

**השפעה של
היבחנות חוזרת
על הישגים במבחנים
פסיכומטריים**

ענת בן-סימון
יואב כהן

הקדמה

פרק זה עוסק בשאלת השיפור בציון המושג על ידי נבחנים הניגשים בפעם השנייה לבחינה הבודקת כשרים סכולסטיים.

לפרק שני חלקים:

החלק הראשון הינו מבוא עיוני ובמסגרתו נסקור ממצאים קודמים, שפורסמו בארץ ובחו"ל, המתייחסים לשאלת שיפור הציון. שלושה נושאים ידונו בהקשר זה:

- I. האם קיים שיפור בבחינה חוזרת, ואם כן מה שיעורו?
- II. אפקט הבחירה העצמית, היינו, מי הם הנבחנים שנבחנים שנית, מה מייחד אותם מאוכלוסיית הנבחנים הכללית, או, במלים אחרות, מהי השפעת הרגרסיה אל הממוצע הנובעת מאפקט הבחירה העצמית, על ציוני השיפור.
- III. האם ובאיזו מידה תלוי שיפור הציון בבחינה חוזרת בסוג הבחינה (התוכן הנבדק) ובדמיון בין הבחינה הראשונה והשנייה.

החלק השני של הפרק מוקדש לבדיקה אמפירית של שאלות אלו. הנתונים לצורך מחקר זה הם ציוני נבחנים חוזרים אשר נבחנו פעמיים לפחות בבחינה הפסיכומטרית של המרכז הארצי לבחינות והערכה.

ביתר פירוט, יבדקו השאלות הבאות:

- I. מהי שכיחות הנבחנים החוזרים בכל שנת בחינה?
- II. מהו שיעורו של השיפור המושג בבחינות חוזרות של המרכז הארצי לבחינות.
- III. מיהם הנבחנים החוזרים, מהם הישגיהם בבחינה, והאם הישגים אלו מעידים על אפקט הבחירה העצמית.
- IV. מהן מהימנויות המבחן החוזר במבחנים השונים של הבחינה הפסיכומטרית. מהימנויות אלו קובעות במידה רבה את אפקט הרגרסיה אל הממוצע בבחינות חוזרות.
- V. הקשר בין השיפור בבחינות חוזרות לבין משתנים שונים: פער הזמן בין בחינות חוזרות, גיל הנבחנים, מינס, ומידת העליה ברמת ההשכלה של הנבחנים בין בחינות חוזרות. חלק מן הבדיקות הנ"ל יושמו גם על הנבחנים ששפת בחינתם לא הייתה עברית.

חלק א' - סקירה כללית

1. מבוא

עשרות אלפי תלמידים בשנת לימודיהם האחרונה בביה"ס התיכון בארה"ב נבחנים כל שנה במבחן ה-SAT (Scholastic Aptitude Test) וזאת לאחר שכבר ביצעו מבחן דומה בשנת הלימודים הקודמת. ציוני ה-SAT הם על סולם שטוח הציונים בו הוא 800-200, ממוצעו בסביבות 460 וסטית התקן היא 100 בקירוב. ה-ETS (Educational Testing Service), שהוא הארגון האחראי לבניית והעברת ה-SAT, מודע להתענינותם של הנבחנים בשאלת ההפרש בין ציוני בחינתם הראשונה והשנייה. בעלון הסבר הנשלח לבתי-ספר ומכללות (ETS, 1982) מצוין כי 65% מן הנבחנים החוזרים משיגים ציון גבוה יותר בבחינתם השנייה; כן מצוין שיש סיכוי של 5% להשיג שיפור של 100 נקודות לפחות (סטית תקן שלימה) בציון, וסיכוי של אחוז אחד להפסיד 100 נקודות לפחות.

רוב השונות של ההפרשים בציונים ניתנת להסבר במונחי טעות המדידה, אך לא כך לגבי ממוצע השיפור בציונים. אין ספק שהשיפור הוא אמיתי אך יש לפחות הטייה אחת במחקרים הבודקים את אפקט השיפור ומקורה בעובדה שהתלמידים עצמם הם המחליטים אם להיבחן שנית או לא. תופעה זו, הקרויה בחירה עצמית (self selection), עשויה להטות את אפקט השיפור כלפי מעלה. אחרי הכל - אם בפעם השנייה נבחנים רק אותם תלמידים המעריכים כי ציונם בבחינה הראשונה אינו משקף את יכולתם האמיתית, ואם אמנם הערכתם היא נכונה, כי אז אפקט הבחירה העצמית עשוי להיות מרכיב חשוב בשיפור הממוצע בבחינה חוזרת.

2. אפקט הבחירה העצמית

אם אפקט הבחירה העצמית אמנם קיים כי אז טעויות המדידה בציוניהם של הנבחנים הבורחים להיבחן שנית תהיינה שליליות במדידה הראשונה. Alderman (1981a), מצא דרך אלגנטית לבדוק אם אמנם כך הדבר. הוא בדק את ביצועם של התלמידים שנבחנו פעם אחת בלבד ("בעלי בחינה אחת") ואת ביצועם של אלה שנבחנו פעמיים ("חוזרים"). בשתי הקבוצות הישווה את הציונים בחלק התפעולי של מבחן ה-SAT (הציונים המדווחים לנבחן) לציוניהם של הנבחנים בחלק נסיוני של המבחן - ציונים שאינם מדווחים לנבחנים. בנוסף, הישווה את ציוני הנבחנים בשתי הקבוצות לעומת ניבוי ציונים אלה על-ידי משתני רקע שונים.

אם אמנם טעויות המדידה בקבוצת החוזרים הן שליליות, כי אז ממוצע ציוני השארית (ההפרש בין הציון הנצפה לבין הציון המנובא) בקבוצה זו צריך אף הוא להיות שלילי.

