סדרת ניירות עבודה
WORKING PAPER SERIES

No. 75 'מס

מבחני הבגרות והפסיכומטרי במתמטיקה ואנגלית
והקשר ביניהם

Matriculation and Psychometric Examinations
in Mathematics and English,
and the Relationship between Them

שלמה יצחקי*, טאינה פודלוב** ואביאל קרנצלר**
Shlomo Yitzhaki*, Taina Pudalov** and Aviel Krentzler **

סיון תשע"ג, מאי May 2013

*הלשכה המרכזית לסטטיסטיקה והאוניברסיטה העברית
**הלשכה המרכזית לסטטיסטיקה
* Central Bureau of Statistics and Hebrew University
** Central Bureau of Statistics

הלשכה המרכזית לסטטיסטיקה
Central Bureau of Statistics
دائرة الإحصاء المركزية

הלשכה המרכזית לסטטיסטיקה (הלמ"ס) מעודדת מחקר המבוסס על נתוני הלמ"ס, כדוגמת עבודה זו. עבודות מחקר אלו אינן פרסומים רשמיים של הלמ"ס, ומכאן שהדעות והמסקנות הבאות בהן לידי ביטוי, הן של המחברים עצמם ואינן משקפות בהכרח את הדעות והמסקנות של הלמ"ס.

הוצאת הלשכה המרכזית לסטטיסטיקה, רח' כנפי נשרים 66, פינת רח' בקי,
ת"ד 34525, ירושלים 91342
טל': 02-6592666; פקס: 02-6521340
אתר הלמ"ס באינטרנט: www.cbs.gov.il
דואר אלקטרוני: info@cbs.gov.il

תקציר

מזה שנים רבות מתקיימות בארץ הן בחינות הבגרות והן הבחינות הפסיכומטריות. הצורך להיבחן בשתי בחינות מהווה נטל הן על התלמידים הנדרשים להתכונן לשני מבחנים שמתכונותיהן שונות, והן על המשק הישראלי. מחקר זה מהווה מחקר מקדים לבחינת שאלת עלות תועלת של קיום 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 שאנחנו שואלים שאלות ובודקים מי ענה עליהם. 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 לעומת 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 שהבחינות בודקות תחומי יידע שונים או שהתכונות הנדרשות להצלחה בבחינה אחת שונות מאשר התכונות הנדרשות להצלחה בבחינה שנייה.

התשובה המתקבלת ממחקר זה הוא 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 לפסיכומטרי גבוה יותר מאשר במתמטיקה.

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

מילות מפתח: בחינות בגרות, מבחן פסיכומטרי, מונוטוניות, רגרסיית ג'יני.

אנחנו מודים לדמיטרי רומנוב ולקורא אלמוני על הביקורת על טיוטא ראשונית של המאמר שסייעו רבות בשיפור ההצגה של המאמר.

תוכן העניינים

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 לצרכי קבלה ללימודים אקדמיים באוניברסיטאות. מערכת בחינות כפולה הנערכות במתכונת שונה, משמעותה עומס כפול על האוכלוסייה הנבחנת. על כן, כדאי לבדוק את המידה בה 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 באם 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[1], וזאת מאחר ואין דרך קלה להבחין בתרומה של קיום אינפורמציה בתחום אבסטרקטי שנקרא יידע.

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 לעומת הציונים במבחן הפסיכומטרי, ואם התוצאה המתקבלת היא שהמתאם הוא מונוטוני לכל אורך התחום של הציונים, אז נסיק שאחד המבחנים מיותר. אם לעומת זאת מדרג הציונים הוא שונה בתכלית אזי המשמעות שהבחינות בודקות תחומי יידע שונים או שהתכונות הנדרשות להצלחה בבחינה אחת שונות מאשר התכונות הנדרשות להצלחה בבחינה שנייה.

בדיקת הקשר בין ציונים של שתי בחינות נתקלת בקושי שלא קיים בבדיקת המתאם בין משתנים כמותיים, כמו למשל המתאם בין המשקל של האדם לגובהו. זאת מאחר ואין אפשרות למדוד יידע בצורה כמותית, אלא רק בצורה עקיפה כגון מספר השאלות שהנבחן הצליח לענות עליהן נכון, או לחילופין בקריאה 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 בה המתאם בין הציונים הוא מונוטוני. במונח מתאם מונוטוני הכוונה היא שלאורך כל תחום הציונים של הבגרות (או הפסיכומטרי) עליה בציון בבגרות (או הפסיכומטרי) מביאה לעליה בממוצע בציון הפסיכומטרי (או בבגרות) בבחינה המקבילה. לצורך זה אנחנו משתמשים בכלי סטטיסטי חדש יחסית המיועד לבדוק קשרים מונוטוניים בתחומים שבהם אין לנו מדידה ישירה של המשתנה.[2]

מבנה העבודה הוא כדלהלן: הסעיף הראשון מכיל תיאור קצר של שאלת המחקר. הסעיף השני מתייחס למשמעויות שיש לייחס למבחן יכולות, וזאת מאחר ואין 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

[1] במחקר שנערך בשנת 2007 על ידי מרכז ארצי 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.

[2] כל השיטות המשמשות במאמר זה נדונות ב Yitzhaki and Schechtman (2013).

בבגרות לבין הצלחה בפסיכומטרי וקשר זה חזק יותר באנגלית מאשר במתמטיקה. במקביל, נמצא שקיים "רעש" אקראי בתוצאות המבחנים השונים, רעש הגורם להקטנת אמינות המבחנים כמנבאים 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 6 בא לענות על שאלה זו שהתעוררה תוך כדי המחקר והוא בודק בעזרת שימוש במדד ג'יני, את השיפור המתקבל משימוש בשני ציונים מאשר 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.

1. תיאור קצר של שאלת המחקר

בחינות הבגרות נערכות על ידי משרד החינוך במתכונת ארצית, במקצועות ליבה שונים ומיועדות לדרג תלמידים על פי רמת התמחות והיכולות בתחומים השונים. לעומת זאת, על פי הגדרת המרכז הארצי לבחינות והערכה – "הבחינה הפסיכומטרית היא כלי לחיזוי סיכויי ההצלחה בלימודים במוסדות להשכלה גבוהה."[3] היא משמשת את מוסדות הלימוד למיון מועמדים לחוגים השונים. הבחינה מאפשרת לדרג את כל המועמדים על סולם הערכה אחיד ובהשוואה לאמצעי מיון אחרים היא מושפעת פחות מהרקע השונה של כל אחד מהם או ממשתנים סובייקטיביים אחרים. 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 ל- 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 שמתכונותיהן שונות, והן על המשק הישראלי וזאת מאחר וקיימת תעשיה לא קטנה שהייתה נעלמת לו נעשתה רק בחינה אחת. מאחר ובחינת הבגרות קדמה למבחן הפסיכומטרי, השאלה הנשאלת היא באם 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, וזאת מאחר והתלמידים מתקבלים ללימודים בתחומי יידע שונים, אולם המתודולוגיה המפותחת בו תוכל לשמש בעתיד לבדיקת הטענה בדבר כושר ניבוי.

מטרת מאמר זה היא לבחון טענה צנועה יותר והיא באיזו מידה בודק המבחן הפסיכומטרי 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.

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

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

2. המשמעות שיש לייחס למבחן יכולות[4]

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 אחרות 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[5].

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[6].

בעזרת שלוש הנחות אלו אנחנו יכולים להגדיר מהו מבחן:

הגדרה: מבחן הוא אוסף של שאלות בתחום יידע 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 יידע ההבדלים בתוצאות המבחן תלויים בהתפלגות הקושי של השאלות במבחן. למעשה ניתן לטעון שמבחן לבדיקת ידע בתחום מסוים כמוהו כמדידת גובה כאשר הנבחנים עומדים מאחרי מסך והבוחן שואל שאלות כגון למי יש את הסנטימטר ה-150, למי יש את הסנטימטר ה-160 וכל אלו שגבוהים מ-150 ס"מ או 160 ס"מ עונים בהתאם לגובהם, באם יש להם או אין להם את 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 לבין מי שגבהו הוא 161. מכאן שניתן לקבוע שתוצאות מבחן הינן כמו מדידת גובה, כאשר האנשים ניצבים מאחרי 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 כמו יצירת פונקציה מתמטית כך שהציון נקבע על ידי היכולת של הנבחן. בונה המבחן, על ידי קביעת התפלגות הקושי של השאלות שולט על צורת הפונקציה המקשרת בין יכולת לציון. מכאן שניתן לראות במבחן כהפעלת טרנספורמציה מונוטונית לא יורדת על

---

[4] להרחבה בנושא זה ראה Yitzhaki and Eisenstaedt (2003). ו Schechtman. and Yitzhaki (2009).
[5] הנחה זו מקובלת במבדקים הבודקים האם השאלה שייכת לנושא המבחן - ראה Lord (1968).
[6] הנחה זו לא קיימת ב-IRT ותוסר על ידנו לאחר מכן.

המשתנה יכולת, לצורך תרגומו לציון. רמות יכולת שבהם יש הרבה שאלות מבחינות גורמים לכך שהבדל קטן ביכולת יתורגם להבדל גדול בציון ואילו רמות יכולת שאין בהם שאלות מבחינות יתורגמו להבדלים אפסיים בציון.

כדאי להעיר שעד עתה המעטנו בהבדל שבין מדידת יידע לבין מדידת גובה כאשר הנמדדים מתייצבים מאחורי פרגוד. זאת מאחר שבמדידת גובה מאחרי פרגוד ידוע לנו המרחק בסנטימטרים בין השאלות השונות ואילו במדידת יידע אין לנו מושג על המרחק ביחידות יידע בין השאלות אלא ידוע לנו רק מספר המשיבים לכל שאלה. מכאן שאין אפשרות לאמוד את פונקציית ההתפלגות של ידע (יכולת) וזאת כי חסרה לנו יחידת מדידה שהיא חיונית להגדרת התפלגות. כל מה שאנו יכולים לקבל ממבחן היא 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 שיש 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 באוכלוסיה. 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 באוכלוסיה והמתאם ביניהן[7]. מאחר ואין 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.[8] 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 יידע הרי שמבחינתנו המבחנים בודקים את אותו תחום יכולות. זאת מאחר ובמצבים מסוג זה אין לנו יכולת להבדיל ביניהם. מאחר ויכולה 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.

---

[7] הקורא המעוניין מופנה ל- Yitzhaki, Itzhaki and Pudalov (2012).

[8] 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 בהם הגדרת השונות המשותפת בין המשתנים היא סימטרית במשתנים: רק במקרה זה מתקיימת האמרה "מה שרואים מכאן הוא מה שרואים משם".

3. 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 מתאם 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 הארוך יותר שהוא: 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 היידע ועל כן, מבחינתנו המבחנים בודקים את אותו תחום יכולות.[9] 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 הוא 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 והשימוש בכלי הוא פשוט יחסית להוכחות הנדרשות כדי לבדוק את נכונות התכונות שלו, ומאחר וההוכחות התפרסמו במספר מאמרים, הרי שלא נלאה את הקורא בתרגום ההוכחות ונפנה את הקורא המתעניין למאמרים עם ההוכחות.[10]

4. שיטת ניתוח נתונים

אוכלוסיית המחקר הם אלה שנבחנו בבגרויות במתמטיקה ואנגלית בשנים 1999-2000 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.

