0.4	0.6	1.2	1.3	1.2	1.4	1.3	1.2	0.9	0.5	10
	7	0.5	0.6	0.8	0.9	1	1.3	0.8	1.8	1.4	0.9	10
	8	0.3	0.4	0.6	0.7	0.8	1.2	1.2	2	1.5	1.4	10
	9	0.2	0.3	0.4	0.6	0.9	1.3	1	1.2	2.4	1.8	10
	10	0.1	0.1	0.2	0.4	0.4	0.8	0.7	1.2	1.9	4.3	10
	סה"כ	10	10	10	10	10	10	10	10	10	10	100

מטריצת המעבר במתמטיקה מדגימה את הגדלת ה"רעש" לעומת מבחנים באנגלית. המסקנה המתקבלת היא שיש מתאם גבוה יותר בין תוצאות המבחנים באנגלית מאשר במתמטיקה. על כן הנזק שייגרם למשק מביטול ההכבדה שבקיום שני מבחנים באנגלית הוא קטן יותר מאשר הנזק שייגרם ממיון מועמדים במתמטיקה.

לוח 5.2.4: מטריצת מעבר: עשירוני ציונים – בגרות במתמטיקה 5 יח' ופסיכומטרי במתמטיקה 1999–2000

פסיכומטרי במתמטיקה	בגרות במתמטיקה 5 יחידות										
	1	2	3	4	5	6	7	8	9	10	סה"כ
1	2.7	1.9	1.5	1.2	0.9	0.7	0.4	0.3	0.3	0.1	10
2	1.8	1.9	1.4	1.2	0.7	0.9	0.8	0.6	0.5	0.2	10
3	1.4	1.2	0.9	1.2	1.2	1.2	0.9	0.9	0.8	0.4	10
4	0.7	1.3	1.1	1.3	1.2	1	1.2	1.1	0.7	0.4	10
5	0.8	1.1	1	1.3	1.1	1	0.9	0.9	1	0.9	10
6	0.7	1	1.4	1.1	1.2	1.1	1	1	0.8	0.8	10
7	0.7	0.7	0.9	0.8	1.1	0.8	1.3	1.3	1.3	1.2	10
8	0.6	0.5	0.8	0.8	1.3	1.3	1.3	1.2	1.3	1.2	10
9	0.3	0.4	0.6	0.8	0.8	1	1.4	1.3	1.5	2	10
10	0.2	0.3	0.3	0.4	0.7	0.9	0.9	1.4	2	2.9	10
סה"כ	10	10	10	10	10	10	10	10	10	10	100

ה.3. נבחנים באנגלית – 4 ו–5 יחידות

בסעיף זה אנו חוזרים על הניתוח שנעשה בסעיף הקודם כאשר הפעם אנחנו כוללים את התלמידים שנבחנו ב-4 וב-5 יחידות, כלומר כל התלמידים שזכאים להתקבל לאוניברסיטה. השקלול של ציוני הבגרות נעשה על פי הכללים המקובלים באוניברסיטאות בעת קבלת סטודנטים.[22] מספר התצפיות המשמשות אותנו הוא 31,392 לשנת 2009–2010 ו-17,867 נבחנים שנבחנו בשנים 1999–2000 ושעומדים בדרישות הסף של האוניברסיטה. כמו בסעיף הקודם אנו מנרמלים את הציונים כך שתחום הציונים בשני המבחנים יהיה זהה.

22 למקצועות אנגלית ומתמטיקה שנלמדו ברמה של 4 יחידות לימוד מוסיפים 12.5 נקודות, ולרמה של 5 יחידות לימוד מוסיפים 25 נקודות.

לוח 5.3.1: פרמטרים מרכזיים של שני המבחנים באנגלית – 2010–2009

N=31,392	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
בגרות באנגלית	105.7	107.0	68	125	12.32	0.066	0.815 (0.000)	0.842 (0.000)	0.843	0.837
פסיכומטרי באנגלית	96.3	96.0	68	125	14.94	0.090				

לוח 5.3.2: פרמטרים מרכזיים של שני המבחנים באנגלית – 2000–1999

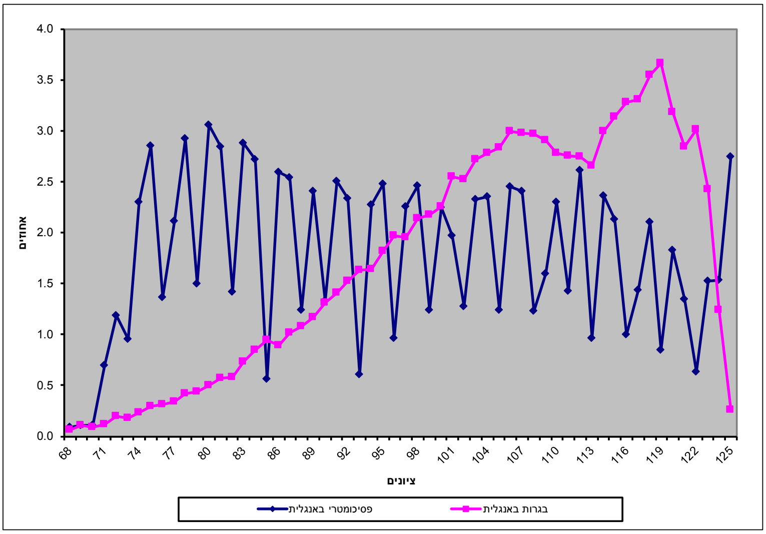
N=17,867	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
בגרות באנגלית	101.5	103	68	125	13.72	0.077	0.82 (0.000)	0.82 (0.000)	0.83	0.83
פסיכומטרי באנגלית	102.3	104	68	125	14.41	0.081				

השוואה בין הלוחות מראה על עלייה דרמטית בין השנים הן בבגרות והן בפסיכומטרי, וזאת הן בממוצע הציון והן בחציון. לעומת זאת קיימת ירידה בפיזור של ציוני הבגרות. סדרת המתאמים עלתה לסדר גודל של 0.84 לעומת מתאם בסדר גודל של 0.83 לפני עשור. השוואה של ההבדלים בין הציונים עם הלוח המקביל של חמש יחידות מראה שהממוצע והחציון בשתי ההתפלגויות כמעט זהים, אולם סטיית התקן של כל מבחן ומקדם אי השוויון של ג'יני גדלו, תוצאה טבעית מכך שהרחבנו את האוכלוסייה הנבחנת. ציורים 5.3.1 ו-5.3.2 מציגים את התפלגות השכיחויות. ניתן לראות שהתפלגות ציוני הבגרות רציפה יותר מאשר התפלגות ציוני הפסיכומטרי. אם "מחליקים" את ציוני הפסיכומטרי על ידי מיצוע של תצפיות סמוכות היינו מקבלים התפלגויות כמעט זהות. יתר על כן, השוואת מקדמי המתאם מראה גם שכל מקדמי המתאם לא ירדו על פני הזמן וחלקם אף עלו, כך שטיב ההתאמה בין שתי מערכות הציונים נעשה גבוה הרבה יותר מאשר בהשוואה של פסיכומטרי עם נבחנים בחמש יחידות בלבד.