מהשוואת נתוני המנבאים והמנובאים של שתי הקבוצות עולה, כי בקבוצת החוזרים היו הציונים בחלק המדווח ובחלק הנסיוני של המבחן נמוכים יותר בממוצע, ופיזורם קטן יותר. כמו-כן, אחוז גבוה יותר של החוזרים השתתפו בתכניות קדם-אקדמיות, התכוונו

להמשיך לימודיהם לקראת תואר ראשון ולמדו מתמטיקה בבית הספר התיכון במשך שלוש שנים ומעלה.

אפקט הבחירה העצמית במחקר זה נבדק על שני מבחנים שונים; בחינת ה-SAT מחולקת למבחן מתמטי (SAT-M) ומבחן מילולי (SAT-V) והציונים של שני מבחנים אלו שימשו כקריטריונים לבדיקת האפקט.

אומדן מקדמי הרגרסיה לניבוי הציון התפעולי על-סמך הציונים בחלק הנסיוני של המבחן נעשה על כשליש מקבוצת בעלי מבחן יחיד. תקפותה של משוואת הרגרסיה נבדקה על שני מדגמים נוספים (שני השלישים הנותרים) של בעלי מבחן יחיד ולא נמצא הבדל מובהק בין שלוש הקבוצות. אותה משוואה שימשה גם לניבוי ציוניהם של נבחנים מקבוצת החוזרים, שחולקו לצורך זה לשלוש קבוצות. גם בין שלוש קבוצות אלו לא נמצא הבדל מובהק בממוצע טעות הניבוי; אך בין קבוצת בעלי בחינה יחידה ובין קבוצת החוזרים נמצא הפרש מובהק בממוצע טעויות הניבוי ובכיוון המצופה - כפי שמוצג בטבלה 2.1. בעוד שבקבוצת בעלי הבחינה היחידה ממוצע טעויות הניבוי הוא חיובי, הרי שבקבוצת החוזרים הממוצע הוא שלילי, וגודל אפקט הבחירה העצמית הוא כ-5% נקודות ב-SAT-V ו-9 נקודות ב-SAT-M: כלומר, בין 5%-ל-10% מסייגת התקן של הציונים עצמם.

טבלה 2.1: ממוצעים וסטיות תקן של ציוני השארית (ציון שיפור נצפה פחות ציון שיפור מנובא).

SAT - M			SAT - V			
ממוצע	סט. ת.	n	ממוצע	סט. ת.	n	
55.75	0.58	13,187	50.96	0.20	13,355	בעלי בחינה יחידה
54.55	-9.19	12,235	49.51	-4.84	12,438	בעלי בחינה חוזרת

לדעת Alderman, מחקר זה מוכיח שאפקט הבחירה העצמית אמנם קיים ויש להביאו בחשבון בניתוח אפקט שיפור הציון. מהו גודלו המדויק של האפקט - התשובה לשאלה זו משתנה, כנראה, ממצב למצב. מחד, האומדן לאפקט הבחירה העצמית הוא בבחינת הערכת יתר, מאחר שהוא משקף גם את טעויות המדידה החיוביות בקרב הנבחנים הבורחים שלא לחזור ולהיבחן שנית. מאידך, ישנו אלמנט מסוים בטעות המדידה שכלל אינו נלקח בחשבון במחקר זה, ועלול לגרום להערכת חסר של אפקט הבחירה העצמית; אותם נבחנים שציונם כלל גורם טעות שלילי גם בחלק הלא-תפעולי של המבחן (לדוגמה, בשל מחלה או עייפות) לא גילו טעות ניבוי שלילית שהיא המשתנה האמור במחקר זה למדוד את אפקט הבחירה העצמית.

בדיווחו של Alderman אין התייחסות לבעיה סטטיסטית שהתעלמות ממנה עלולה להביא להערכת יתר של אפקט הבחירה העצמית. שורש הבעיה נעוץ בכך שמלכתחילה השיגו בעלי הבחינה היחידה ציונים גבוהים יותר. מקדמי הרגרסיה נאמדו על נתוני קבוצה זו וברור שהפעלתם על קבוצה כלשהי של נבחנים שמוצאם בחינתם נמוך יותר, תביא לטעויות ניבוי שליליות. כאמור, אין התייחסות לבעיה זו במאמרו של Alderman, ואף כי אין בכך משום הסבר מלא של אפקט הבחירה העצמית, התעלמות מן הבעיה בוודאי תורמת להערכת יתר של האפקט.

3. בחינות חוזרות - תוצאות על מבחני SAT

מבחני ה-SAT הם אולי המקור האמין ביותר לקבלת אומדנים על הפרמטרים השונים, הקשורים לאפקט הבחינות החוזרות. זאת משום איכותם הגבוהה של המבחנים, ההקפדה על סילום נכון של הציונים והמספרים הגדולים של נוסחי בחינה ונבחנים.

נתונים שנאספו על בחינות חוזרות על-ידי ה-ETS במשך מספר שנים (Angoff, 1971) מראים כי השיפור הממוצע בציוני ה-SAT הוא כ-23 נקודות, כלומר, כרבע עד חמישית מסטיית התקן של האוכלוסייה הנבחנת. נתונים אלו אינם יציבים כי אם משתנים משנה לשנה, בין היתר משום שבהשוואת ציוני שתי בחינות נכללת גם טעות בכיול הנוסחים השונים של הבחינה. טעות הכיול אינה גדולה בהקשר של ציוני נבחנים בודדים, אך היא קבועה ועל-כן משפיעה על ממוצעי קבוצות נבחנים. סטיית התקן של טעות הכיול היא כ-5 נקודות בסולם ציוני ה-SAT. כאשר מחשבים ציוני הפרש, מרכיב הטעות הנובע מן הכיול יכול להגיע לכ-7 נקודות וכאשר משווים ציוני הפרש בין שתי קבוצות טעות זו מגיעה לכ-10 נקודות. אך כאשר נעשות ההשוואות מעבר לנוסחים רבים, ניתן להגיע להכללות מדויקות.