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.

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

---

[9] 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.

[10] ראה Yitzhaki and Schechtman (2013).

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: [11]

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 B, A ציוני המינימום והמקסימום של ציוני הבגרות ואילו D , C ציוני המינימום והמקסימום של הציון הפסיכומטרי המקורי. יהי Y הציון הפסיכומטרי המנורמל.

נוסחת המעבר הינה:

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

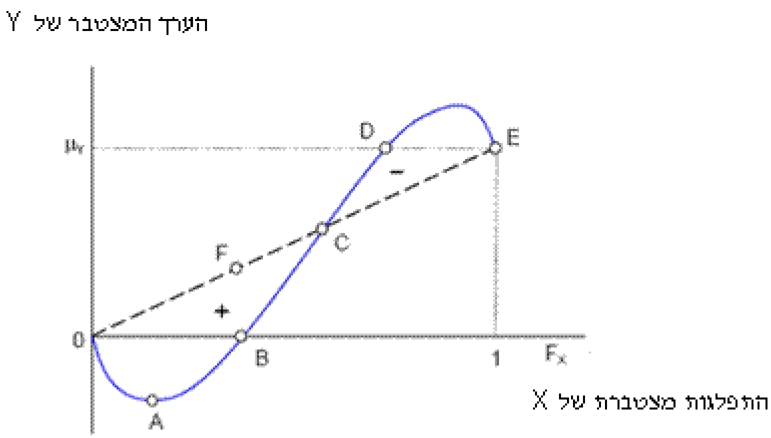
התוצאה של הנרמול הינה שתחום ציוני הבגרות ותחום ציוני הפסיכומטרי זהים.

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 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.

<u>ציור 4.1: עקומת הריכוז וקו אי תלות</u>



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).

---

[11] כדאי להעיר שאין בשיטת הנירמול בכדי להשפיע על הממצאים. השפעתה העיקרית היא על הפרשנות שיש לתת לנקודה במבחן אחד יחסית לנקודה במבחן השני.

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[12].

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:

(א) **האם העקומה חותכת את הציר האופקי**. במידה וכן אזי הקשר בין המשתנה X ל- Y אינו מונוטוני כאשר X צריך להסביר את Y. כלומר, קיים תחום בו 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 ציר האופקי יכולה לשנות את סימנו של מקדם המתאם של פירסון בין שני המשתנים. זאת מאחר וטרנספורמציה 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 סותרת את קיום ההנחה של מונוטוניות של מתאם בין המשתנים.

(ב) **האם קיים קטע שבו העקומה משנה את קעירותה מקעורה לקמורה או להיפך**. 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 דרוג 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, 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.[13]

לבסוף נעיר, שבניגוד לשונות המשותפת המבוססת על השונות שהיא סימטרית בין המשתנים (נובע מכך ש $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$ הרי שהשונות המשותפת על פי מדד ג'יני אינה סימטרית ועל כן יש לבדוק באם מה

[12] הערך המצטבר הינו סכום ערכי המשתנה עד לנקודה מסוימת. X מסודר בצורה עולה, אולם אין זה מחייב את הערך של Y בתנאי X.
[13] את התכונות של מדד ג'יני ניתן למצוא ב Yitzhaki (2003) Yitzhaki and Schechtman (2013) .

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.

5. בדיקות אמפיריות.

אוכלוסיית המחקר הם זכאי בגרות שנבחנו בבגרויות במתמטיקה ואנגלית בשנים 1999-2000 והשתתפו במבחן פסיכומטרי לאחר מכן[14]. כלומר אנו מגבילים עצמנו לאותה אוכלוסייה שהמיון לצורך קבלה לאוניברסיטה הוא רלוונטי עבורה.

סה"כ נבחנים (עוברים) בבגרות במתמטיקה ואנגלית בשנים 1999-2000 הוא 23,069, מתוכם זכאים לתעודת בגרות – 20,235 (זכאים לבגרות הם אלה שעמדו בדרישות משרד החינוך לקבלת תעודת בגרות. זה כולל בין השאר 20 יחידות לימוד, והיבחנות במקצועות שמוגדרים כמקצועות חובה). מתוכם נבחרו נבחנים שעומדים בדרישות סף של האוניברסיטאות – 17,867 (בין היתר, אלה שעברו את המבחן במתמטיקה ברמה של 3 יחידות לפחות ובנוסף לכך עבר את המבחן באנגלית ברמה של 4 יחידות לפחות). הגודל של המדגם (למעשה מפקד של הנבחנים בשנה מסויימת) הוא גדול יחסית כך שכמעט כל סטיה באומדן של פרמטר תהיה מובהקת.

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

5.1 . הנבחנים ב 5 יחידות באנגלית

לוח 5.1.1 מרכז את הפרמטרים המרכזיים המאפיינים את התפלגויות הציונים והמתאמים ביניהם. כפי שניתן לראות, הציון במבחן הפסיכומטרי (המנורמל), הממוצע והחציוני גבוהים יותר מאשר בבגרות, תוצאה שניתן להסבירה בכך שהפסיכומטרי הוא מבחן יחיד ואילו הבגרות נערך בהקבצות לפי יחידות לימוד. סטיית התקן והג'יני של הבגרות גבוהים יותר מאשר בסטיית התקן ומקדם פיזור של ג'יני במבחן הפסיכומטרי באנגלית, שוב תוצאה שניתן להסביר אותה כנגרמת מההקבצות השונות. ארבעת הטורים האחרונים מתארים את מדדי המתאם המקובלים: מדד פירסון שהוא מדד המתבסס על קשר ליניארי , מדד מתאם הדרגות (ספירמן) ושני מקדמי המתאם על פי ג'יני[15], וניתן לראות מתאם גבוהה בין תוצאות המבחנים, העומדים על 0.7.

<u>לוח 5.1.1 : פרמטרים מרכזיים של שני המבחנים באנגלית לנבחנים ב 5 יחידות*</u>

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 (0.000)	0.687 (0.000)	0.693	0.693
פסיכומטרי	87.28	89.00	55	100	8.55	0.055				

*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.

---

[14] לכל נבחן נלקח המבחן הפסיכומטרי עם הציון הכולל הגבוה ביותר.

[15] 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 וכל מקדמי המתאם דומים אחד לרעהו ומאחר והמדגם גדול, לא חישבנו טעות דגימה.

מלוח 5.1.1 ניתן ללמוד שממוצע וחציון הציונים בפסיכומטרי המנורמל, גבוהים יותר מציוני בגרות מה שמעיד שבמבחן הפסיכומטרי יש יותר ציונים גבוהים מאשר בבחינת הבגרות בחמש יחידות. גם סטיית התקן וגם מקדם הפיזור של ג'יני מעידים שבבגרות השונות בציונים גבוהה יותר מאשר בפסיכומטרי ואילו מקדמי המתאם השונים מראים את אותו סדר גודל של מתאם, שהוא יחסית נמוך, מאחר ומדובר בציונים שמקבל תלמיד בשני מבחנים שמיועדים לבדוק יכולת באותו תחום.

בטרם נעמיק בניתוח הממצאים כדאי לחזור ולהסביר את ההבדלים בין ציור התפלגות של משתנה כמותי מדיד, כגון גובה של בני אדם לבין משתנה שהוא יכולת של אדם:

כאשר מודדים גובה של אדם יש לנו משתנה (גובה) וכל עוד לא נשנה את יחידת המדידה, סנטימטרים, גובהו הנמדד של אדם לא ישתנה, מלבד אולי טעויות רישום ומדידה שהן קטנות ביחס. כאשר אנחנו בודקים יכולות הרי שיש כמה שאלות שיש צורך להתייחס אליהן:

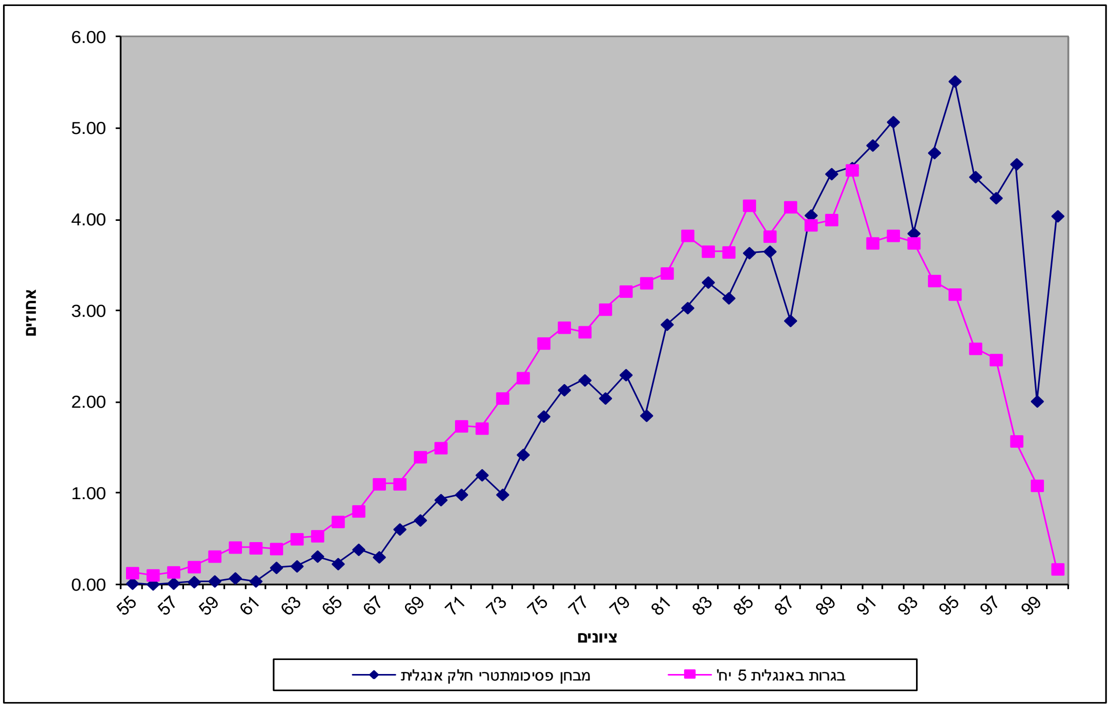
(א). האם יכולת ניתנת למדידה כמו גובה? כלומר האם יכולת היא משתנה עצמאי כמו גובה או שהיא תלויה במספר גורמים כגון יכולת זיכרון, יכולת חשיבה מתמטית, וכדומה. לשאלה זו נתייחס בהמשך. בשלב זה נניח שקיים משתנה יכולת שאנו יכולים למדוד אותו כמו שאנו מודדים גובה.