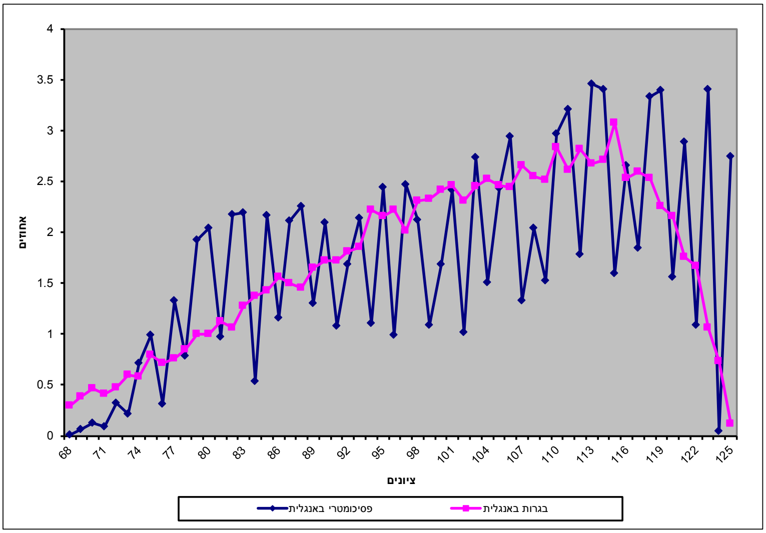
ציורים 5.3.3 ו-5.3.4 מציגים עקומות NLMA של ציון פסיכומטרי כפונקציה של ציוני הבגרות. כפי שנראה מהציורים העקומה הינה קעורה לכל ארכה ועל כן, גם במקרה זה אנו מקבלים שציוני הפסיכומטרי מהווים פונקציה מונוטונית עולה של ציוני הבגרות, מה שמעיד שאנחנו עוסקים באותו תחום.

כמו בסעיף הקודם, גם כאן אנחנו הופכים את המשתנה התלוי והמשתנה הבלתי תלוי ומציגים את עקומת ה-NLMA של ציון בגרות כפונקציה של הפסיכומטרי. גם כאן העקומה היא קעורה וחלקה, מה שמעיד על קשר מונוטוני בין הציונים. על כך מעידים גם מקדמי המתאם של הג'יני בין שני המשתנים שהם בגודל דומה.

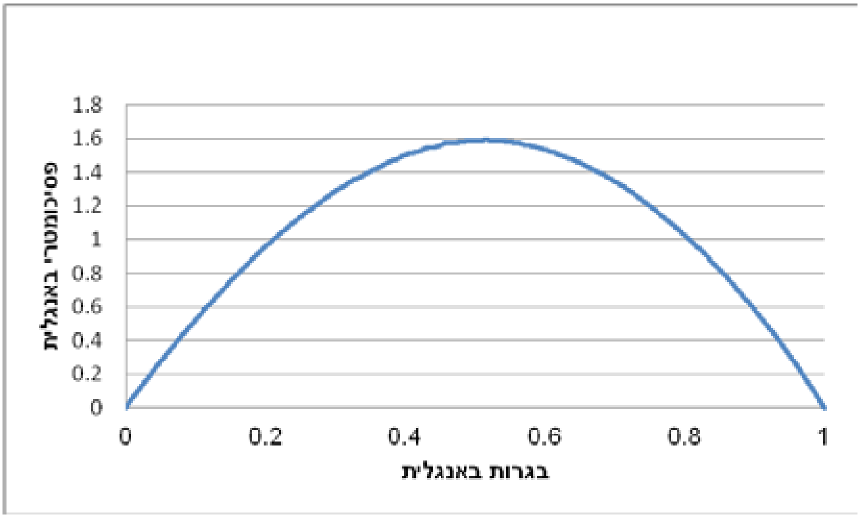
ציור 5.3.1: התפלגות ציוני פסיכומטרי ובגרות באנגלית – 2010–2009



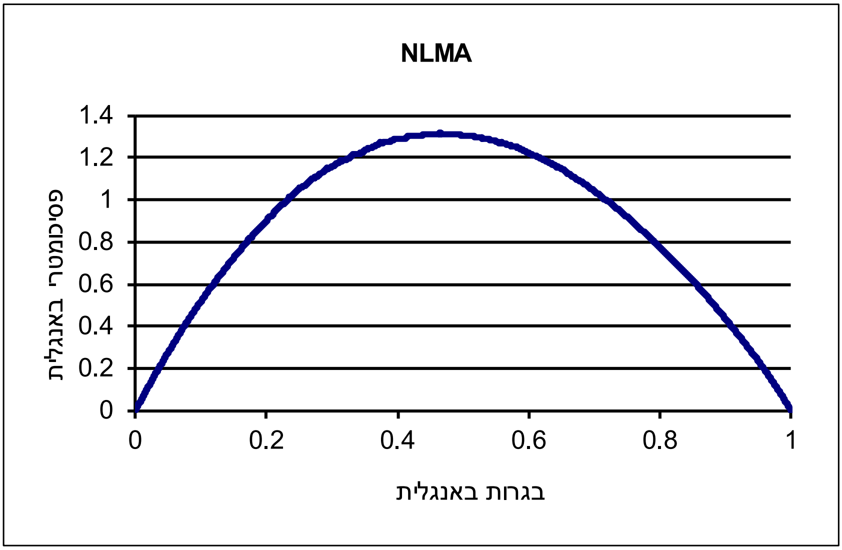
ציור 5.3.2: התפלגות ציוני פסיכומטרי ובגרות באנגלית – 2000–1999



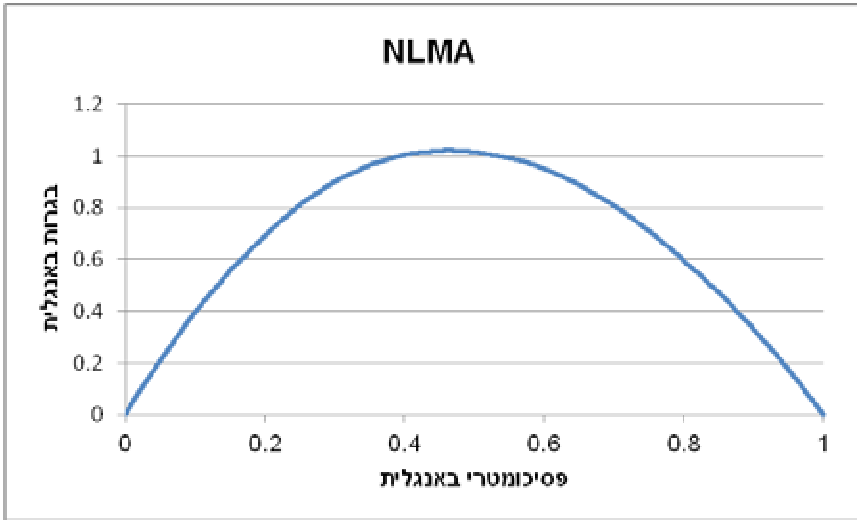
ציור 5.3.3: ציון פסיכומטרי באנגלית כפונקציה של ציון בגרות באנגלית – 2010–2009