טעות התקן של המדידה בציוני ה-SAT היא בסביבות 30 נקודות, טעות התקן של מדידת ציוני ההפרש היא כ-42 נקודות בסולם ה-SAT. מהימנותם של ציוני ההפרש נמוכה - בסביבות 0.20, אך המתאם של ציוני בחינה ראשונה ושניה מגיע ל-0.88.

מנתוני Angoff ניתן לחשב את השפעת הזמן שחולף מן הבחינה הראשונה לשניה על השיפור בציון (טבלאות 2.20 ו-2.21; Angoff, 1971). אם השיפור בציון בבחינה חוזרת הוא כולו תוצאה של חשיפה לבחינה הראשונה, כי אז ניתן לשער שבמהלך הזמן אפקט השיפור ילך וירד. לעומת זאת, אם אפקט השיפור נובע גם מצמיחה אישית של הנבחנים (רכישת ידע ונסיון בבית הספר או מחוצה לו) אזי ניתן לצפות לגידול בשיפור עם חלוף הזמן בין הבחינה הראשונה לשניה. בטבלה 3.1 מוצג השיפור כפונקציה של מספר החודשים שחלפו בין בחינת SAT ראשונה לשניה. הנתונים מבוססים על מיצוע של מאות-אלפי תלמידים שנבחנו בבחינה חוזרת בין השנים 1962 ו-1968.

טבלה 3.1: אפקט השיפור בציוני SAT כפונקציה של מספר החודשים בין הבחינה הראשונה לשנייה.

	7	8	9	10
SAT-V	17	22	25	30
SAT-M	18	24	22	29

הנתונים מצביעים על גידול באפקט השיפור ככל שמשך הזמן בין הבחינה הראשונה לשנייה מתארך, לפחות בטווחי הזמן שנבדקו - 7 עד 10 חודשים. ניתן, איפוא, להסיק, שמרכיב חשוב בשיפור הציון בבחינה חוזרת הוא הצמיחה האישית של הנבחנים בזמן שחלף בין בחינה לבחינה.

Angoff מצייין שכל הנסיונות לתאם ולתקף את אפקט השיפור עם קריטריון חיצוני לא הביאו לתוצאות כלשהן. המתאם העקבי היחיד הוא של ציון השיפור עם הציון בבחינה הראשונה. מתאם זה הוא בסביבות 0.25- וכפי שניתן ללמוד בטבלה 2.1, הוא מוסבר כולו ע"י הרגרסיה אל הממוצע.

יש לזכור שהנתונים המובאים כאן הם נתונים תיאוריים גולמיים ללא נסיון לנכות מאפקט השיפור בבחינה חוזרת הטיות הנובעות מן הבחירה העצמית ומן הרגרסיה אל הממוצע.

4. אפקט הבחירה העצמית ואפקט הרגרסיה אל הממוצע

בנסיון לנכות שני אפקטים אלו מאפקט הבחינה החוזרת השתמש Alderman (1981b) בשיטה שפותחה ע"י F. Lord. הרעיון המרכזי בשיטה זו הוא לנסות לאמוד את ההפרש בין הציונים האמיתיים בבחינה הראשונה והשנייה באמצעות שימוש בציוני הנבחנים על מבחנים נסיוניים המודדים אותה יכולת (ר' לעיל סעיף 2); יהא X הציון הנצפה ב-SAT-V או SAT-M המדווח לנבחן. יהא Y ציונו של הנבחן, במבחן נוסף, המועבר יחד עם המבחן התפעולי אך אינו מדווח לנבחן.

X מורכב מציון אמיתי T וטעות מדידה E:

$$X = T + E \quad (4.1)$$

בדומה, Y מורכב מציון אמיתי aT+b, כאשר a ו-b הם קבועי הסילוס, ומטעות מדידה F. מובלעת כאן הנחה ש-X ו-Y מודדים אותה יכולת על סולמות זהים עד כדי טרנספורמציה לינארית:

$$y = aT + b + F \quad (4.2)$$

קבוצת נבחנים המקבלים ציונים X_1 ו- Y_1 , בבחינה הראשונה, מחליטים להיבחן שנית, ומקבלים בבחינה השנייה ציונים X_2 ו- Y_2 . כמו כן, קבוצה נוספת של נבחנים

הנדגמים בצורה אקראית מתוך אלו שנבחנו בבחינה הראשונה מקבלים ציונים X_0 ו- Y_0 . ניתן להניח שעבור כל הציונים, תוחלת טעויות המדידה שווה לאפס. פרט ל- E_1 - טעות המדידה בבחינה הראשונה של הנבחנים פעמיים; זאת משום שהחלטתם של אלו להבחן שנית מושפעת כנראה מטעות מדידה שלילית.