(ב). מהו המאפיין את מדידת היכולת? יכולת באנגלית, בניגוד לגובה, איננה ברת מדידה בעזרת סרגל. כל שאנו יכולים לעשות, הוא לשאול שאלה וכל אלו שהם בעלי יכולת לענות נכון על השאלה, יענו נכון, להוציא טעויות אקראיות. הדבר דומה למדידת גובה של אנשים כאשר האנשים נמצאים אחרי מסך. הבוחן שואל שאלה מהסוג למי יש את הסנטימטר ה 160 וכל אלו שגובהם מעל 160 יענו שיש להם את הסנטימטר ה 160. לאחר מכן ישאל הבוחן למי יש את הסנטימטר ה 180 וכל אלו שגובהם מעל 180 יענו בחיוב. אולם, כפי שהוזכר, בעוד שבסנטימטרים אנחנו יודעים למדוד מרחק באופן אובייקטיבי, הרי שבמדידת יכולת אין לנו ידיעה באם המרחק בין השאלות הסמוכות בדרגת קושי הוא 20 סנטימטר או שווה ערך ל 40 סנטימטר. כל אשר בידינו הוא השכיחות היחסית של אלו שענו נכון על השאלה.

המשמעות המעשית של הטענה היא, שכאשר אנחנו משווים בין התפלגויות של ציונים במבחנים שונים, הרי שככל שהשכיחות היחסית בתחום מסוים גבוהה יותר הרי שה"מרחק" בין השאלות "גדול" יותר. זאת כיוון שבפער שבין השאלות "נכנסו" נבחנים במספר רב יותר. מובן שגם מספר השאלות קובע את יכולת ההבחנה של המבחן. ככל שמספר השאלות גדול יותר, כן יכולת ההבחנה של המבחן גבוהה יותר.

ציור מס. 5.1.1 מציג את ההתפלגויות של הציונים בבגרות מול פסיכומטרי מנורמל. מהציור בולט שהמבחן הפסיכומטרי מבחין טוב יותר בתחום היכולות הנמוך ואילו הבגרות מתרכז בתחום היכולות הגבוהות, תוצאה שניתן להסביר אותה מחד בהקבצות שקיימות בבגרות ובכך שהפסיכומטרי הוא מבחן יחיד המיועד לכל. בנוסף, קל לראות שהציונים אינם מתפלגים נורמלית וזאת מאחר וההתפלגויות הן א-סימטריות. .

ציור 5.1.1: התפלגות נבחנים באנגלית 5 יחידות*



* התצפיות חוברו בעזרת קטעים ישרים

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.[16] אם לעומת זאת אנחנו רואים שהקשר משנה את סימנו לאורך ההתפלגות הרי שההיסק שלנו יהיה שהמבחנים אינם מודדים את אותה יכולת. זאת מאחר וקשר מונוטוני משמעותו שאיננו יכולים לזהות בין שני התחומים של יכולת.

לצורך בחינת המונוטוניות אנחנו נעזרים בשני תרשימים של עקומת NLMA [17] 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 הציונים בין המבחנים היו בלתי תלויים סטטיסטית, לבין הערך המצטבר בפועל לאותם נבחנים.

ציור 5.1.2 בוחן את המידה בה ציוני המבחן הפסיכומטרי יוצרים קשר מונוטוני עם ציוני 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[18] שיכול להפוך את סימנו של המתאם בין ציוני שני המבחנים.

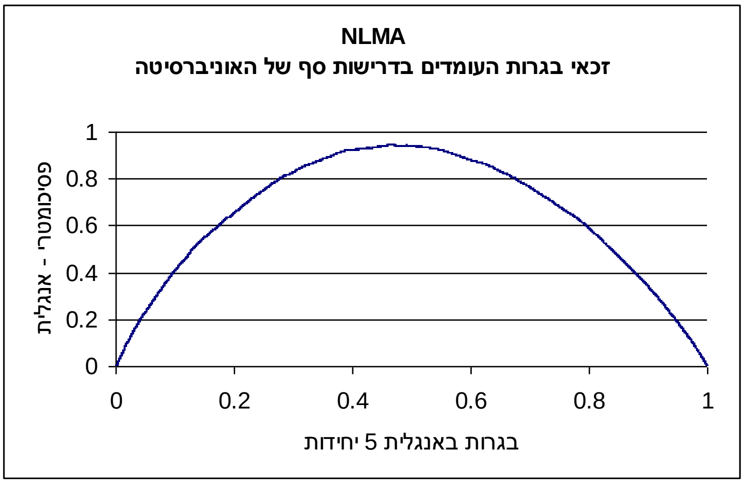
מציור 5.1.2 אנחנו למדים שהעקומה עולה בתחילה ולאחר מכן יורדת – כלומר שבתחום שבו ציוני הבגרות נמוכים גם ציוני הפסיכומטרי נמוכים ואילו בתחום שבו ציוני הבגרות גבוהים גם ציוני הפסיכומטרי הם גבוהים.

---

[16] ייתכן כמובן שההצלחה במבחנים השונים תלויה בגורמים שונים אולם אין המבחנים מסוגלים להבחין ביניהם.
[17] על תכונות העקומה והשימוש בתחום החינוך – ראה Yitzhaki, Itzhaki and Pudalov (2012) .
[18] במונח מבחן לגיטימי אחר הכוונה היא למבחן אחר עם התפלגות שונה של קושי השאלות בו.

אשר לקעירות העקומה אנחנו רואים שהעקומה היא קעורה, כך שאנחנו יכולים להסיק שציוני הפסיכומטרי יוצרים קשר מונוטוני לציוני הבגרות.

ציור 5.1.2 : ציון פסיכומטרי באנגלית כפונקציה של בגרות באנגלית 5 יח'



לבדיקה נוספת של הממצא שלנו חזרנו על בדיקת הקשר בעזרת רגרסיית הג'יני המורחב[19]. רגרסיית הג'יני המורחב דומה לרגרסיית ריבועים פחותים מלבד בדבר אחד. מריץ הרגרסיה מציין פרמטר המאפשר להדגיש תחום מסוים לאורך המשתנה המסביר. ככל שהפרמטר גבוה יותר כך מודגש יותר תחום הציונים התחתון. לוח 5.1.2 מציג את הרגרסיות של ציון פסיכומטרי (מנורמל) על ציון הבגרות. העמוד השמאלי בלוח מדגיש את החלק התחתון בציוני הבגרות ואילו הטור הימני מדגיש את החלק העליון. תחום השיפועים המתקבלים הם בין 0.615 לבין 0.647 כלומר הקשר הוא כמעט לינארי, אולם עם שיפוע הולך וגדל, כלומר קשר קמור במקצת. המשמעות היא שעליה של נקודה בבגרות מביאה לעליה בממוצע של בין 0.615 לבין 0.647 בציון הפסיכומטרי המנורמל ושככל שמדגישים ציונים גבוהים בבגרות כך ההשפעה על ציון בפסיכומטרי גבוהה יותר. תוצאה זו מוסברת בכך שככל שציון הבגרות גבוה יותר, בגלל שהפסיכומטרי מבחין יותר בתחום העליון, כן ההשלכה של נקודה בבגרות על הציון הפסיכומטרי גבוהה יותר.

לוח 5.1.2: רגרסית ג'יני מורחב – ציון פסיכומטרי באנגלית כפונקציה של ציון בגרות באנגלית 5 יח'.

v=7	v=5	v=3	v=2	v=1.5	v =1	v=0.5		
35.90	35.40	34.68	34.21	33.94	33.63	33.25	α (ממוצע)	1
36.58	36.05	35.31	34.84	34.56	34.24	33.86	α (חציון)	2
0.615 (0.008)	0.621 (0.008)	0.630 (0.007)	0.636 (0.007)	0.638 (0.006)	0.642 (0.006)	0.647 (0.006)	β (אנגלית 5 יח')	3
0.451	0.536	0.652	0.709	0.721	0.693	0.548	מתאם ג'יני Γyx	4
0.996	0.865	0.722	0.669	0.662	0.693	0.881	מתאם ג'יני Γxy	5
0.275	0.275	0.275	0.276	0.276	0.276	0.276	GR	6
87.26	87.26	87.32	87.22	87.22	87.24	87.28	α (ממוצע)+ $\beta(v)*\bar{X}$	7
-17.53	-17.70	-17.96	-18.13	-18.19	-18.30	-18.45	$\beta(v)*(X_{min}-\bar{X})$	8
10.14	10.24	10.39	10.49	10.52	10.59	10.67	$\beta(v)*(X_{max}-\bar{X})$	9

[19] לתיאור תכונות רגרסיה זו ראה Schechtman et al. (2008).

בלוח 5.1.2 אנחנו בודקים את טעות התחזית של שימוש ברגסיות ג'יני מורחב כאשר ציון הבגרות משמש כבסיס לחיזוי הציון הפסיכומטרי. הרעיון מאחורי הרגרסיות הוא הרעיון הבא: Yitzhaki (1996) הראה שאומדן של מקדם רגרסיה של ריבועים פחותים או של רגרסיית ג'יני מהווה ממוצע משוקלל של שיפועי עקום הרגרסיה בין נקודות סמוכות של המשתנה הבלתי תלוי. המשקלות נקבעים על ידי שיטת הרגרסיה והתפלגות המשתנה הבלתי תלוי. רגרסיית הג'יני המורחב מאפשרת לנו לשלוט במערכת המשקלות על ידי בחירת פרמטר v שככל שהוא גבוה יותר כן מודגש החלק התחתון של המשתנה הבלתי תלוי. מאחר וכל רגרסיה משמעותה קו ישר, ומאחר וכל הרגרסיות עוברות דרך הממוצע (או החציון בהתאם לבחירתנו) אזי ניתן להעריך את הסטייה המקסימלית של קו רגרסיה אחד מרעהו. כך למשל הנוסחה של כל קו רגרסיה היא:

$$Y(x,v) = \beta(v)(x-\bar{x})$$

כך שהסטיה של קו רגרסיה אחד מרעהו היא:

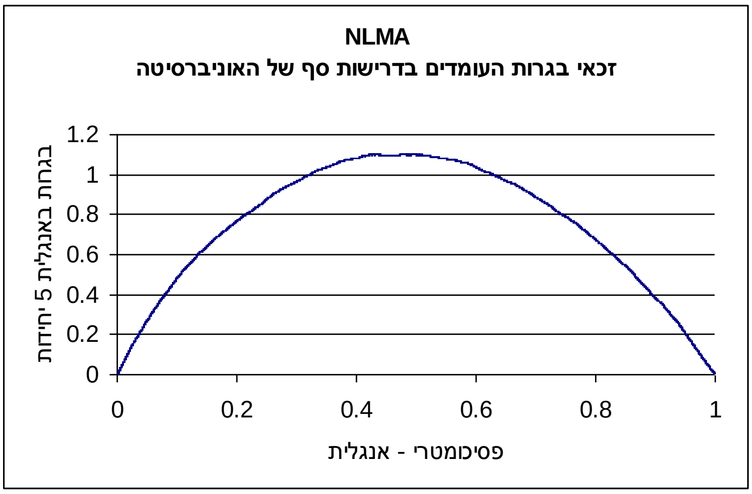
$$Y(x,v_1) - Y(x,v_2) = [\beta(v_1) - \beta(v_2)](x-\bar{x})$$

המשמעות היא שבנקודת הממוצע של המשתנה הבלתי תלוי כל קוי הרגרסיה מתלכדים וככל שמתרחקים מממוצע המשתנה הבלתי תלוי וככל שהקשר הוא קעור או קמור יותר נצפה להגדלה של סטיית התחזית בין קו אחד לרעהו.