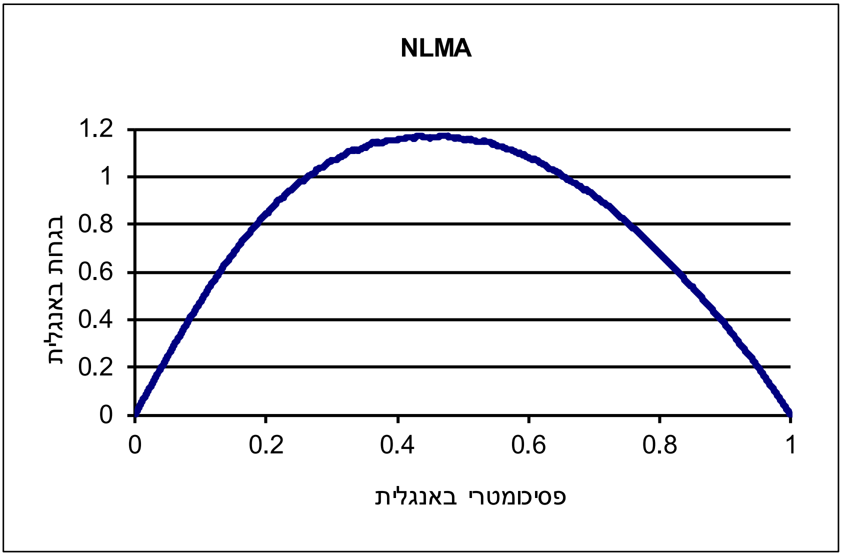
ציור 5.3.4: ציון פסיכומטרי באנגלית כפונקציה של ציון בגרות באנגלית – 2000–1999



ציור 5.3.5: ציון בגרות באנגלית כפונקציה של ציון פסיכומטרי באנגלית – 2010–2009



ציור 5.3.6: ציון בגרות באנגלית כפונקציה של ציון פסיכומטרי באנגלית – 2000–1999



לסיום ההשוואה, לוחות 5.3.3 ו-5.3.4 מציגים את מטריצת המעבר בין הציון הפסיכומטרי לציון הבגרות.

לוח 5.3.3: מטריצת מעבר: עשירוני ציונים – בגרות באנגלית ופסיכומטרי באנגלית – 2009–2010

		בגרות באנגלית										
פסיכומטרי באנגלית		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	סה"כ
	1	4.9	2.6	1.4	0.6	0.3	0.1					10
	2	2.8	3.1	1.9	1.2	0.6	0.3	0.1				10
	3	1.5	2.1	2.4	2	1.2	0.6	0.2	0.1			10
	4	0.5	1.3	2.1	2.2	1.8	1.2	0.5	0.2	0.1		10
	5	0.2	0.5	1.2	1.9	2.1	1.8	1.4	0.6	0.2		10
	6	0.1	0.2	0.6	1.2	1.8	2.1	1.9	1.2	0.6	0.2	10
	7		0.1	0.2	0.6	1.3	1.9	2	2	1.3	0.5	10
	8			0.1	0.2	0.5	1.1	2	2.4	2.3	1.3	10
	9				0.1	0.3	0.6	1.4	2.2	2.9	2.5	10
	10						0.2	0.5	1.2	2.6	5.9	10
	סה"כ	10	10	10	10	10	10	10	10	10	10	100

לוח 5.3.4: מטריצת מעבר: עשירוני ציונים – בגרות באנגלית ופסיכומטרי באנגלית – 1999–2000

		בגרות באנגלית										
פסיכומטרי באנגלית		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	סה"כ
	1	5.2	2.6	1.3	0.5	0.3	0.1	0.1				10
	2	3.1	3.1	2.1	1.2	0.4	0.2	0.1				10
	3	1.1	2.2	2.5	2.1	1.2	0.6	0.2	0.1			10
	4	0.4	1	1.6	1.9	2.1	1.6	0.9	0.4	0.1		10
	5	0.1	0.6	1.3	1.8	1.9	1.9	1.3	0.7	0.3	0.1	10
	6	0.1	0.3	0.7	1.3	1.7	1.9	1.8	1.3	0.8	0.2	10
	7		0.2	0.3	0.8	1.3	1.9	1.8	1.7	1.4	0.6	10
	8			0.1	0.2	0.7	1	1.7	2.4	2.4	1.5	10
	9			0.1	0.1	0.4	0.7	1.4	2	2.6	2.6	10
	10				0.1	0.1	0.3	0.8	1.4	2.3	5	10
	סה"כ	10	10	10	10	10	10	10	10	10	10	100

כפי שניתן לראות, כ-30 אחוזים מהנבחנים נשארים באותו עשירון של ציונים לעומת 28.45 אחוזים בשנת 1999–2000. שאר הנבחנים מתחלקים בצורה כמעט שווה מעל ומתחת לאלכסון. את הגדלת המתאם בין מערכות הציונים ניתן לראות גם מכך שבניגוד לסעיף הקודם שבו עברו אנשים מעשירון עליון באחד הציונים לעשירון תחתון בציון השני הרי שהתאים המרוחקים מהאלכסון נותרו ריקים. מכאן שכאשר אנחנו מצרפים תלמידי 5 יחידות לתלמידי 4 יחידות המתאם בין שתי מערכות הציונים עולה.

השונות של סכום או ממוצע של שני משתנים נמוכה יותר ככל שהמתאם בין המשתנים קטן יותר. המתאמים הגבוהים בין שתי מערכות הציונים מעידים על כך שאין "רווח" גדול ממיצוע הציונים, וזאת משום שהרווח ממיצוע גבוה יותר ככל שהמתאם בין הציונים נמוך יותר.

ה.4. מתמטיקה – 3, 4 ו–5 יחידות בגרות

לאור המתאם הגבוה שקיבלנו באנגלית, החלטנו לבדוק את כל התלמידים שהשתתפו בבחינה במתמטיקה של 3, 4 ו-5 יחידות, כלומר אנו מצרפים את ציוני כל התלמידים שעומדים בדרישות הכניסה לאוניברסיטה במתמטיקה. המשמעות היא שאנו משווים תוצאות מבחן כללי אחד במבחן יחיד שני הנערך לכלל התלמידים הזכאים. מספר התלמידים שנבחנו בשתי הבחינות ושעומדים בדרישות הקבלה לאוניברסיטאות הוא 31,392 ו-17,867 בשנים 2009–2010 ו-1999–2000, בהתאמה. השקלול של הציונים הוא בהתאם לשקלול שמעניקות האוניברסיטאות.