כאשר ממצעים מעבר לנבחנים ניתן לנסח את המשוואות הבאות:

$$\bar{X}_0 = \bar{T}_0; \bar{Y}_0 = a\bar{T}_0 + b \quad (4.3)$$

$$\bar{X}_1 = \bar{T}_1 + E_1; \bar{Y}_1 = a\bar{T}_1 + b \quad (4.4)$$

$$\bar{X}_2 = \bar{T}_2; \bar{Y}_2 = a\bar{T}_2 + b \quad (4.5)$$

משוואות אלו ניתן לחלץ את קבועי הסילום ואת \bar{T}_1 כלהלן: מחיסור (4.3) מ- (4.5) נקבל את ערכו של a :

$$a = (\bar{Y}_2 - \bar{Y}_0) / (\bar{X}_2 - \bar{X}_0) \quad (4.6)$$

מתוך (4.5) ניתן לקבל את ערכו של b :

$$b = \bar{Y}_2 - a\bar{X}_2 \quad (4.7)$$

מתוך (4.4) נחלץ את:

$$\bar{T}_1 = (\bar{Y}_1 - b) / a \quad (4.8)$$

ועכשיו ניתן לחשב את ההפרש האמיתי בין ממוצעי ציוני הנבחנים החוזרים:

$$\bar{T}_2 - \bar{T}_1 = \frac{\bar{X}_2 - \bar{X}_0}{\bar{Y}_2 - \bar{Y}_0} (\bar{Y}_2 - \bar{Y}_1)$$

ההפרש בין ממוצעי הציונים האמיתיים ניתן, איפוא, לאמידה, על-ידי ממוצעי הנבחנים בחלקים התפעוליים והנסיוניים של המבחן. אמדן זה ניתן להשוות להפרש בין הציונים

הנצפים: $\bar{X}_0 - \bar{X}_1$ ותוצאת ההשוואה היא אמדן ל- E_1 : ממוצע מרכיב הטעות בציונם הראשון של הנבחנים החוזרים.

עבור נבחן בודד אמדן ציון השיפור האמיתי הוא:

$$T_2 - T_1 = \frac{\bar{X}_2 - \bar{X}_0}{\bar{Y}_2 - \bar{Y}_0} (Y_2 - Y_1) + F_1 - F_2 \quad (4.10)$$

עבור קבוצת מועמדים בעלי אותו ציון התחלתי X_1 , ניתן למצע את (4.10) ולקבל אומדן של $T_2 - T_1$ עבור ציון התחלתי קבוע. אומדן זה הוא אפקט הציון ההתחלתי על השיפור בציון, כשהוא מנוכה מהשפעת הרגרסיה אל הממוצע.

כל החישובים וההשוואות במחקרו של Alderman נעשו בנפרד על שני חלקי מבחן ה-SAT, עבור כל אחד מהם נמדדו X_0, Y_0 ו- X_1, Y_1 ו- X_2, Y_2 . התוצאות מובאות בטבלה 4.1:

טבלה 4.1: ממוצעי הציונים על החלקים התפעוליים (X) והנסיוניים (Y), והשיפור הממוצע בבחינה חוזרת.

	Y ממוצע ס.ת.		X ממוצע ס.ת.		n	
SAT-V	7.94	16.26	107.33	429.76	1,800	$:Y_0, X_0$
	7.43	16.87	98.24	439.93	1,325	$:Y_1, X_1$
	7.76	18.63	98.21	452.42	1,325	$:Y_2, X_2$
				12.49	1,325	$:X_2 - X_1$
SAT-M	6.06	10.25	114.81	481.75	1,755	$:Y_0, X_0$
	5.69	10.24	106.58	478.45	1,312	$:Y_1, X_1$
	5.68	10.84	105.79	495.55	1,312	$:Y_2, X_2$
				17.10	1,312	$:X_2 - X_1$

השיפור הנצפה בנתונים אלו הוא, איפוא, בין שישית לשמינית סטיית התקן של הציונים. המתאם $r_{X_1X_2}$ הוא 0.88 עבור SAT-V ו- 0.87 עבור SAT-M. ומתקרב למהימנות ה-SAT כפי שנמדדת על נוסח יחיד. גם המתאם בין המבחנים הנסיוניים לבין המבחנים התפעוליים הוא גבוה: $r_{X_1Y_1} = 0.85$.

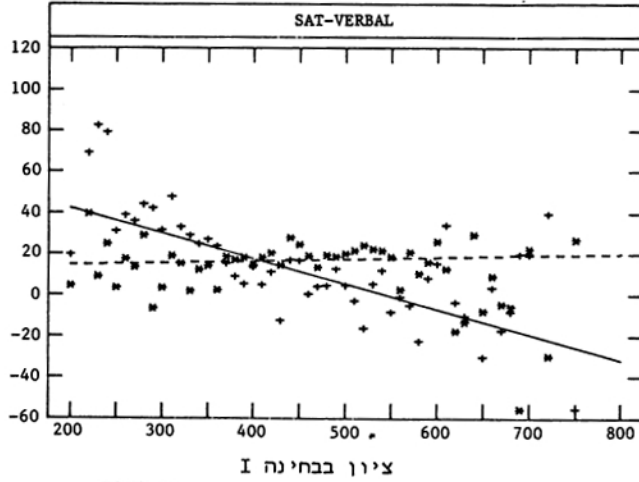
כאשר אומדים את השינוי האמיתי בציונים מתקבלות התוצאות הבאות: עבור SAT-V, בעוד שהשיפור הנצפה הוא 12.49 אמדן השיפור האמיתי הוא 16.84; פירושו של דבר ש- E_1 הוא חיובי ($E_1 = 4.35$) - דבר המצביע על כך שאפקט הבחירה העצמית הוא הפוך מן המצופה במקרה זה.

עבור ה- SAT-M, השיפור הנצפה הוא 17.10 ואילו אמדן השיפור האמיתי הוא 14.03. במקרה זה אמדן E_1 הוא שלילי ($E_1 = -3.07$) ובכיוון המצופה.