מהשורות התחתונות של לוח 5.1.2 ניתן לראות את סטיית התחזית שמתקבלת תוך שימוש בשיטות שקלול שונות לחישוב קו הרגרסיה. בשורה מספר 7 מודגמת הנקודה בה עובר קו הרגרסיה בנקודת הממוצע של ציון הבגרות, בשורה 8 - התוספת למשתנה התלוי בנקודת הקיצון התחתונה של הציון הנחזה ואילו בשורה 9 - הערך בנקודת הקיצון העליונה. ניתן לראות שאין ההבדלים עולים על נקודה אחת הציון המשוקלל. לדוגמה כאשר משתמשים ב V=7 או ב V=0.5 החיוץ של קו הרגרסיה הלינארי לציון הגבוה ביותר הוא 10.14 ו 10.67 בהתאמה כלומר בערך חצי נקודה. הבדלים בתחזיות קוי הרגרסיה הם קטנים מועל כן ניתן להעריך שהקו הלינארי המתאר את עקום הרגרסיה מהווה קירוב טוב לקשר בין המשתנים.

כדי לבדוק באם הציון של הבגרות יוצר קשר מונוטוני סימטרי עם הציון הפסיכומטרי, אנחנו חוזרים על ציור 5.1.2 ולוח 5.1.2 בהיפוך הצירים. על הציר האופקי הציון הפסיכומטרי ואילו הציר האנכי מציג את ציון הבגרות. ציור 5.1.3 מציג את עקומת NLMA של ציוני הבגרות כפונקציה של ההישגים במבחן הפסיכומטרי. 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 5.1.3 : ציון בגרות באנגלית 5 יחידות כפונקציה של ציון פסיכומטרי באנגלית



לוח 5.1.3 : מקדמי רגרסיה של ציוני אנגלית 5 יחידות כפונקציה של ציוני פסיכומטרי באנגלית

	v=0.5	v=1	v=1.5	v=2	v=3	v=5	v=7
α (ממוצע)	17.46	18.29	19.02	19.64	20.65	22.01	22.88
α (חציון)	18.27	19.09	19.81	20.41	21.43	22.81	23.68
β (פסיכומטרי אנגלית)	0.757 (0.007)	0.747 (0.007)	0.739 (0.007)	0.731 (0.007)	0.720 (0.008)	0.704 (0.008)	0.694 (0.009)
מתאם ג'יני Γyx	0.545	0.693	0.722	0.711	0.653	0.536	0.450
מתאם ג'יני Γxy	0.892	0.693	0.657	0.662	0.708	0.843	0.991
GR	0.277	0.277	0.277	0.277	0.277	0.276	0.275
α (ממוצע)+ β(v)* $\overline{X}$	83.53	83.49	83.52	83.44	83.49	83.45	83.45
β(v)*(X_{min}- $\overline{X}$)	-24.44	-24.11	-23.85	-23.60	-23.24	-22.73	-22.40
β(v)*(X_{max}- $\overline{X}$)	9.63	9.50	9.40	9.30	9.16	8.95	8.83

לוח 5.1.3 מציג את הקשר ההופכי לקשר המוצג בלוח 5.1.2 . במקום לנסות לנבא את תוצאות הפסיכומטרי מציוני הבגרות בעזרת קו רגרסיה לינארי, אנו מנסים לנבא את ציוני הבגרות על סמך תוצאות הפסיכומטרי.[20] כפי שניתן לראות טיב האומדנים דומה כך שאנו מקבלים סימטריה בין שתי מערכות הציונים. כאשר מסתכלים על נקודת הממוצעים אנו רואים שכל קוי הרגרסיה עוברים דרכה בהתאם למגבלה שהטלנו אולם המעניין הוא שההבדלים בסטיית התחזית בתחום העליון של הציונים קטנים יותר מאשר ההבדלים בסטיות התחזית בתחום התחתון של הציונים. ההסבר האינטואיטיבי לכך הוא שאנו עוסקים בתלמידים שנבחנו ב 5 יחידות באנגלית ועל כן הם קרובים יותר לתחום הציונים הגבוה בפסיכומטרי, מה שבולט מכך הוא שהציון הממוצע בפסיכומטרי המנורמל הוא גבוה.

כדי לבדוק בצורה אי-פרמטריות את הסימטריה בקשר בין ציוני המבחנים, ליווחנו את הציונים בעזרת מטריצת מעבר. לוח 5.1.4 מציג את מטריצת מעבר בין עשירוני הציונים. 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

[20] כדאי להעיר שברגרסיית ריבועים פחותים הסימנים של מקדם רגרסיה ומקדם הרגרסיה ההופכית חייבים להיות זהים וזאת מאחר ובשני האומדים המונה מבוסס על קווארינס שמוגדר בצורה סימטרית. ברגרסיית הג'יני אין הדבר מחוייב המציאות והדבר תלוי בסימטריה ביחס לקו הרגרסיה בין ההתפלגות המשתנה התלוי והבלתי תלוי.

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 הוא שהמבחנים בודקים את אותו נושא אולם יש הרבה רעש בתוצאות.

<u>לוח 5.1.4 : מטריצת מעבר: עשירוני ציונים – בגרות באנגלית 5 יח' ופסיכומטרי באנגלית</u>

		בגרות באנגלית 5 יח'										
פסיכומטרי באנגלית		1	2	3	4	5	6	7	8	9	10	סה"כ
	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

קיומו של "רעש" בתוצאות המבחנים צריך להטריד מאחר והמשמעות 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.

5.2 . 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 7,059. כדאי לשים לב שמספר הנבחנים במתמטיקה 5 יחידות נמוך בכ 40 אחוז מאשר מספר הנבחנים המקביל באנגלית 5 יחידות (כ 7000 לעומת כ 12000) . המשך הניתוח דומה במבנהו לניתוח שנעשה עבור המבחנים באנגלית, ועל כן אנו מתרכזים בממצאים בלבד.

<u>לוח 5.2.1 : פרמטרים מרכזיים של שני המבחנים במתמטיקה לנבחנים ב 5 יחידות</u>

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				

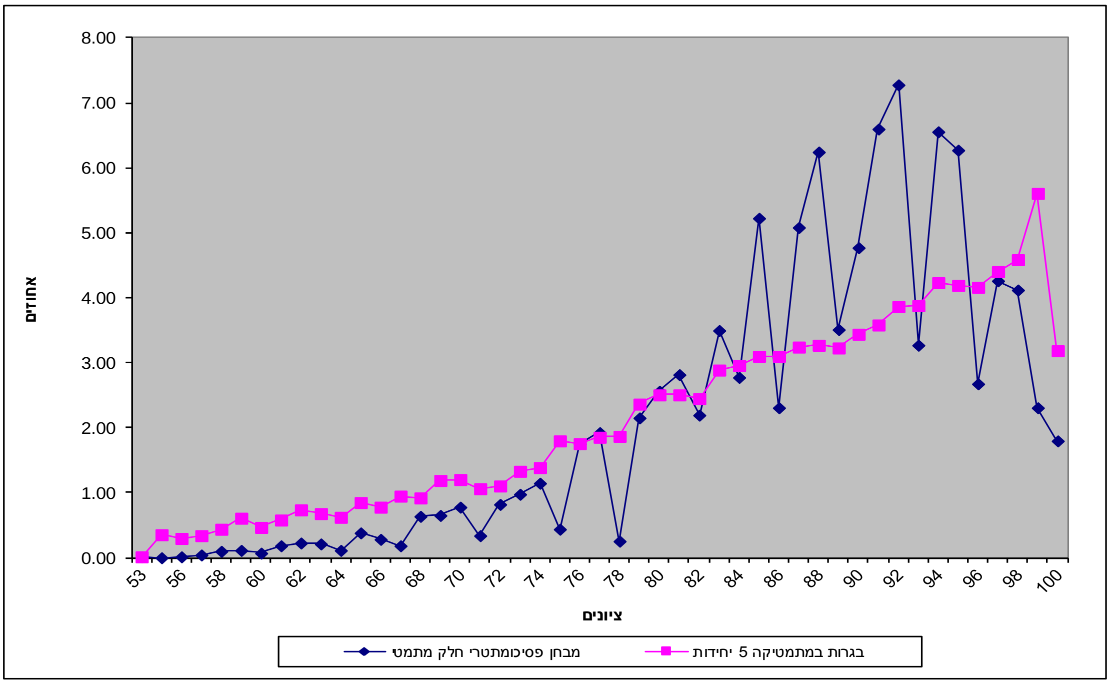
כפי שניתן לראות מלוח 5.2.1**,** ממוצע הציון והחציון (המנורמלים) בפסיכומטרי גבוהים יותר מממוצע והחציון של הציונים בבגרות (5 יחידות). לעומת זאת, שני מדדי הפיזור שהשתמשנו בהם, סטיית התקן ומדד ג'יני גבוהים יותר במבחן הבגרות מאשר במבחן הפסיכומטרי. המשמעות היא שציוני הבגרות הם עם פיזור גדול יותר מאשר ציוני 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 ובפסיכומטרי יש מבחן יחיד הרי שטבעי הדבר שמי שניגש ל 5 יחידות במתמטיקה יהיה במקבץ הגבוה של הציונים בפסיכומטרי.[21] הממצא המפתיע כאן הוא המתאם הנמוך יחסית בין שני מבחנים שהועברו לאותם אנשים, שהינם בעלי ההשכלה הרחבה ביותר במתמטיקה. כל מקדמי המתאם בין תוצאות שני המבחנים לא עולים על 0.4 וזאת כאשר באנגלית הם עומדים על 0.7. ממצא מפתיע נוסף, הוא שבדומה לתוצאות המבחן באנגלית, גם כאן מתקבלת סימטריה בהתפלגות המשותפת של הציונים בבגרות ובפסיכומטרי, ממצא מפתיע שני הוא הדמיון בין שני מקדמי המתאם של הג'יני, 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.

ציור 5.2.1 מציג 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[22]. מאחר וקיים 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 הנירמול שנערך בבגרות עבור הציונים המתקבלים בין נבחני 5 יחידות בנפרד.

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

[22] כדאי לשים לב שמאחר שההתפלגות של ציונים תלויה בהתפלגות הקושי של השאלות, אין יכולת לאמוד את התפלגות האמיתית של האוכלוסיה.

ציור 5.2.1 :   התפלגות ציוני פסיכומטרי ובגרות במתמטיקה של נבחני מתמטיקה 5 יחידות*

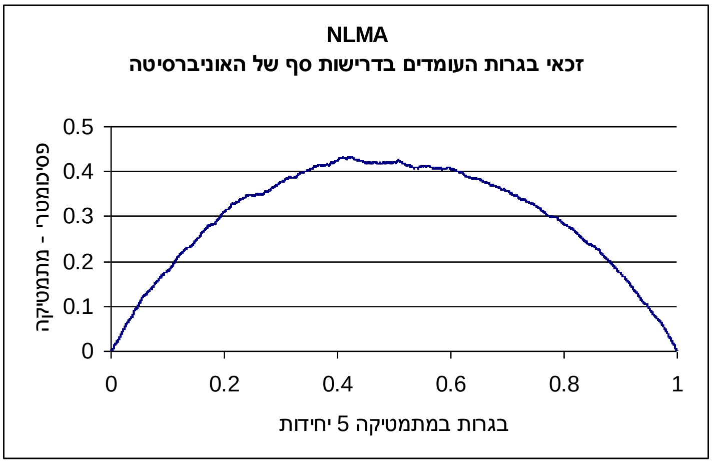


* בבגרות במתמטיקה 5 יחידות יש סה"כ 10 שאלות: 6 שאלות של $16\frac{2}{3}$ נקודות (3 יחידות), ו-4 שאלות של 25 נקודות (השלמה ל-5 יחידות). לאחר מכן משקללים את התוצאה. בפסיכומטרי במתמטיקה יש 50 שאלות (שני חלקים של 25 כל אחד). חוסר השיאים הרבים בתוצאות הבחינה הפסיכומטרית נראים לנו מוזרים אבל אלו הן התוצאות שקיבלנו.