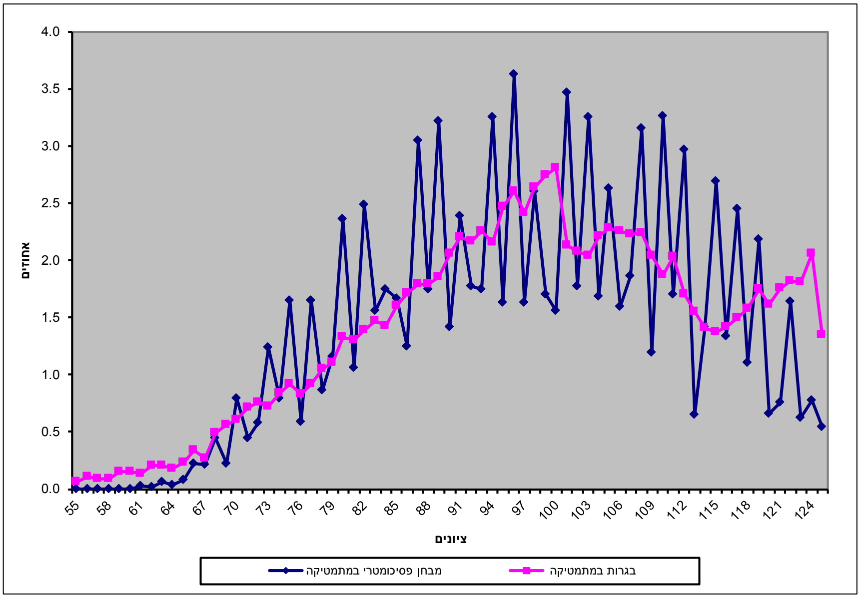
לוח 5.4.1: פרמטרים מרכזיים של שני המבחנים במתמטיקה ב-3, 4 ו-5 יחידות – 2009–2010

N=31,392	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם גיני Γyx	מתאם ג'יני Γxy
בגרות במתמטיקה	98.5	99.0	55	125	15.33	0.088	0.70 (0.000)	0.71 (0.000)	0.706	0.708
פסיכומטרי במתמטיקה	97.7	97.7	55	125	14.12	0.083				

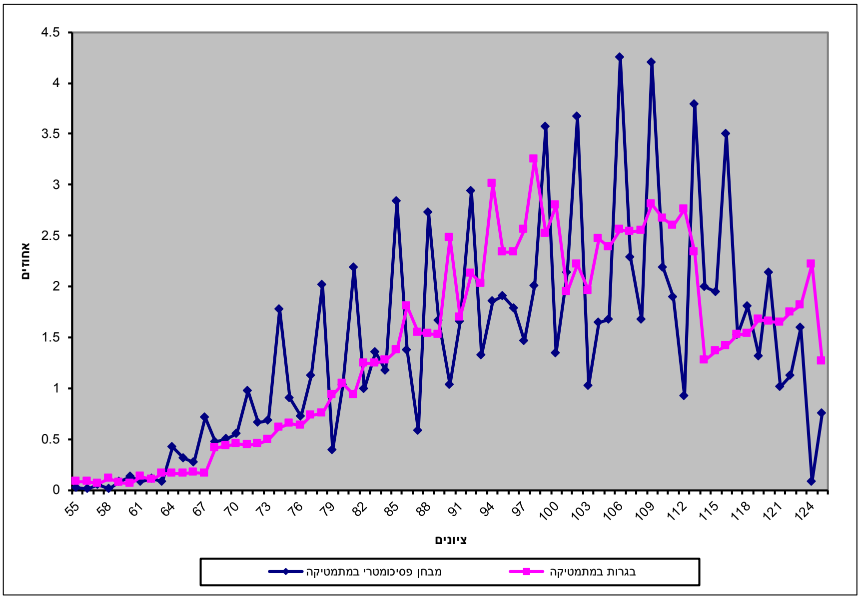
לוח 5.4.2: פרמטרים מרכזיים של שני המבחנים במתמטיקה ב-3, 4 ו-5 יחידות – 1999–2000

N=17,867	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם גיני Γyx	מתאם ג'יני Γxy
בגרות במתמטיקה	100.3	100.5	55	125	14.14	0.082	0.59 (0.000)	0.60 (0.000)	0.60	0.60
פסיכומטרי במתמטיקה	98.7	101.0	55	125	15.13	0.088				

ציור 5.4.1: התפלגות ציוני פסיכומטרי ובגרות במתמטיקה – 2009–2010



ציור 5.4.2: התפלגות ציוני פסיכומטרי ובגרות במתמטיקה – 1999–2000



השוואת הציון הממוצע והחציוני בבגרות מראה על ירידה בציון על פני הזמן הן בבגרות והן בפסיכומטרי. השוואת הציונים שהתקבלו בתקופה המוקדמת יותר (לוח 5.4.2) בין הבגרות לפסיכומטרי מראה שהציון בבגרות גבוה בממוצע ב-1.5 נקודות, אולם ממצא זה אינו משמעותי מאחר שהוא תלוי בצורת הנרמול שנעשתה. סטיית התקן ומקדם ג'יני מראים שהפיזור של הציון גדול יותר בפסיכומטרי, אולם ממצא זה משתנה על פני השנים. מקדמי המתאם מראים מידה קטנה יותר של מתאם בין הציונים מכפי שמצאנו באנגלית, אם כי גם כאן ניתן להבחין בעלייה משמעותית של מקדמי המתאם בין התקופות. מקדמי המתאם של הג'יני שווים, מה שמעיד שיש לצפות לסימטריה בהתפלגות הציונים בשתי הבחינות.

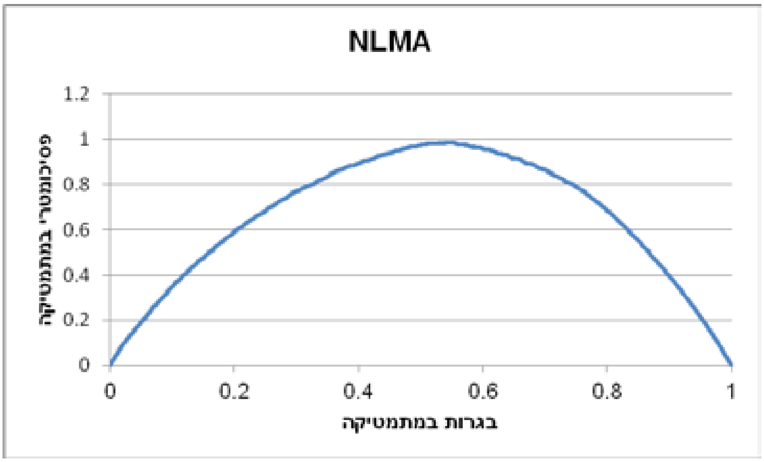
ציורים 5.4.1 ו-5.4.2 מצביעים על שכיחות גבוהה יותר בציוני הבגרות הגבוהים ועל כך שציוני הבגרות מתפלגים בצורה רציפה יותר מאשר ציוני הפסיכומטרי.

ציורים 5.4.3 ו-5.4.4 מציגים את עקומת ה-NLMA של ציון הפסיכומטרי במתמטיקה כפונקציה של ציון הבגרות. ניתן לראות עקומה קעורה וחלקה, מה שמעיד על קשר מונוטוני חזק של הציון הפסיכומטרי כפונקציה של ציון הבגרות.