בהערת אגב כדאי להוסיף, כי הפרש הזמן בין שתי הבחירות עליהן בוצע המחקר הוא שישה חודשים, ואפקט השיפור המדווח מתאים לנתונים המוצגים בטבלה 3.2 המתארים את אפקט השיפור (הנצפה) כפונקציה של מרווח הזמן בין הבחירות.

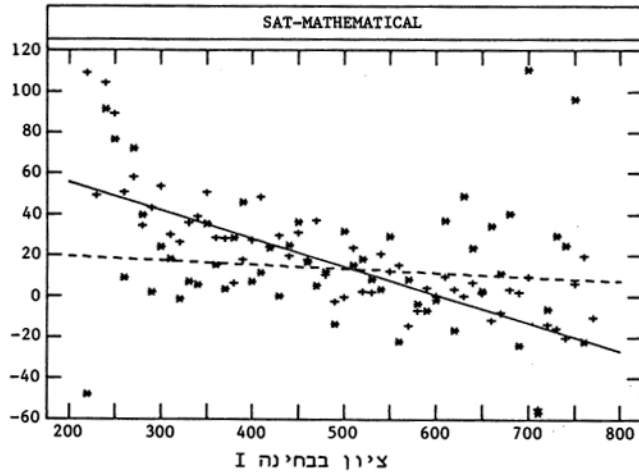
למרות שבמחקר זה לא נתקבל אפקט בחירה עצמית עקבי, הרי שאומדני השיפור האמיתיים בציוני SAT-V ו- SAT-M קרובים זה לזה יותר מן השיפורים הנצפים. נוסף על כך, ניתן, בעזרת אמדני השיפור האמיתי, לבדוק את הקשר בין הציון בבחינה הראשונה לבין השיפור בציון. התרשים המצורף מובא מתוך מאמרו של Alderman (p.9, 1981b) ובו מוצגים השיפורים הנצפים ($X_2 - X_1$) והאמדנים לשיפורים האמיתיים ($T_2 - T_1$) כפונקציה של הציון ההתחלתי (X_1).

שיפור נצפה ואומדן לשיפור אמיתי
כפונקציה של ציון בבחינה I של ה- SAT-V.



+ נצפה
* אומדן

שיפור נצפה ואומדן לשיפור אמיתי
כפונקציה של ציון בבחינה I של ה- SAT-M.



+ נצפה
* אומדן

הקו האחיד הוא קו הרגרסיה הלינארית של ציוני השיפור הנצפים על X_1 ; הקו המרוסק הוא קו הרגרסיה הלינארית של ציוני השיפור האמיתיים. בעוד ששיפועי הקוים עבור השיפור הנצפה הם שליליים - ומבטאים בכך את אפקט הרגרסיה אל הממוצע - הרי שקו הרגרסיה של השיפור האמיתי הם כמעט מאוזנים ומצביעים על כך שאין כנראה כל קשר בין הציון ההתחלתי לבין השיפור האמיתי.

5. גורמי מבחן המשפיעים על אפקט השיפור

עד כמה תלוי אפקט השיפור בדמיון בין שתי הבחינות ובאיזו מידה הוא תלוי בסוג הבחינות עצמן? תשובות על שאלות אלו עשויות להבחיר מהם הגורמים השונים המעורבים באפקט השיפור, מעבר להטיות הנובעות מאפקט הבחירה העצמית ומן הרגרסיה אל הממוצע.

על השאלה הראשונה - השפעת הדמיון בין הבחינה הראשונה והשניה על אפקט השיפור - יש תשובה - אצל Nevo (1976). לטענתו, קשה להשוות תוצאות ממחקרים שונים שטיפלו באפקט השיפור מאחר שהמחקרים נבדלים זה מזה במידת השוני שבין הבחינה הראשונה והשניה. אפקט שיפור הנמדד על שני מבחנים השונים זה מזה בתוכנם, מספק אומדן למידת האימון הכללי של הנבחנים בסיטואציה של מבחן. אם המבחנים הם נוסחים מקבילים, אזי אפקט השיפור מודד גם אימון כללי וגם אימון שהוא ספציפי למבחן, כמו הבנת ההוראות הניתנות עם המבחן ואימוץ אסטרטגיות תשובה ספציפיות. אם שני המבחנים הם זהים, כי אז אפקט השיפור הוא אומדן לאימון כללי ובנוסף לו, אימון ספציפי ברמת המבחן וגם אימון ספציפי ברמת הפריט הבודד. על מנת לבדוד אפקטים אלה, הושוו תוצאותיהם של תלמידים שנבחנו בשתי סוללות מבחנים כאשר עשרה חודשים מפרידים בין בחינה לבחינה. שתי סוללות הבחינה היו דומות אך לא זהות; שניים מן המבחנים בסוללה השניה היו מבחנים חדשים שכלל לא הופיעו בסוללה הראשונה, שניים אחרים היו נוסחים מקבילים למבחנים מן הסוללה הראשונה, ושניים נוספים היו מבחנים זהים לאלה שהופיעו בסוללה הראשונה.

תוצאות המחקר מוצגות בטבלה 5.1. מן הטבלה עולה בבירור כי אפקט השיפור הולך וגדל עם הגידול בדמיון בין המבחן הראשון והשני.

טבלה 5.1: אפקט אימון כללי וספציפי

הציונים באוכלוסית הבסיס הם על סולם אחיד שסטיית התקן שלו היא 3.0.