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 המונוטוני אינו משנה את סימנו לאורך ההתפלגות של 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.

ציור עקומת 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 ולהעלמות של תנודות אקראיות.

ציור 5.2.2: ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה 5 יח'



לוח 5.2.2 מציג את הרגרסיות של ציון פסיכומטרי (מנורמל) על ציון הבגרות. תחום השיפועים המתקבלים הם בין 0.26 לבין 0.308 כלומר שהקשר, בדומה לתחום האנגלית, הוא כמעט לינארי אולם עם שיפוע הולך וגדל כלומר קשר קמור במקצת.

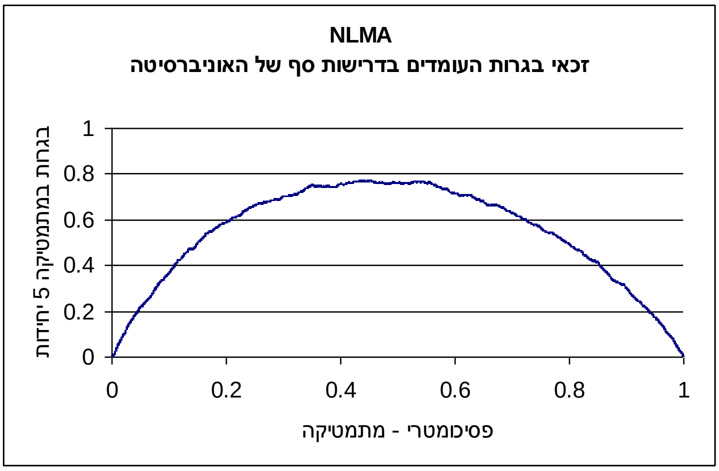
לוח 5.2.2: רגרסית ג'יני מורחב – ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה 5 יח'.[23]

	v=0.5*	N=1	v=1.5*	v=2	v=3	v=5	v=7
α (ממוצע)	61.46	62.51	63.20	63.67	64.31	65.06	65.52
α (חציון)	62.62	63.76	64.48	64.96	65.59	66.36	66.82
β (מתמטיקה 5 יח')	0.308 (0.008)	0.296 (0.008)	0.288 (0.008)	0.282 (0.009)	0.275 (0.009)	0.266 (0.010)	0.261 (0.011)
מתאם ג'יני Γyx		0.409		0.414	0.380	0.312	0.261
מתאם ג'יני Γxy		0.392		0.369	0.392	0.463	0.541
GR		0.074		0.075	0.075	0.075	0.075
α (ממוצע)+ β(v)* $\overline{X}$	87.92	87.93	87.94	87.89	87.93	87.91	87.94
B(v)*(X_{min}- $\overline{X}$)	-10.13	-9.74	-9.47	-9.28	-9.05	-8.75	-8.59
B(v)*(X_{max}- $\overline{X}$)	4.34	4.17	4.06	3.98	3.88	3.75	3.68

*חושב בעזרת Gini– SAS, ואין אפשרות לחשב **Γyx** ו-**Γxy** .

ציור 5.2.3 מציג את עקומת 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 בגרות במתמטיקה 5 יח' כפונקציה של ציון פסיכומטרי במתמטיקה



23 סטיות תקן חושבו בעזרת Fast Jackknife. בשיטה זו מחושבת השונות של האומד תוך הורדת תצפית, אולם אין מחשבים את כל החישוב מחדש. בחלק מהלוח חסרים סטיות תקן בגלל אי קיום תוכנה מתאימה.

לוח 5.2.3:   מקדמי רגרסיה על פי ג'יני מורחב - ציון בגרות במתמטיקה 5 יח' כפונקציה של פסיכומטרי במתמטיקה.[24]

	v=0.5*	v=1	v =1.5*	v=2	v=3	v=5	v=7
α (ממוצע)	37.13	38.26	39.02	39.57	40.37	41.41	42.13
α (חציון)	39.12	40.27	41.90	41.55	42.35	43.41	44.15
β (פסיכומטרי מתמטיקה)	0.555 (0.015)	0.541 (0.016)	0.533 (0.015)	0.526 (0.016)	0.518 (0.016)	0.506 (0.017)	0.498 (0.018)
מתאם ג'יני Γyx		0.392		0.405	0.376	0.316	0.269
מתאם ג'יני Γxy		0.409		0.386	0.407	0.474	0.548
GR		0.079		0.079	0.079	0.079	0.079
α (ממוצע)+ β(v)*$\overline{X}$	85.94	85.84	85.89	85.83	85.92	85.91	85.92
β (v)*(X_{min}- $\overline{X}$)	-19.39	-18.90	-18.62	-18.34	-18.10	-17.68	-17.40
β (v)*(X_{max}- $\overline{X}$)	6.69	6.52	6.43	6.34	6.25	6.10	6.01

*חושב בעזרת SAS –Gini, ואין אפשרות לחשב **Γyx** ו-**Γxy** .

לוח 5.2.3 מציג את הבגרות כפונקציה של הפסיכומטרי. גם כאן ככל שמדגישים את התחום העליון, מקדם הרגרסיה גבוה יותר וההבדל בין מקדם הרגרסיה הנמוך ביותר לגבוה ביותר הוא 0.056, כלומר סדר גודל של שינוי של 10 אחוזים בשיפוע.

מבחינה זו הממצא שלמרות שהקשרים הם מונוטוניים מקדמי המתאם נמוכים מטריד במקצת. כדי להבין את משמעותו מציג לוח 5.2.4 מטריצת 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 2.7 אחוז מהתלמידים הם גם בעשירון התחתון של בגרות וגם בעשירון התחתון של הפסיכומטרי. כלומר, לגבי העשירון התחתון יש הסכמה בין המבחנים רק לגבי 2.7 אחוזים מהתלמידים (שהם 27 אחוזים מהקבוצה). המפתיע הוא שכ-18 אחוזים מהקבוצה של העשירון התחתון של הפסיכומטרי נמצאים מעל החציון בבגרות (0.1 + 0.3 +0.3 +0.4+0.7) וכ-25 אחוז מהעשירון התחתון של הבגרות נמצאים מעל החציון בפסיכומטרי. ממצא דומה ניתן לקבל אם מתרכזים בעשירון העליון שרק 29 אחוז 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. מטריצת המעבר נראית כמעט סימטרית 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 לעוצמת ההצלחה במבחן השני. לא בדקנו את כושר הניבוי להצלחה בלימודים אקדמאיים, אולם המתאם הנמוך בין הבגרות והפסיכומטרי אינו מנבא טובות לכושר הניבוי לגבי מבחנים מסוג נוסף.

[24] ראה הערה 19.

<u>לוח 5.2.4: מטריצת מעבר: עשירוני ציונים – בגרות במתמטיקה 5 יח' ופסיכומטרי במתמטיקה</u>

פסיכומטרי במתמטיקה	בגרות במתמטיקה 5 יח'										
	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	1.0	1.2	1.1	0.7	0.4	10
5	0.8	1.1	1.0	1.3	1.1	1.0	0.9	0.9	1.0	0.9	10
6	0.7	1.0	1.4	1.1	1.2	1.1	1.0	1.0	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	1.0	1.4	1.3	1.5	2.0	10
10	0.2	0.3	0.3	0.4	0.7	0.9	0.9	1.4	2.0	2.9	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 שיגרם למשק מביטול ההכבדה שקיימים שני מבחנים באנגלית הוא קטן יותר מאשר הנזק שיגרם במיון מועמדים במתמטיקה.

5.3 . נבחנים באנגלית – 4 ו-5 יחידות

בסעיף זה אנו חוזרים על הניתוח שנעשה בסעיף הקודם כאשר הפעם אנחנו כוללים את התלמידים שנבחנו ב 4 ו ב 5 יחידות, כלומר כל התלמידים שזכאים להתקבל לאוניברסיטה. השקלול של ציוני הבגרות נעשה על פי הכללים המקובלים באוניברסיטאות בעת קבלת סטודנטים.[25] מספר התצפיות המשמשות אותנו הוא 17,867 נבחנים שנבחנו בשנים 1999-2000 ושעומדים בדרישות הסף של אוניברסיטה, מתוכם 5,767 – 4 יחידות, 12,100 – 5 יחידות. כמו בסעיף הקודם אנו מנרמלים את הציונים כך שתחום הציונים בשני המבחנים יהיה זהה.

<u>לוח 5.3.1 : פרמטרים מרכזיים של שני המבחנים באנגלית</u>

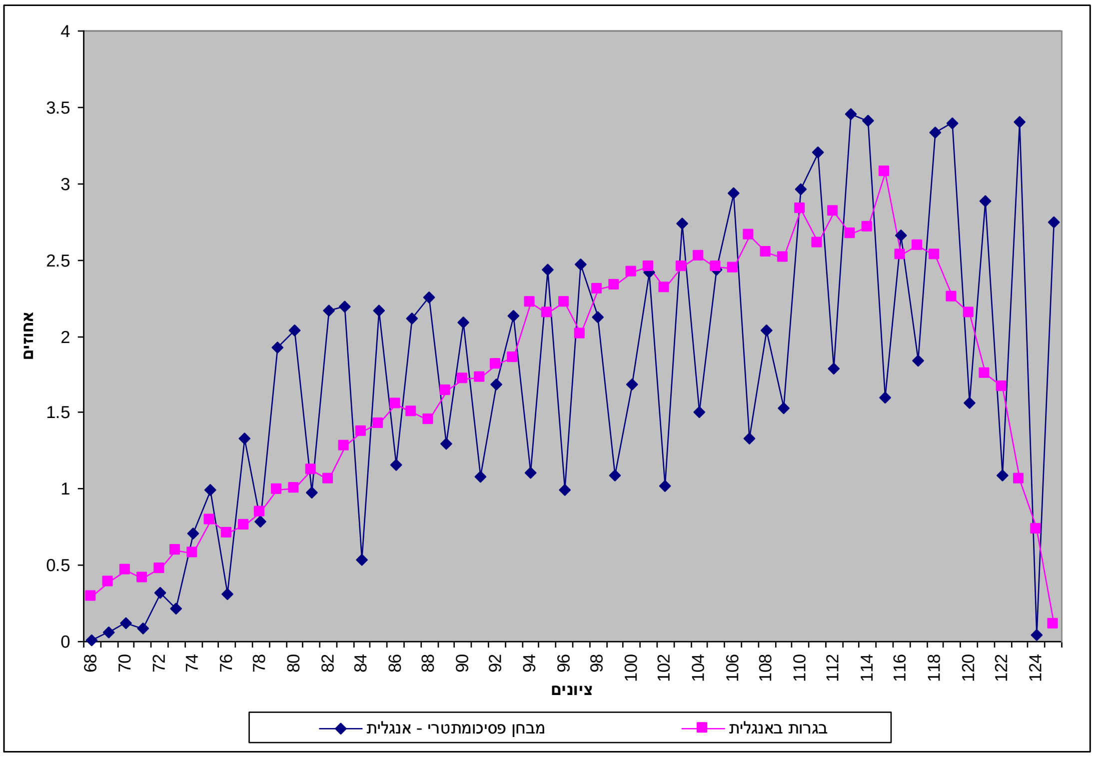
N=17,867	ממוצע	חציון	מינימום	מקסימום	סטיית תקן	ג'יני	פירסון	ספירמן	מתאם ג'יני Γyx	מתאם ג'יני Γxy
בגרות באנגלית	101.5	103	68	125	13.72	0.077	0.82 (0.000)	0.82 (0.000)	0.83	0.83
פסיכומטרי באנגלית	102.3	104	68	125	14.41	0.081				