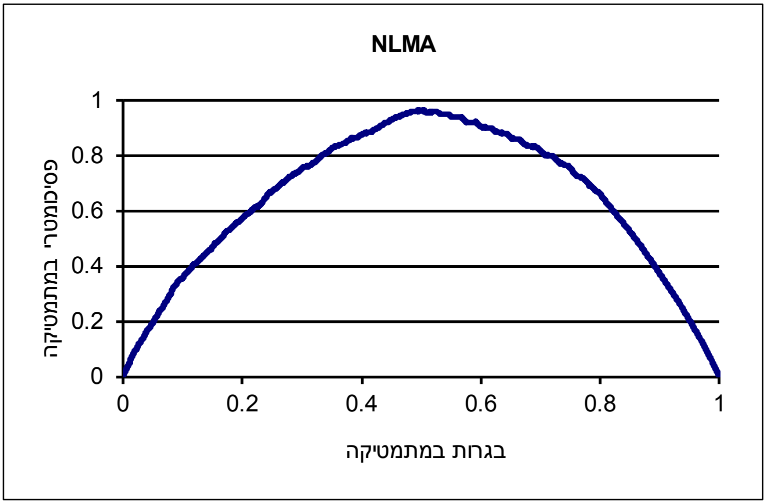
ציורים 5.4.5 ו-5.4.6 מציגים את עקומת ה-NLMA של הציון במתמטיקה בבגרות כפונקציה של הציון הפסיכומטרי. כצפוי קיבלנו קשר מונוטוני יציב שמשמעותו ששני המבחנים בוחנים את אותו הנושא.

לוחות 5.4.3 ו-5.4.4 מציגים את מטריצות המעבר ממערכת ציונים אחת לשנייה. ניתן לראות שרק 23.8 (20.2 ב-1999–2000) אחוזים מהנבחנים נשארים על האלכסון, כלומר רק כ-20 אחוזים מהתלמידים נשארים באותו עשירון של ציונים. השאר נמצאים באופן דומה מעל ומתחת לאלכסון. ניתן לראות שהפיזור של התלמידים מעבר לאלכסון גבוה יותר במתמטיקה מאשר באנגלית.

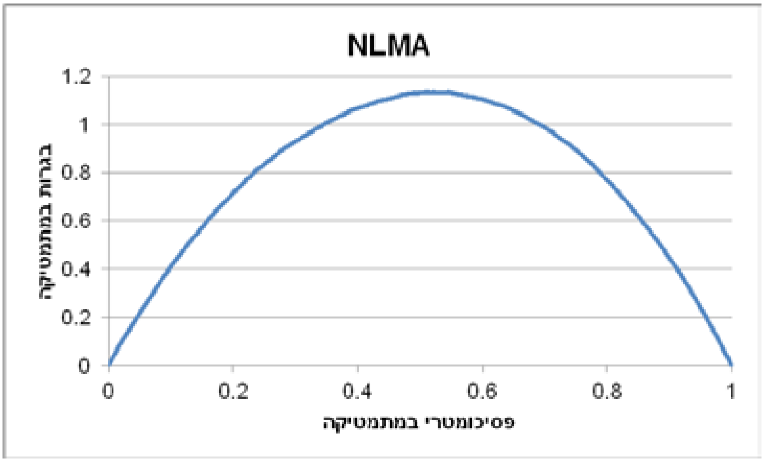
ציור 5.4.3: ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה – 2009–2010



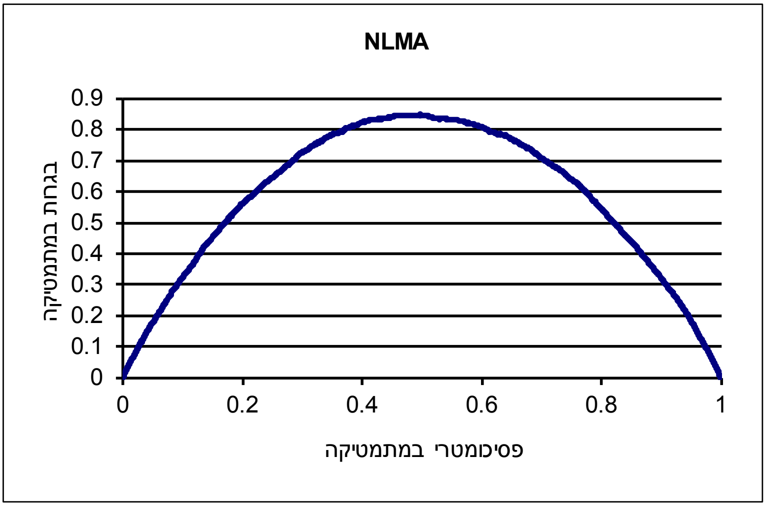
ציור 5.4.4: ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה – 2000–1999



ציור 5.4.5: ציון בגרות במתמטיקה כפונקציה של ציון פסיכומטרי במתמטיקה – 2010–2009



ציור 5.4.6: ציון בגרות במתמטיקה כפונקציה של ציון פסיכומטרי במתמטיקה – 2000–1999



לוח 5.4.3: מטריצת מעבר: עשירוני ציונים – בגרות במתמטיקה ופסיכומטרי במתמטיקה – 2010–2009

פסיכומטרי במתמטיקה	בגרות במתמטיקה										
	1	2	3	4	5	6	7	8	9	10	סה"כ
1	3.7	2.2	1.6	1.1	0.8	0.4	0.2	0.1			10
2	2.5	2.2	1.8	1.4	1	0.6	0.4	0.2	0.1		10
3	1.6	1.8	1.6	1.5	1.5	1	0.6	0.3	0.1	0.1	10
4	0.9	1.3	1.6	1.6	1.8	1.3	0.8	0.5	0.2	0.1	10
5	0.7	1.1	1.2	1.4	1.5	1.5	1.2	0.9	0.4	0.1	10
6	0.3	0.6	0.9	1.2	1.2	1.5	1.7	1.4	0.8	0.3	10
7	0.2	0.5	0.8	0.9	1.1	1.7	1.7	1.6	1.1	0.6	10
8	0.1	0.2	0.3	0.5	0.8	1	1.7	2.2	2.1	1.2	10
9		0.1	0.2	0.4	0.4	0.8	1.2	1.7	2.4	2.7	10
10			0.1	0.1	0.2	0.3	0.6	1.1	2.7	5.2	10
סה"כ	10	10	10	10	10	10	10	10	10	10	100

לוח 5.4.4: מטריצת מעבר: עשירוני ציונים – בגרות במתמטיקה ופסיכומטרי במתמטיקה – 1999–2000

פסיכומטרי במתמטיקה	בגרות במתמטיקה										
	1	2	3	4	5	6	7	8	9	10	סה"כ
1	3.5	2	1.6	1.1	0.9	0.4	0.3	0.2			10
2	2.3	1.6	1.7	1.5	1.3	0.6	0.5	0.4	0.1		10
3	1.3	1.8	1.5	1.5	1.4	0.9	0.7	0.6	0.3	0.1	10
4	1.1	1.5	1.4	1.3	1.2	1.2	1.1	0.7	0.3	0.2	10
5	0.6	0.9	1.1	1.3	1.3	1.3	1.3	1.1	0.6	0.4	10
6	0.4	0.7	0.9	1	1.3	1.3	1.3	1.3	1.1	0.6	10
7	0.4	0.7	0.8	1.1	1.3	1.6	1.1	1.2	1.1	0.8	10
8	0.2	0.4	0.4	0.5	0.5	1.1	1.8	2.1	1.7	1.3	10
9	0.2	0.3	0.4	0.5	0.5	0.9	0.9	1.3	2.5	2.7	10
10	0.1	0.2	0.2	0.3	0.3	0.7	0.8	1.2	2.3	4	10
סה"כ	10	10	10	10	10	10	10	10	10	10	100