סוג החשוואה	אפקט שיפור ממוצע	סטית תקן של השיפור	מתאם הבחינה החוזרת
מבחנים חדשים: "ספירת קוביות" ו"הבנת קטע" מול סוללת המבחנים הראשונה כולה	.03	1.38	.68
נוסחים מקבילים: מבחן אוצר מילים ומבחן ידע כללי בנושאי תברה	.41	1.75	.67
נוסחים זהים: מבחן ידע כללי ומבחן מטריצות	.84	1.68	.66

מתוך נתונים אלו, ע"י הפחתת השורות זו מזו, ניתן לאמוד את האפקט של סוגי האימון השונים. האימון הכללי מביא לשיפור מזערי בלבד של 0.03. נקודות ואינו מובהק סטטיסטית; האימון הספציפי למבחן מביא לשיפור של $0.38 = 0.03 - 0.41$. נקודות, ואילו האימון הספציפי לפריטים מביא לשיפור של $0.43 = 0.41 - 0.84$ נקודות. אמדן השיפור על נוסחים מקבילים בשיעור של 0.41 הוא כחמישית מסטיית התקן של ציוני הנבחנים החוזרים ומבחינה זו, תואם את הנתונים על אפקט השיפור במבחני ה-SAT. לא ניתן ללמוד ממחקר זה על התלות של אפקט השיפור בסוג המבחן עצמו, זאת מאחר שאפקטים כאלה (אם הם קיימים כלל) מעורבים בתוך מערך המחקר בצורה שקשה להפרידם מן האפקטים של כלליות האימון.

על אפקט שיפור דיפרנציאלי בתחומי בחינה שונים ניתן ללמוד ממחקר שנעשה על בחינות ה-GRE (Graduate Record Examinations) הנערכות גם הן ע"י ה-ETS (Kingston & Turner, 1984).

בחינת ה-GRE כוללת שלושה מבחנים: מבחן מילולי - V, מבחן חשיבה כמותית - Q ומבחן חשיבה אנליטית - A. Kingston & Turner בדקו את ציוניהם של אלפי נבחנים שחזרו על בחינת ה-GRE בשנים 1980-1982 במרווחי זמן שונים. בשנים אלו הוכנסו למבחן ה-GRE שינויים מבניים שונים, כך שלמעשה, בתקופה מסוימת היו שני מבחני GRE: "חדש" ו-"ישן". עקב כך היתה אפשרות לבדוק אם שיפור הציון מושפע משינויים מבניים בבחינה. תוצאות החשוואה מובאות בטבלה 5.2; השיפור מבוטא במונחי סולמות הציונים של מבחני ה-GRE הדומים לסולמות ציוני ה-SAT (טווח 200-800, ממוצע - 500 בקירוב, סטיית-תקן 100 בקירוב).

טבלה 5.2: השיפור בבחינה חוזרת של ה-GRE

מבחן	טיפוס הבחינה				מובהקות החפרש בין חדש וישן
	"ישן" (n = 5072) ממוצע ס.ת.		"חדש" (n = 3475) ממוצע ס.ת.		
V	22.2	56.2	26.6	58.4	.00
Q	26.2	65.5	29.5	69.1	.12
A.	41.0	70.0	22.0	82.5	.00

בנתונים אלו יש כדי לאשש ממצאים לאמור, שאפקט החזרה הוא בסדר גודל של רבע מסטיית התקן. מנתונים אלו עולה גם שבשניים (A,V) מתוך שלושת הסולמות, אפקט המבחן החוזר מושפע משינויים במתכונת הבחינה. כדאי לסקור את השינויים במתכונת הבחינה; לפחות אחד מהם רלבנטי להבנת התוצאות בלוח 5.2.

בבחינה הישנה הוכנסו שלושה שינויים:

- א. הזמן המוקצב למבחנים שונה - על-מנת להפחית מגורם הזמן (Speededness) במבחן הוורבלי ובמבחן החשיבה האנליטית, ועל-מנת להגדיל את מספר הפריטים במבחן החשיבה הכמותית. התוצאות מרמזות כי מבחן מהיר (Speeded) עמיד יותר בפני שיפור בבחינה חוזרת.
- ב. במבחן החשיבה האנליטית הוכנס שינוי מהותי באמצעות השמטת שני סוגי פריטים שבהם נתגלה אפקט אימון קצר טווח (בתוך המבחן עצמו) ואפקט אימון ארוך טווח; כלומר, ניתן היה, על-ידי הדרכה מתאימה, ללמוד אסטרטגיות לפתרונם. התוצאות מראות שהשינוי השיג את מטרתו: במבחן החשיבה האנליטית "החדש" הצטמצם אפקט השיפור במידה ניכרת.
- ג. שיטת הצייון של המבחנים שונתה כולה; מצייון עפ"י נוסחה בה מתקנים עבור ניהוש, לצייון עפ"י מספר התשובות הנכונות. ההוראות לנבחנים בדבר כדאיות הניהוש שונו בהתאם.

ממצאים נוספים במחקר זה תומכים בממצאים על מבחני SAT: ראשית, ככל שהזמן בין הבחנות מתארך, כן גדל שיפור הצייון בבחינה החוזרת. שנית, ככל שגדל הדמיון בין שני מבחנים, כן גדל שיפור הצייון. בדיקה זו התאפשרה משום העובדה שחלק מן הנבחנים חזרו על אותו נוסח בחינה. פירושו של דבר - שיש השפעה לספציפיות האימון. הנתונים אף מראים שספציפיות האימון שונה ממבחן למבחן - היא קטנה ביותר במבחן הוורבלי וגדולה יותר במבחן החשיבה האנליטית. הנתונים מוצגים בטבלה 5.3 ומהם אפשר ללמוד כי יש סימנים לאינטראקציה בין שני הגורמים; ככל שספציפיות האימון גדלה כן פוחת האפקט שיש למרווח הזמן בין הבחנות על השיפור בצייון.

טבלה 5.3: ממוצעי השיפור בבחינה חוזרת במבחני GRE.