השוואה 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. ציור 5.3.1 מציג את התפלגות השכיחויות. ניתן לראות שהתפלגות ציוני הבגרות רציפה יותר מאשר התפלגות ציוני הפסיכומטרי. אם "מחליקים" את ציוני הפסיכומטרי על ידי מיצוע

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

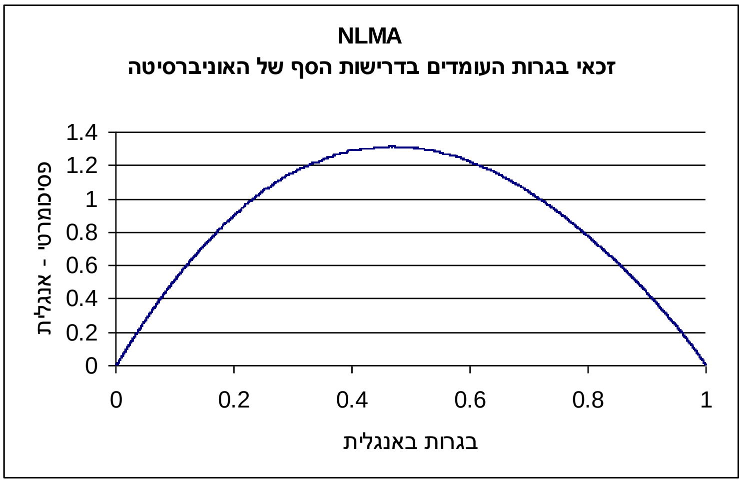
של תצפיות סמוכות היינו מקבלים התפלגויות כמעט זהות. ייתר על כן, השוואת מקדמי המתאם, מראה גם שכל מקדמי המתאם גדלו בצורה משמעותית, כך שטיב ההתאמה בין שתי מערכות הציונים, נעשה גבוה הרבה יותר מאשר בהשוואה של פסיכומטרי עם נבחנים בחמש יחידות בלבד.

<u>ציור 5.3.1 : התפלגות ציוני פסיכומטרי ובגרות באנגלית</u>



ציור מס. 5.3.2 מציג עקומת NLMA של ציון פסיכומטרי כפונקציה של ציוני הבגרות. כפי שנראה מהציור 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.

<u>ציור 5.3.2 : ציון פסיכומטרי באנגלית כפונקציה של ציון בגרות באנגלית</u>



לוח 5.3.2 מציג את תוצאות הפסיכומטרי באנגלית כפונקציה של ציון הבגרות באנגלית בעזרת רגרסיות הג'יני המורחב. השורה השלישית מציגה את מקדם הרגרסיה המתקבל בהרצות השונות. ככל שמדגישים יותר את

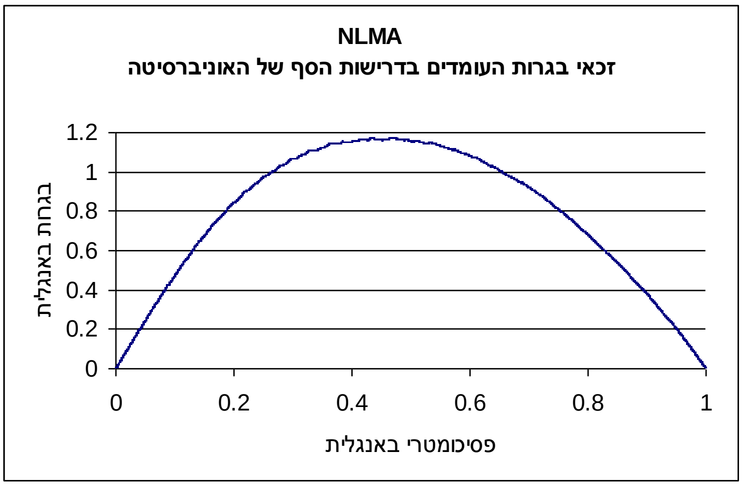
הציונים הגבוהים בבגרות, כן מתקרב מקדם הרגרסיה לאחד וכן גדל מקדם המתאם. המשמעות היא שככל שאנו מתרכזים בציונים הגבוהים, כן הדמיון בתוצאות שני המבחנים גבוה יותר. תוצאה זו גם מתקבלת מהשוואת סטיות קווי הרגרסיה, שמראים שהסטיה המתקבלת עבור הקצה התחתון של ההתפלגות גדולה יותר מהסטיה המתקבלת בקצה העליון של ההתפלגות. אולם כדאי להדגיש שבכל הרגרסיות אנו מקבלים שעליה של נקודה בציון בבגרות מביאה לעליה ממוצעת של בין 0.83 לבין 0.89 בפסיכומטרי.

לוח 5.3.2 : רגרסית ג'יני מורחב– ציון פסיכומטרי באנגלית כפונקציה של ציון בגרות באנגלית.

	v=0.5	N=1	v =1.5	v=2	v=3	v=5	v=7
α (ממוצע)	12.32	12.70	13.15	13.64	14.65	16.56	18.17
α (חציון)	12.66	13.04	13.52	14.01	15.04	16.96	18.62
β (בגרות באנגלית)	0.885 (0.004)	0.881 (0.004)	0.877 (0.004)	0.872 (0.004)	0.862 (0.004)	0.843 (0.005)	0.827 (0.005)
מתאם ג'יני Γyx	0.828	0.826	0.825	0.825	0.824	0.821	0.819
מתאם ג'יני Γxy	0.824	0.827	0.827	0.825	0.821	0.809	0.796
GR	0.428	0.445	0.452	0.454	0.451	0.436	0.420
α (ממוצע)+ β(v)* $\overline{X}$	102.15	102.12	102.17	102.15	102.14	102.12	102.11
β (v)*(X_{min}- $\overline{X}$)	-29.65	-29.51	-29.38	-29.21	-28.88	-28.24	-27.70
β (v)*(X_{max}- $\overline{X}$)	20.80	20.70	20.61	20.49	20.26	19.81	19.43

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 5.3.3: ציון בגרות באנגלית כפונקציה של ציון פסיכומטרי באנגלית



לוח 5.3.3 מציג את רגרסיות הג'יני המורחב של ציון הבגרות כפונקציה של הציון הפסיכומטרי. הממצא הבולט העיקרי הוא שהדגשת תחומים שונים בציון הפסיכומטרי כמעט ואינה משנה את השיפוע של קו הרגרסיה, כך שאנו יכולים להתייחס אל הקשר בין הציונים כמעט כלינארי. ההבדל בין השיפוע כאשר מדגישים את הקצה התחתון של הפסיכומטרי יחסית להדגשת הקצה העליון של ההתפלגות מסתכם בהבדל שבין 0.788 לבין 0.771 שמשמעותו היא שעליה של נקודה אחת בציון הפסיכומטרי תביא בין 0.77 ל 0.788 עליה

בציון הבגרות. גם שאר חלקי הלוח מעידים על הבדלים קטנים כאשר מדגישים חלקים שונים לאורך התפלגות הציון הפסיכומטרי.

<u>לוח 5.3.3 : רגרסית ג'יני מורחב– ציון בגרות באנגלית כפונקציה של ציון פסיכומטרי באנגלית.</u>

	v=0.5	v=1	v =1.5	v=2	v=3	v=5	v=7
α (ממוצע)	22.72	22.33	22.01	21.75	21.37	21.00	20.98
α (חציון)	23.39	22.99	22.64	22.36	21.95	21.55	21.53
β (פסיכומטרי באנגלית)	0.771 (0.004)	0.775 (0.004)	0.778 (0.004)	0.781 (0.004)	0.784 (0.004)	0.788 (0.005)	0.788 (0.005)
מתאם ג'יני Γyx	0.824	0.827	0.827	0.825	0.821	0.809	0.796
מתאם ג'יני Γxy	0.828	0.826	0.825	0.825	0.824	0.821	0.819
GR	0.427	0.440	0.446	0.449	0.450	0.446	0.441
α (ממוצע)+ β(v)* $\overline{X}$	101.59	101.61	101.60	101.65	101.57	101.61	101.59
β (v)*($X_{min}-\overline{X}$ **)**	-26.45	-26.58	-26.69	-26.79	-26.89	-27.03	-27.03
β (v)*($X_{max}-\overline{X}$ **)**	17.50	17.59	17.66	17.73	17.80	17.89	17.89

לסיום ההשוואה, מציג לוח 5.3.4 את מטריצת המעבר בין הציון הפסיכומטרי לציון הבגרות.

<u>לוח 5.3.4 : מטריצת מעבר: עשירוני ציונים – בגרות באנגלית ופסיכומטרי באנגלית</u>

פסיכומטרי באנגלית	בגרות 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	1.0	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	1.0	1.7	2.4	2.4	1.5	10
9			0.1	0.1	0.4	0.7	1.4	2.0	2.6	2.6	10
10				0.1	0.1	0.3	0.8	1.4	2.3	5.0	10
סה"כ	10	10	10	10	10	10	10	10	10	10	100

כפי שניתן לראות, כ 28.45 אחוזים מהאוכלוסייה של הנבחנים נשארים באותו עשירון ציונים, כ 35 אחוזים נמצאים מתחת לאלכסון וכ 36 אחוזים נמצאים מעל האלכסון. 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 עולה המתאם בין שתי מערכות הציונים.

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 בגלל שהרווח ממיצוע גבוה יותר ככל שהמתאם בין הציונים נמוך יותר.

5.4. מתמטיקה – 3, 4 ו-5 יחידות בגרות

לאור המתאם הגבוה שקיבלנו באנגלית, החלטנו לבדוק מתמטיקה של תלמידים שנבחנו 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 17,867 נבחנים בשנים 1999-2000, מתוכם 4,092 – 3 יחידות, 6,716 – 4 יחידות, 7,059 – 5 יחידות. השיקלול של הציונים הוא בהתאם לשקלול שמעניקות האוניברסיטאות.