ו. מידת השיפור בעמידות (robustness) הציון

בסעיף הקודם הראינו שמבחני הבגרות והפסיכומטרי בתחומי המתמטיקה והאנגלית בודקים את אותו תחום כישורים, וזאת עד כמה שיכולת הבחינה הסטטיסטית מאפשרת לבדוק. מסקנתנו התבססה על כך שהציונים במבחנים מהווים טרנספורמציה מונוטונית האחד של השני: כלומר ככל שהציון בבגרות גבוה יותר כן יעלה הציון בפסיכומטרי, ולהפך: ככל שהציון בפסיכומטרי גבוה יותר כן יש לצפות שהציון בבגרות יהיה גבוה יותר. בסעיף זה נרצה לבדוק את הטענה שהצירוף של ציונים של שני המבחנים – הפסיכומטרי והבגרות – משפר את העמידות (robustness) של הציון, וזאת בבחינת "טובים השניים מן האחד". עמידות מבחינתנו משמעותה הקטנת השונות בציון שמקבל התלמיד.

לצורך בדיקה של השערה זו אנחנו מבצעים הערכה של התרומה של הציון בבחינה הפסיכומטרית ובבגרות לציון הממוצע של הנבחן בשתי הבחינות. אם תוצאות הבחינות הן בלתי תלויות סטטיסטית אזי המיצוע של הציונים יקטין את הפיזור (שונות או ג'יני) של ממוצע הציונים בחצי יחסית לשונות ההתחלתית. לעומת זאת בהנחה שהמתאם בין ציוני שתי הבחינות הוא אחד אזי אין צירוף של ציוני המבחנים מקטין את הפיזור של הציון הממוצע אלא משאיר אותו באותה רמה. היות שהמתאמים שמצאנו בין הציון שהנבחן מקבל בפסיכומטרי לבין הציון שהוא מקבל בבגרות הם גבוהים יחסית, וגדולים יותר באנגלית מאשר במתמטיקה, אזי הצפייה שלנו היא שהתרומה של צירוף המבחנים באנגלית להורדת הפיזור של הציון הממוצע לתלמיד בשני המבחנים תהיה קטנה יותר מהצירוף של הציונים במתמטיקה. כמו כן, המתאמים לשנת 2009–2010 נמצאו גבוהים יתר בשני המקצועות יחסית לשנת 1999–2000, ועל כן גם אם היה בסיס לטיעון זה בעבר הרי שעם הזמן ירדה תקפותו.

הנוסחה שעל פיה ניתן לפרק את מדד ג'יני היחסי כוללת כמקרה פרטי את המבנה של הפירוק של מקדם ההשתנות, ועל כן נשתמש בה.

הנוסחה היא:[23]

יהי $Y = X_1 + X_2$ ותהי $\delta_k = \mu_k/\mu_Y$ התרומה היחסית של הציון במבחן לסכום הציונים של שני המבחנים (הציון הכולל), אזי

$$G_Y^2 - G_Y\sum_{k=1}^{2}\delta_k D_{kY} G_k = \sum_{k=1}^{2}\delta_k^2 G_k^2 + \delta_1\delta_2 G_k G_j(\Gamma_{12} + \Gamma_{21}) \quad (6.1)$$

כאשר $D_{kY} = \Gamma_{kY} - \Gamma_{Yk}$ ו $\Gamma_{12} = \dfrac{\text{cov}(X_1,F(X_2))}{\text{cov}(X_1,F(X_1))}$ הוא מקדם המתאם המותאם למדד ג'יני.

אם $D_{kY} = 0$ וכן $\Gamma_{21} = \Gamma_{12}$ אזי הנוסחה המתקבלת היא

$$G_Y^2 = \sum_{k=1}^{2}\delta_k^2 G_k^2 + 2\,\delta_1\delta_2 G_1 G_j \Gamma_{12} \quad (6.2)$$

כאשר נוסחה (6.2) מקבילה לנוסחה שהייתה מתקבלת אם היינו מבקשים לראות את התרומה של כל מרכיב למקדם ההשתנות של הציון הכולל.[24] נוסחה (6.1) המתאימה לג'יני כוללת כמקרה פרטי את נוסחה (6.2), והיתרון שלה הוא שהיא אינה מניחה מקדם מתאם סימטרי בין שתי ההתפלגויות של הציונים.

לוח 6.1 מציג את המרכיבים של מדד ג'יני בציון הכולל של שני המבחנים באנגלית לשנת 2009–2010 ואילו לוח 6.2 מבצע את אותו פירוק לשנת 1999–2000. כפי שניתן לראות, הפיזור על פי מדד ג'יני בפסיכומטרי בשתי התקופות גדול יותר מאשר בבגרות, ופער זה הלך והתרחב על פני הזמן, ואילו הציון הממוצע גבוה מעט בבגרות מאשר בפסיכומטרי בשנת 2009–2010, ואילו ב-1999–2000 המצב הפוך, דבר המעיד על זחילת הציון כלפי מעלה בבגרות. בשתי השנים המדד לציון המשוקלל קטן רק במידה מזערית מהמדד בבגרות, תוצאה שניתן לייחס למתאם הגבוה שנמצא בין ציוני שני המבחנים.

החלק השני בכל לוח מציג את מקדמי המתאם בין הציון שקיבל הנבחן בבגרות מול הציון בפסיכומטרי. כפי שניתן לראות, הבדלי המתאמים הם קטנים ביחס, מה שמעיד על סימטריה בהתפלגויות.[25] גם המתאם עם הציון הכללי נראה דומה בשתי מערכות הציונים, ומכאן אנו מקבלים שמאחר שהציון הכולל מתבסס על מתאמים גבוהים בין הציונים הבודדים הרי שאין רווח גדול מקיומם של שני מבחנים הדומים בתוצאות הדירוג של הנבחנים.

השורה התחתונה מציגה את התרומות של כל מרכיב בפירוק של מדד ג'יני בציון הכולל. התרומה מתחלקת בערך לכ-60 אחוזים למרכיבים העצמאיים של הציון ו-40 אחוזים למתאם בין הציונים. המתאם בין תוצאות שני המבחנים משפיע על תוצאת הציון הכולל.

23 פיתוח הנוסחה נמצא ב-Yitzhaki and Schechtman (2013).

24 מקדם ההשתנות הוא סטיית התקן מחולקת בממוצע של המשתנה.

25 תוצאה זו יכולה לנבוע גם מכך ששתי מערכות הציונים עוברות תהליך דומה של נרמול.