אפקט מרווח הזמן, סוג המבחן והדמיון בין המבחן הראשון

לשני

סוג המבחן			n		הזמן עד לבחינה חוזרת
A	Q	V			
47	36	36	253	נוסח זהה	2 חודשים
17	16	23	498	נוסח מקביל	
30	20	13		הפרש	
49	46	40	251	נוסח זהה	4 חודשים
29	31	31	539	נוסח מקביל	
20	15	9		הפרש	

6. סיכום

הממצא הבולט ביותר בכל המחקרים שסוכמו לעיל הוא, כי במבחני כושר סכולסטי לסוגיהם קיים אפקט שיפור ממוצע מובהק בבחינות חוזרות. חלקו של האפקט הוא שיפור אמיתי הנובע כנראה מאימון, תרגול, ורכישת ידע. עם זאת - חלק מן השיפור הנצפה ניתן לייחוס לאפקט הבחירה העצמית של הנבחנים, אשר בשילוב עם תופעת הרגרסיה אל הממוצע, מביא להערכת יתר של אפקט השיפור.

בנתונים ישנם רמזים לכך שאפקט שיפור הציון עולה במעט עם חלוף הזמן בין הבחינות.

כצפוי, אפקט השיפור שונה ממבחן למבחן, והוא תלוי במידה רבה בדמיון שבין המבחן הראשון לבין המבחן החוזר. ככל שהמבחנים דומים יותר - כך גדל אפקט השיפור. כאמור, חלקו הבא של הפרק יוקדש לבדיקה אמפירית של שאלות אלו (ואחרות) תוך בהסתמך על הציונים בבחינות הפסיכומטריות של המרכז הארצי לבחינות.

מאז דצמבר 1983, עורך המרכז הארצי לבחינות את בחינת הכניסה הפסיכומטרית למועמדי כל האוניברסיטאות בישראל (הטכניון הצטרף בתחילת 1986). בנוסף על המועמדים לאוניברסיטאות, נבחנים בבחינות של המרכז הארצי מועמדים למכינות הקדם-אקדמיות ומועמדים למכללות שונות.

מדצמבר 1983 ועד דצמבר 1986 נבחנו על ידי המרכז הארצי 97,715 נבחנים. הנסיון מלמד כי נבחנים רבים נבחנים בבחינה הפסיכומטרית פעמיים ואף יותר, בנסיון לשפר את ציונם על מנת להתקבל לאוניברסיטה בכלל או לחוג מועדף בפרט.

מטרתו של מחקר זה היא לבדוק שאלות שונות הקשורות בשיפור הציון המושג בין הבחינה הפסיכומטרית הראשונה לשניה. המחקר מבוסס על מאגר הנתונים הקיים במרכז הארצי לבחינות ולהערכה, היינו, ההישגים בבחינה הפסיכומטרית על מרכיביה ונתוניו האישיים (ביוגרפיים) של הנבחן, כפי שהם מופיעים בשאלון הסטנדרטי לנבחן.

בפועל, ידון מחקר זה בנושאים הבאים:

1. התפלגות שכחות של נבחנים חוזרים לפי שנתוני בחינה.
2. התפלגות ציון השיפור (ממוצע וסטיית תקן) בין בחינה ראשונה ושניה, תוך הבחנה בין השיפור בציון הפסיכומטרי הסופי ובמבחנים השונים המרכיבים את הבחינה הפסיכומטרית.
3. התפלגות הציונים בבחינה הפסיכומטרית, הראשונה, השניה והשלישית של נבחנים חוזרים.
4. מהימנות מבחן חוזר, היינו, מתאמים בין בחינות חוזרות.
5. הקשר בין ציון השיפור לבין משתנים שונים (הישגים בבחינה הראשונה, פער הזמן בין בחינות עוקבות, גיל ושיפור ברמת השכלה).
6. הבדלים בציון השיפור בין נבחנים בשפות שונות והקשר בין שינוי שפת הבחינה לבין גובה ציון השיפור.
7. הבדלים בין המינים בגובה ציון השיפור.

1. מידגם

מחקר זה מקיף את כל הנבחרים אשר נבחנו בבחינה הפסיכומטרית פעמיים או יותר בין דצמבר 1983 לדצמבר 1986. סך כל הנבחרים בתקופה זו הוא 97,715 איש. מתוכם נבחנו 79,882 איש בשפה העברית. פיזור הנבחרים החוזרים (כל הנבחרים) הוא כדלהלן:

10,656 איש נבחנו לפחות פעמיים בבחינה הפסיכומטרית.

720 איש נבחנו לפחות שלוש פעמים בבחינה הפסיכומטרית.

22 איש נבחנו ארבע פעמים בבחינה הפסיכומטרית.

התוצאות שיובאו להלן מתייחסות לשתי קבוצות הנבחרים הראשונות:

נבחרים "כפולים" - נבחרים אשר נבחנו לפחות פעמיים בבחינה הפסיכומטרית.

נבחרים "משולשים" - נבחרים אשר נבחנו לפחות שלוש פעמים בבחינה הפסיכומטרית.

היות ומרבית הנבחרים נבחנו בשפה העברית ומספר הנבחרים בשפות השונות הוא קטן יחסית, נעשתה הבחנה בין נבחני עברית לבין נבחנים שנבחנו לפחות בחלק מהבחינות בשפות זרות. מרבית העיבודים נעשו על נבחני עברית בלבד.