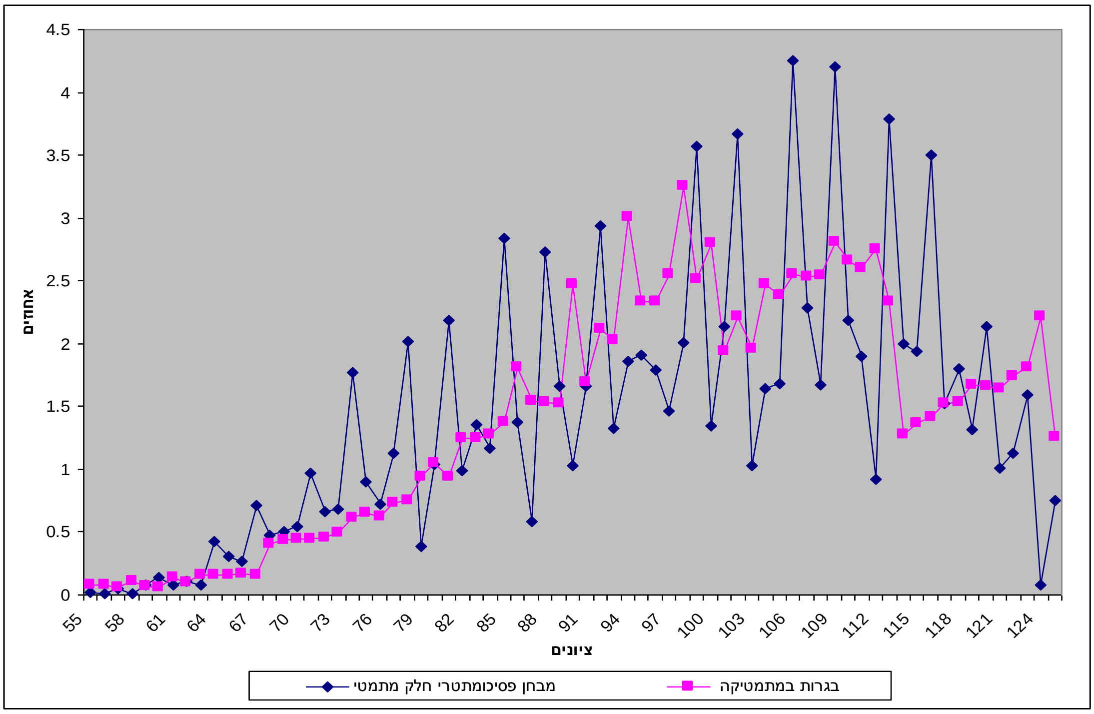
לוח 5.4.1: פרמטרים מרכזיים של שני המבחנים במתמטיקה 3 ו 4 ו5 יחידות

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				

השוואת הציונים שהתקבלו (לוח 5.4.1) בין הבגרות לפסיכומטרי מראה שהציון בבגרות גבוה בממוצע ב 1.5 נקודות אולם ממצא זה אינו משמעותי מאחר והוא תלוי בצורת הנרמול שנעשתה. סטיית התקן ומקדם ג'יני מראים שהפיזור של הציון גדול יותר בפסיכומטרי ואילו מקדמי המתאם מראים מידה קטנה יותר של מתאם בין הציונים מאשר באנגלית. מקדמי המתאם של הג'יני שווים מה שמעיד שיש לצפות לסימטריה בהתפלגות הציונים בשתי הבחינות.

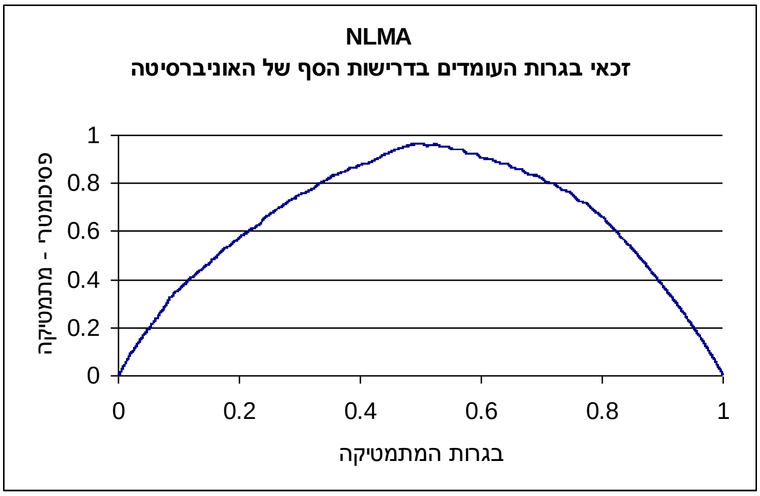
ציור 5.4.1 מצביע על שכיחות גבוהה יותר בציוני הבגרות הגבוהים ועל כך שציוני הבגרות מתפלגים בצורה רציפה יותר מאשר ציוני הפסיכומטרי.

ציור 5.4.1 : התפלגות ציוני פסיכומטרי ובגרות במתמטיקה



ציור 5.4.2 מציג 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.

ציור 5.4.2 : ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה



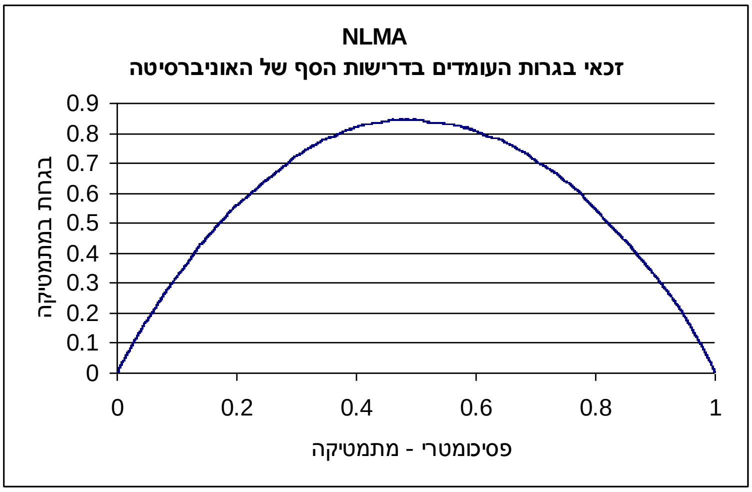
לוח 5.4.2 מציג את רגרסיות הג'יני המורחב עבור הציון בפסיכומטרי במתמטיקה כפונקציה של ציון הבגרות. שורה מספר 3 בלוח מראה את השיפוע של קו הרגרסיה. כאן בולטת חוסר הלינאריות של הקשר כי כאשר מדגישים את התחום התחתון של הבגרות השיפוע הממוצע הוא 0.558 ואילו כאשר מדגישים את החלק העליון, השיפוע הוא 0.65 והשיפוע עולה בצורה מונוטונית, מה שמעיד על עקום רגרסיה קמור יחסית לציר האופקי. כל השיפועים המתקבלים נמוכים יותר מאשר הקשרים המתקבלים באנגלית מה שמעיד שיכולת הניבוי של הציון במתמטיקה נמוכה יותר מיכולת הניבוי באנגלית והאלמנט האקראי במתמטיקה גדול יותר מהאלמנט האקראי באנגלית.

לוח 5.4.2: רגרסית ג'יני מורחב– ציון פסיכומטרי במתמטיקה כפונקציה של ציון בגרות במתמטיקה.

	v=0.5	v=1	v =1.5	v=2	v=3	v=5	v=7
α (ממוצע)	33.51	34.93	36.14	37.21	38.95	41.33	42.79
α (חציון)	34.51	35.97	37.23	38.30	40.09	42.48	43.97
β (בגרות במתמטיקה)	0.650 (0.006)	0.636 (0.006)	0.624 (0.006)	0.613 (0.007)	0.596 (0.007)	0.572 (0.007)	0.558 (0.007)
מתאם ג'יני Γyx	0.621	0.603	0.588	0.577	0.560	0.540	0.532
מתאם ג'יני Γxy	0.611	0.605	0.599	0.594	0.584	0.565	0.551
GR	0.311	0.205	0.211	0.212	0.210	0.200	0.191
α (ממוצע)+ β(v)* $\overline{X}$	98.71	98.72	98.73	98.69	98.73	98.70	98.76
β (v)*($X_{min}-\overline{X}$**)**	-29.45	-28.81	-28.27	-27.77	-27.00	-25.91	-25.28
β (v)*($X_{max}-\overline{X}$**)**	16.06	15.71	15.41	15.14	14.72	14.13	13.78

ציור 5.4.3 מציג את עקומת ה NLMA של הציון במתמטיקה בבגרות כפונקציה של הציון הפסיכומטרי. כצפוי קיבלנו קשר מונוטוני יציב שמשמעותו ששני המבחנים בוחנים את אותו נושא.

ציור 5.4.3 : ציון בגרות במתמטיקה כפונקציה של ציון פסיכומטרי במתמטיקה



לוח 5.4.3 מציג את רגרסיות הג'יני המורחב של הציון בבגרות כפונקציה של הציון בפסיכומטרי במתמטיקה. גם כאן בולט הקמור אם כי הוא חלש יותר ברמתו. ההבדלים בין השיפוע כאשר מדגישים את החלק התחתון (0.525) לבין השיפוע כאשר מדגישים את החלק העליון (0.585) מעידים על קשר קמור. מאחר וגם כאשר אנו מסתכלים על הבגרות כפונקציה של הפסיכומטרי אנו מקבלים קשר קמור וגם כאשר אנו הופכים את הקשר אנו מקבלים קשר קמור המסקנה המתקבלת היא שככל שמדגישים את החלק העליון של ההתפלגויות קטן חלקה של הטעות האקראית ואנו מתכנסים לקשר חזק יותר בין ההצלחה במבחן האחד לבין ההצלחה במבחן השני.

לוח 5.4.3 : רגרסית ג'יני מורחב – ציון בגרות במתמטיקה כפונקציה של ציון פסיכומטרי במתמטיקה.

	v=0.5*	v=1	v =1.5*	N=2	v=3	v=5	v=7
α (ממוצע)	42.56	43.67	44.49	45.15	46.17	47.54	48.44
α (חציון)	43.75	44.88	45.73	46.37	17.39	48.77	49.67
β (פסיכומטרי במתמטיקה)	0.585 (0.006)	0.573 (0.006)	0.565 (0.006)	0.558 (0.006)	0.548 (0.006)	0.534 (0.007)	0.525 (0.007)
מתאם ג'יני Γyx	0.612	0.605	0.599	0.594	0.584	0.565	0.551
מתאם ג'יני Γxy	0.662	0.603	0.588	0.577	0.560	0.540	0.532
GR	0.204	0.205	0.201	0.197	0.190	0.180	0.173
α (ממוצע)+ $\beta(v)*\overline{X}$	100.30	100.23	100.26	100.22	100.26	100.25	100.26
$\beta(v)*(X_{min}-\overline{X})$	-25.56	-25.04	-24.69	-24.38	-23.95	-23.34	-22.94
$\beta(v)*(X_{max}-\overline{X})$	15.39	15.07	14.86	14.68	14.41	14.04	13.81

לוח 5.4.4 מציג את מטריצת המעבר ממערכת ציונים אחת לשניה. ניתן לראות שרק 20.2 אחוזים מהנבחנים נשארים על האלכסון כלומר רק כ 20 אחוזים מהתלמידים נשארים באותו עשירון של ציונים. כ 39 אחוזים נמצאים מתחת לאלכסון ואילו 41 אחוזים נמצאים מעל לאלכסון. ניתן לראות שהפיזור של התלמידים מעבר לאלכסון גבוה יותר במתמטיקה מאשר באנגלית. אולם יש לציין שמתקבל הרושם שאחוז הטעויות האקראיות של כל בחינה הוא גבוה. כך למשל, רק 35 אחוזים מהתלמידים שהיו בעשירון העליון של אחד המבחנים

נמצאים בעשירון העליון גם במבחן השני, ולעומת זאת רק 40 אחוזים מהתלמידים שנמצאו בעשירון התחתון באחד המבחנים נשארים בעשירון התחתון גם במבחן שני.