לוח 6.1: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות באנגלית מול פסיכומטרי באנגלית – 2009–2010

	בגרות באנגלית	פסיכומטרי באנגלית	סה"כ
G	0.066	0.090	0.074
משקל יחסי δ	0.523	0.477	
Γ_{ij}	**בגרות באנגלית**	**פסיכומטרי באנגלית**	**סה"כ**
בגרות באנגלית		0.843	0.950
פסיכומטרי באנגלית	0.844		0.968
סה"כ	0.952	0.968	
G_O^2	**$G_O \Sigma \delta_i D_{iO} G_i$**	**$\delta_i^2 G_i^2$**	**$\Sigma\Sigma\ \delta_i \delta_j G_i G_j \Gamma_{ij}$**
0.0055	0.000006	0.0030	0.0025

לוח 6.2: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות באנגלית מול פסיכומטרי באנגלית – 1999–2000

	בגרות באנגלית	פסיכומטרי באנגלית	סה"כ
G	0.077	0.082	0.076
משקל יחסי δ	0.498	0.502	
Γ_{ij}	**בגרות באנגלית**	**פסיכומטרי באנגלית**	**סה"כ**
בגרות באנגלית		0.827	0.953
פסיכומטרי באנגלית	0.834		0.962
סה"כ	0.953	0.959	
G_O^2	**$G_O \Sigma \delta_i D_{iO} G_i$**	**$\delta_i^2 G_i^2$**	**$\Sigma\Sigma\ \delta_i \delta_j G_i G_j \Gamma_{ij}$**
0.00578	-0.00001	0.00316	0.00262

לוחות 6.3 ו-6.4 בודקים את מדד ג'יני של הציון במתמטיקה בבגרות לעומת הציון בבגרות באנגלית לשתי התקופות. המשקל היחסי מציג את הציון בכל מבחן בממוצע של סכום תוצאות המבחנים, מכאן שניתן לראות שעליית הציון הממוצע על פני הזמן באנגלית גבוהה יותר מאשר עליית הציון במתמטיקה. עוד ניתן לראות שאי השוויון בציוני המתמטיקה גבוה יותר מאי השוויון באנגלית, ואילו אי השוויון בסכום תוצאות המבחנים נמוך מאי השוויון בציונים בכל מבחן. הסבר לתוצאה זו ניתן לקבל מהמתאם היחסית נמוך בין הציונים (0.5), אם כי מתאם זה עלה בין שתי תקופות הזמן. הסימטריה של המתאם מעידה שההתפלגות נורמלה בשני המבחנים בצורה דומה. קיימת גם

סימטריה במתאם בין כל ציון לסכום הציון בשני המבחנים. המתאם הנמוך יחסית אומר שקיים טעם בהסתכלות על סכום הציונים במתמטיקה ואנגלית בבגרות, מאחר שהם מוסיפים אינפורמציה שונה על התלמיד. המתאם הנמוך יחסית גם מסביר מדוע קיימת ירידה באי השוויון בסכום הציונים יחסית לכל ציון בנפרד, מכאן שלקיחה בחשבון של הצלחה במתמטיקה ובאנגלית יכולה לתרום יותר להבנת היכולות של התלמידים בפסיכומטרי ובבגרות בכל תחום.

לוח 6.3: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות באנגלית מול בגרות במתמטיקה – 2010–2009

	בגרות באנגלית	בגרות במתמטיקה	סה"כ
G	0.066	0.089	0.069
משקל יחסי δ	0.517	0.483	
Γ_{ij}	**בגרות באנגלית**	**בגרות במתמטיקה**	**סה"כ**
בגרות באנגלית		0.593	0.864
בגרות במתמטיקה	0.601		0.920
סה"כ	0.864	0.919	
G_O^2	**$G_O \Sigma \delta_i D_{iO} G_i$**	**$\delta_i^2 G_i^2$**	**$\Sigma\Sigma\ \delta_i \delta_j G_i G_j \Gamma_{ij}$**
0.0048	-0.000003	0.0030	0.0018

לוח 6.4: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות באנגלית מול בגרות במתמטיקה – 2000–1999

	בגרות באנגלית	בגרות במתמטיקה	סה"כ
G	0.077	0.082	0.068
משקל יחסיδ	0.503	0.497	
Γ_{ij}	**בגרות באנגלית**	**בגרות במתמטיקה**	**סה"כ**
בגרות באנגלית		0.496	0.861
בגרות במתמטיקה	0.496		0.874
סה"כ	0.853	0.871	
G_O^2	**$G_O \Sigma \delta_i D_{iO} G_i$**	**$\delta_i^2 G_i^2$**	**$\Sigma\Sigma\ \delta_i \delta_j G_i G_j \Gamma_{ij}$**
0.00468	-0.00003	0.00315	0.00156

לוח 6.5: פירוק מדד ג'יני בציון הכולל לג'יני של פסיכומטרי באנגלית מול פסיכומטרי במתמטיקה 2009–2010

	פסיכומטרי באנגלית	פסיכומטרי במתמטיקה	סה"כ
G	0.090	0.083	0.078
משקל יחסי δ	0.496	0.504	
Γ_{ij}	**פסיכומטרי באנגלית**	**פסיכומטרי במתמטיקה**	**סה"כ**
פסיכומטרי באנגלית		0.637	0.911
פסיכומטרי במתמטיקה	0.647		0.904
סה"כ	0.912	0.898	
G_O^2	$G_O\Sigma\delta_iD_{iO}G_i$	$\delta_i^2G_i^2$	$\Sigma\Sigma\ \delta_i\delta_jG_iG_j\Gamma_{ij}$
0.0061	-0.00002	0.0037	0.0024

לוח 6.6: פירוק מדד ג'יני בציון הכולל לג'יני של פסיכומטרי באנגלית מול פסיכומטרי במתמטיקה 1999–2000

	פסיכומטרי באנגלית	פסיכומטרי במתמטיקה	סה"כ
G	0.082	0.087	0.077
משקל יחסי δ	0.509	0.491	
Γ_{ij}	**פסיכומטרי באנגלית**	**פסיכומטרי במתמטיקה**	**סה"כ**
פסיכומטרי באנגלית		0.675	0.914
פסיכומטרי במתמטיקה	0.680		0.922
סה"כ	0.911	0.916	
G_O^2	$G_O\Sigma\delta_iD_{iO}G_i$	$\delta_i^2G_i^2$	$\Sigma\Sigma\ \delta_i\delta_jG_iG_j\Gamma_{ij}$
0.00596	-0.00003	0.00357	0.00242

בלוחות 6.5 ו-6.6 מבוצעת חזרה על לוחות 6.3 ו-6.4, כאשר הפעם ההסתכלות היא על סכום הציונים בפסיכומטרי באנגלית ובמתמטיקה. אי השוויון בציוני הפסיכומטרי גם באנגלית וגם במתמטיקה גבוה יותר יחסית לאי השוויון המקביל בציוני הבגרות, אולם הציון הממוצע בפסיכומטרי באנגלית עלה פחות על פני השנים יחסית לציון בפסיכומטרי במתמטיקה, והתוצאה היא עליית המשקל של הציון הממוצע במתמטיקה בשנת 2009 יחסית לשנת 1999. גם המתאם בציונים באנגלית ובמתמטיקה בפסיכומטרי גבוה יותר מאשר בציוני הבגרות, וכתוצאה מכך גם המתאם בין כל מרכיב מול הציון הכולל. ייתכן שניתן לייחס ממצא זה לדמיון שבמבחני הפסיכומטרי במתמטיקה ובאנגלית יחסית לבגרויות באותם תחומים. בכל מקרה התוצאה הינה שמשקל כל מרכיב בציון הכולל גבוה יותר מהמשקל בציוני הבגרות המקבילים.

לוחות 6.7 ו-6.8 חוזרים על הלוחות הקודמים, כאשר הפעם ההתייחסות היא לציון של בגרות במתמטיקה מול פסיכומטרי במתמטיקה. כמעט לא חל שינוי במשקל היחסי של הציונים, מה שכנראה מעיד על נרמול של הציונים על פני זמן. בשתי השנים, אי השוויון בסכום הציונים נמוך יחסית מאי השוויון בתוך כל מרכיב, אם כי ההבדל שמתקבל נמוך יותר מאשר המקבילה באנגלית. המתאם בין הציונים עלה על פני השנים, מה שגורם שההקטנה באי השוויון בציון הכולל קטנה על פני הזמן.

המסקנות המתקבלות מתוך הניתוח דלעיל הן שהמבחנים הפסיכומטריים במתמטיקה ובאנגלית דומים יותר בממצאיהם מאשר הציונים המקבילים בבגרות. לעומת זאת אי השוויון המתקבל בציוני הפסיכומטרי, גם באנגלית וגם במתמטיקה, הוא גדול יותר בפסיכומטרי מאשר בבגרות. ייתכן שתוצאה זו מוסברת בכך ששיטות הבחינה במתמטיקה ובאנגלית בפסיכומטרי קרובות יותר מאשר שיטות הבחינה באותם מקצועות בבגרות, ואם ההשערה דלעיל נכונה אזי ניתן להסיק ששיטת הבחינה יש לה השפעה על המתאם המתקבל בהצלחות בבחינות השונות.

לוח 6.7: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות מתמטיקה מול פסיכומטרי במתמטיקה 2010–2009

	בגרות במתמטיקה	פסיכומטרי במתמטיקה	סה"כ
G	0.089	0.083	0.080
משקל יחסי δ	0.502	0.498	
Γ_{ij}	**בגרות במתמטיקה**	**פסיכומטרי במתמטיקה**	**סה"כ**
בגרות במתמטיקה		0.706	0.929
פסיכומטרי במתמטיקה	0.718		0.924
סה"כ	0.929	0.918	
G_O^2	**$G_O \Sigma \delta_i D_{iO} G_i$**	**$\delta_i^2 G_i^2$**	**$\Sigma\Sigma\ \delta_i \delta_j G_i G_j \Gamma_{ij}$**
0.0064	-0.00002	0.0037	0.0026

לוח 6.8: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות מתמטיקה מול פסיכומטרי במתמטיקה 1999–2000

	בגרות במתמטיקה	**פסיכומטרי במתמטיקה**	**סה"כ**
G	0.082	0.087	0.076
משקל יחסי δ	0.504	0.496	
Γ_{ij}	**בגרות במתמטיקה**	**פסיכומטרי במתמטיקה**	**סה"כ**
בגרות במתמטיקה		0.612	0.894
פסיכומטרי במתמטיקה	0.617		0.908
סה"כ	0.889	0.903	
G_O^2	$G_O \Sigma \delta_i D_{iO} G_i$	$\delta_i^2 G_i^2$	$\Sigma\Sigma\, \delta_i \delta_j G_i G_j \Gamma_{ij}$
0.00574	-0.00003	0.00358	0.00219

ז. סיכום הממצאים

מחקר זה מהווה מחקר מקדים לשאלה רחבה יותר שצריכה להטריד את מעצבי המדיניות בתחום החינוך: באיזו מידה מהווה הקיום של שתי בחינות באותם תחומים נטל עודף על המשק הישראלי. תשובה לשאלה זו יכולה להתקבל ממבחן של עלות-תועלת מקיומם של שני מבחנים. במחקר זה בדקנו שאלה מקדימה לבחינה של עלות-תועלת והיא באיזו מידה המבחנים בודקים את אותו תחום יכולות.

התשובה המתקבלת ממחקר זה היא שהמבחנים בודקים את אותו תחום יכולות, אולם קיימת בהם טעות אקראית גדולה. הטעות האקראית במבחני המתמטיקה גדולה יותר מאשר במבחנים באנגלית, ועל כן ניתוח של עלות-תועלת כדאי לבצע עבור המבחן באנגלית, כי שם הכפילות שיוצרים שני המבחנים גדולה יותר.

המסקנות המתקבלות מהתרומות של שתי הבחינות לציון הכולל של כל תלמיד מחזקות את ההשערה שהתרומה של כפל הבחינות אינו תורם רבות, ועל כן מתעורר הספק לגבי התרומה של קיום שני מבחנים לבחינת אותם מקצועות. המסקנה דלעיל מתקיימת ביתר שאת לגבי מקצוע האנגלית שבו המתאם בין הציון בבגרות לפסיכומטרי גבוה יותר מאשר במתמטיקה.

מסקנה זו מתחזקת על פני הזמן, וזאת כי המתאמים בין תוצאות המבחנים עולים על פני הזמן.

כדאי להדגיש שאין במחקר כדי להכריע בשאלה הבאה: נניח שקיבלנו את תוצאות המחקר ומצאנו גם שקיים נטל מיותר על האוכלוסייה, על איזה מבחן כדאי לוותר? תשובה לשאלה זו מחייבת בדיקת המטרות והשימושים של המבחנים, הצרכים של מערכת ההשכלה הגבוהה ועוד שיקולים נוספים. מה שהמחקר קובע הוא שמבחני הבגרות והפסיכומטרי במקצועות המקבילים (אנגלית ומתמטיקה) אינם בודקים תחומים שונים שיש לנו יכולת להבחין ביניהם, וזאת על סמך ניתוח מערכת התוצאות של המבחנים.

מקורות

Lord F.M. and Novick M.R. (1968), *Statistical Theories of Mental Test Scores*, Readind, Massachusetts.

Schechtman E. and Yitzhaki S. (2009), "Ranking Groups’ Abilities – Is It Always Reliable?", *International Journal of Testing* 9(3), 195–214.

Yitzhaki S. (2003), "Gini’s Mean Difference: A Superior Measure of Variability for Non-normal Distributions", *Metron* LXI(2), 285–316.

Yitzhaki S. and Eisenstaedt M. (2003), "Ranking Groups' versus Individuals' Ranking", Y. Amiel and J.A. Bishop (eds.), *Fiscal Policy, Inequality, and Welfare, Research on Economic Inequality*, 10, Amsterdam, 101–123.

Yitzhaki S., Itzhaki R. and Pudalov T. (2012), "A Nonparametric ICC Using the Gini's Mean Difference Approach", Unpublished, http://ssrn.com

Yitzhaki S. and Schechtman E. (2012), "Identifying Monotonic and Non-monotonic Relationship", *Economics Letters* 116, 23–25.

Yitzhaki S. and Schechtman E. (2013), *The Gini Methodology: A Primer on a Statistical Methodology*, New York.