<u>לוח 5.4.4 : מטריצת מעבר: עשירוני ציונים – בגרות במתמטיקה ופסיכומטרי במתמטיקה</u>

פסיכומטרי במתמטיקה	בגרות במתמטיקה										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	סה"כ
	1	3.5	2.0	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	1.0	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	4.0	10
	סה"כ	10	10	10	10	10	10	10	10	10	10	100

האלטרנטיבות שיש לבחון הן: קיום מבחן אחד בלבד, קיום מבחן אחד אולם עם דרישה מכל תלמיד להיבחן פעמיים או קיומם של המבחנים במתכונת הנוכחית. במקביל, כדאי לחזור עם נתוני שנים שונות על מחקר זה כדי לבדוק את המידה בה הממצאים שנמצאו יציבים על פני זמן.

6. מידת השיפור בעמידות (robustness) הציון

בסעיף הקודם הראינו שמבחני הבגרות והפסיכומטרי בתחומי המתמטיקה ואנגלית 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 ולהיפך: ככל שהציון בפסיכומטרי גבוה יותר כן יש לצפות שהציון בבגרות יהיה גבוה יותר. בסעיף זה נרצה לבדוק את הטענה שהצרוף 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 באותו רמה. היות המתאמים 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 קטנה יותר מהצירוף של הציונים במתמטיקה.

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

הנוסחה היא:[26]

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

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

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

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

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

כאשר נוסחה (6.2) מקבילה לנוסחה שהיתה מתקבלת אם היינו מבקשים לראות את התרומה של כל מרכיב למקדם ההשתנות של הציון הכולל.[27] נוסחה (6.1) המתאימה לג'יני כוללת כמקרה פרטי את נוסחה (6.2) והיתרון שלה 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. כפי שניתן לראות הפיזור על פי מדד ג'יני בפסיכומטרי גדול יותר מאשר בבגרות (0.082 לעומת 0.077) ואילו הציון הממוצע גבוה במעט בפסיכומטרי מאשר בבגרות. המדד לציון המשוקלל קטן רק במידה מזערית מהמדד בבגרות, תוצאה שניתן לייחס אותה למתאם הגבוה שנמצא בין ציוני שני המבחנים.

<u>לוח 6.1: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות באנגלית מול פסיכומטרי באנגלית</u>

	בגרות באנגלית	פסיכומטרי באנגלית	סה"כ
G	0.077	0.082	0.076
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

החלק השני בלוח 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.[28] גם המתאם עם הציון הכללי

[26] פיתוח הנוסחה נמצא ב Yitzhaki and Schechtman (2013)
[27] מקדם ההשתנות הוא סטיית התקן מחולקת בממוצע של המשתנה.

נראה דומה בשתי מערכות הציונים ומכאן שאנו מקבלים שמאחר והציון 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 גיני בציון הכולל. התרומה מתחלקת בערך כ 60 אחוזים למרכיבים העצמאיים של הציון ו-40 אחוזים למתאם בין הציונים. המתאם בין תוצאות שני המבחנים משפיע על תוצאת הציון הכולל.

לוח 6.2 בודק את מדד ג'יני של הציון במתמטיקה בבגרות לעומת הציון בבגרות באנגלית. כפי שניתן לראות אי השוויון בציוני המתמטיקה גבוה יותר מאי השוויון באנגלית ואילו האי השוויון בסכום תוצאות המבחנים נמוך מאי השוויון בציונים בכל מבחן. הסבר לתוצאה זו ניתן לקבל מהמתאם היחסית נמוך בין הציונים (0.5). הסימטריה של המתאם מעידה שההתפלגות נורמלה בשני המבחנים בצורה דומה. גם קיימת סימטריה במתאם בין כל ציון לסכום הציון בשני המבחנים. המתאם הנמוך יחסית אומר שקיים טעם בהסתכלות על סכום הציונים במתמטיקה ואנגלית בבגרות, מאחר והם מוסיפים אינפורמציה שונה על התלמיד. המתאם הנמוך יחסית גם מסביר מדוע קיימת ירידה באי שוויון בסכום הציונים יחסית לכל ציון בנפרד.

<u>לוח 6.2: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות באנגלית מול בגרות במתמטיקה</u>

	בגרות באנגלית	בגרות במתמטיקה	סה"כ
G	0.077	0.082	0.068
δ	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 6.3 מבצע חזרה על לוח 6.2 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 גם כאן הציון הממוצע באנגלית גבוה יותר מאשר הציון במתמטיקה. גם המתאם בציונים באנגלית ומתמטיקה גבוה יותר מאשר בציוני הבגרות וכתוצאה מכך גם המתאם בין כל מרכיב מול הציון הכולל. התוצאה הינה שמשקל כל מרכיב בציון הכולל גבוה יותר מהמשקל בציוני הבגרות המקבילים.

---

[28] תוצאה זו יכולה לנבוע גם מכך ששתי מערכות הציונים עוברות תהליך דומה של נירמול.

<u>לוח 6.3 : פירוק מדד ג'יני בציון הכולל לג'יני של פסיכומטרי באנגלית מול פסיכומטרי במתמטיקה</u>

	פסיכומטרי באנגלית	**פסיכומטרי במתמטיקה**	**סה"כ**
G	0.082	0.087	0.077
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	**$G_O \Sigma \delta_i D_{iO} G_i$**	**$\delta_i^2 G_i^2$**	**$\Sigma\Sigma\ \delta_i \delta_j G_i G_j \Gamma_{ij}$**
0.00596	-0.00003	0.00357	0.00242

לוח 6.4 חוזר על הטבלות הקודמות כאשר הפעם ההתייחסות היא לציון של בגרות במתמטיקה מול פסיכומטרי במתמטיקה. אי השוויון בסכום הציונים נמוך יחסית מאי השוויון בתוך כל מרכיב אם כי ההבדל שמתקבל נמוך יותר מאשר המקבילה באנגלית. זאת כי המתאם נמוך יותר מאשר המתאם בין פסיכומטרי באנגלית ובמתמטיקה.

<u>לוח 6.4: פירוק מדד ג'יני בציון הכולל לג'יני של בגרות מתמטיקה מול פסיכומטרי במתמטיקה</u>

	בגרות במתמטיקה	**פסיכומטרי במתמטיקה**	**סה"כ**
G	0.082	0.087	0.076
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

המסקנות המתקבלות מתוך הניתוח דלעיל הן שהמבחנים הפסיכומטריים במתמטיקה ואנגלית 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 באותן מקצועות בבגרות ואם ההשערה דלעיל נכונה אזי ניתן להסיק ששיטת הבחינה יש לה השפעה על המתאם המתקבל בהצלחות בבחינות השונות.

7. 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 אקראית 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.

כדאי להדגיש שאין במחקר בכדי 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) לא בודקים תחומים שונים שיש לנו את היכולת להבחין ביניהם, וזאת על סמך ניתוח מערכת התוצאות של המבחנים.

ספרות:

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*, Readind, Massachusetts: Adison-Wesley.

Schechtman, E., S. Yitzhaki and Yevgeny Artsev (2008). Who Does Not Respond in the Household Expenditure Survey: An exercise in Extended Gini Regressions, *Journal of Business & Economic Statistics*, 26, 3, July, 329-344.

Schechtman, E. and S. Yitzhaki (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 Maggie Eisenstaedt (2003). Ranking Groups' versus Individuals' Ranking. In Amiel, Yoram and John A. Bishop (eds.) *Fiscal Policy, Inequality, and Welfare , Research on Economic Inequality*, 10, Amsterdam: JAI, 101-123.

Yitzhaki, S. and E. Schechtman (2012). Identifying monotonic and non-monotonic relationship, *Economics Letters,* 116, 23-25.

Yitzhaki, S; Rinat Itzhaki and Taina Pudalov (2012). A nonparametric ICC using the Gini's Mean Difference Approach. Unpublished, http://ssrn.com

Yitzhaki, S. and E. Schechtman (2013). *The Gini Methodology: A Primer on a Statistical Methodology.* Springer:N.Y. Forthcoming.

Abstract

Matriculation and psychometric examinations have been conducted in Israel for many years. The need to take both exams is a burden on students who are required to prepare for two different tests, as well as a burden on the Israeli economy. This research is a preliminary study of the costs and benefits of having two examinations (matriculation and psychometric) in the same fields. In addition, the study investigates the extent to which these examinations test the same abilities.

The contribution of this study lies in the research method, which is non-parametric. This method takes into consideration that in contrast to measurement of quantitative variables, there is no possibility for direct measurement of variables in the field of education. Rather, it is only possible to ask questions that can broaden our knowledge about the respondents. Therefore, scores on research questionnaires depend on the difficulty distribution of the questions.

Our methodology is based on the following argument: We examine the correlation between the scores of examinees on matriculation and psychometric exams. If the correlation is monotonic throughout the set of the scores, we can conclude that one of the tests is redundant. However, if the sign of the correlation changes throughout the set of scores, we can conclude that the tests examine different fields of knowledge, or that the characteristics needed for success in one exam are different from those needed for success in the other exam.

In response to the research question, it can be concluded that the matriculation and psychometric examinations test the same abilities, but there is a large random error. The contributions of both exams to the total score for each student support the hypothesis that the existence of two examinations does not have a substantial impact. This conclusion applies all the more to the English exams, where the correlation between the matriculation and psychometric scores was higher than in the mathematics exams.

It should be emphasized that the findings of this research do not provide a basis for deciding which of the exams should be eliminated. The findings only suggest that matriculation and psychometric tests in corresponding subjects (English and Mathematics) do not examine different areas that can be distinguished from one another.

Key words: Matriculation Examination, Psychometric Examination, Monotony, Gini Regression.

We wish to thank Dmitri Romanov and antonymic reviewer for their helpful comments.

The Central Bureau of Statistics (CBS) encourages research based on CBS data, such as this work. Works of research of this sort are not official publications of the CBS, and therefore the opinions and conclusions expressed in these publications are those of the authors and do not necessarily represent those of the CBS.

Published by the Central Bureau of Statistics, 66 Kanfe Nesharim St.,

Corner Bachi St., P.O.B 34525, Jerusalem 91342, Israel

Tel. 972-2-6592666; Fax: 972-2-6521340

Internet Site: www.cbs.gov.il

E-Mail: info@cbs.gov.il

WORKING PAPER SERIES

NO. 75

Matriculation and Psychometric Examinations in Mathematics and English, and the Relationship between Them

Shlomo Yitzhaki*, Taina Pudalov** and Aviel Krentzler**

May, 2013

* Central Bureau of Statistics and Hebrew University
** Central Bureau of Statistics