2. שיטה

הנבחרים החוזרים אותרו מתוך קובץ הכולל את כל הנבחרים שנבחנו בבחינה הפסיכומטרית במרכז הארצי לבחינות בתקופה שפורטה לעיל. תהליך זיהוי הנבחרים החוזרים נעשה בשלב ראשון בעזרת התאמת מספר תעודת הזהות. בשלב השני נבדקה זהות שמית. בדיקת הזהות השמית נעשתה באופן ידני; בכל מקרה שנתגלתה אי-התאמה כלשהי בין השמות, הופעל שיקול דעת על מנת להחליט האם שני השמות שונים לחלוטין או האם קיים דמיון מספיק על מנת להניח כי שני השמות שייכים לאותו אדם. בכל מקרה של ספק בהתאמה השמית, הושמטו הנבחרים מקובץ הנבחרים החוזרים.

בשלב השלישי אותרו נבחרים אשר עשו בחינה חלקית בלבד ונבחרים אשר מסיבות שונות לא סיימו את הבחינה; נבחרים אלה הושמטו מקובץ.

מתוך תהליך המיון הנ"ל הוכנו שני קבצים:

הקובץ הראשון כלל את כל הנבחרים שנבחנו בבחינה הפסיכומטרית לפחות פעמיים. עבור נבחרים שנבחנו יותר מפעמיים, כולל הקובץ את שתי הבחינות הראשונות. הקובץ השני כלל את כל הנבחרים שנבחנו שלוש פעמים או יותר. במקרה זה, נכללו בקובץ שלוש הבחינות הראשונות של הנבחרים שנבחנו ארבע פעמים.

בטרם נדון בממצאים, יש להזכיר מספר עובדות המתייחסות למדיניות המרכז הארצי לבחינות ומדיניות האוניברסיאות, אשר עשויות להיות בעלות השלכות על תוצאות מתקר זה:

1. אין נבחן רשאי לגשת פעם נוספת לבחינה הפסיכומטרית תוך פרק זמן הקטן מ-10 חודשים מאז נבחן בפעם האחרונה, אלא אם קיבל אישור מיוחד לכך.
2. לכל נבחן הנבחן בפעם השנייה מחושב ציון ההפרש בין שתי הבחינות; היה וציון זה עולה על 150 נקודות CG, מוזמן הנבחן לבחינה שלישית האמורה לקבוע איזה משני הציונים אמין יותר, נוהל זה הופעל לראשונה בדצמבר 1984. החל מתאריך זה ועד יוני 1986, שולבו תוצאות הבחינה השלישית של הנבחנים הנ"ל בקובץ הנתונים הכללי, כלומר, חלקם של הנבחנים ה"משולשים" הם למעשה נבחנים "כפויים". החל מיוני 1986 הוחלט לשמור את תוצאות הבחינה בקובץ נפרד.
3. ציוני הבחינה הפסיכומטרית תקפים לפרק זמן של 5 שנים. משמעות הדבר הוא, כי רק החל מדצמבר 1988 ניתן יהיה למצוא בקרב הנבחנים החוזרים נבחנים שניגשו לבחינה בפעם השנייה היות ותוצאות הבחינה הראשונה אינן תקפות יותר.

3. תוצאות

3.1 פיזור ושכיחות נבחנים חוזרים לפי שנתוני בחינה

אחוז הנבחנים הניגשים לבחינה בפעם השנייה בשנתון בחינה נתון נע בין 13% ל-17.6% (ראה טבלה מס. 3); אחוז הנבחנים הניגשים לבחינה בפעם השלישית בשנתון נתון נע בין אחוז אחד ל-2% בקירוב (ראה טבלה מס. 4). עם זאת, יש לזכור כי מתקר זה מקיף פרק זמן בן שלוש שנים בלבד; יתכן מאוד כי במחזוריים הראשונים היו נבחנים חוזרים אשר המרכז הארצי לבחינות אינו יכול לזהותם; כמו כן, תיתכן עליה במספר הנבחנים בתום 5 שנים מאז הועברה הבחינה הראשונה (דצמבר 1988). עובדות אלה מצביעות על האפשרות, שאחוז הנבחנים החוזרים בשנתון הוא למעשה גבוה מעט מהנצפה לעיל, אם כי לא במידה ניכרת. עיון בטבלה מס. 5 מצביע על כך שהשכיחות היחסית הגבוהה ביותר של נבחנים חוזרים על הבחינה הפסיכומטרית תוך 9-14 חודש (כ-71%). 95.3% מהנבחנים החוזרים נבחנים בפעם השנייה בבחינה הפסיכומטרית תוך שנתיים; ממוצע פער הזמן הממוצע בין בחינה ראשונה ושנייה הוא 13.8 חודשים. מנתונים אלה ניתן להניח כי אחוז הנבחנים החוזרים בכל שנתון בחינה לא יעלה באופן ניכר, אף אם יגדל טווח הזמן של המתקר. כמו כן, ניתן להניח כי השנתון המייצג ביותר את אחוז הנבחנים החוזרים הוא שנתון 1986 לנבחנים "כפולים", בעוד שעל מנת להעריך נכונה את השכיחות היחסית של הנבחנים ה"משולשים" בכל שנתון, יש להמתין לפחות עוד שנתיים נוספות.

פיזור נבחנים "כפולים" לפי שנתון הבחינה -
 כל הנבחנים ונבחני עברית בלבד (בסוגריים)

טבלה מס. 1

בחיינה ראשונה	בחיינה שנייה	1984	1985	1986	1987*	סה"כ
1984		59 (50)	4109 (2956)	1620 (1274)	75 (74)	5863 (4354)
	1985		184 (38)	4265 (3383)	71 (68)	4520 (3489)
	1986			164 (148)	109 (106)	273 (254)
	סה"כ	59 (50)	4293 (3003)	6049 (4804)	255 (248)	10656 (8097)

* נבחני 1987 כוללים את הנבחנים אשר נבחנו בדצמבר 1986 בלבד.

פיזור נבחנים "משולשים" לפי שנתוני הבחינה
 כל הנבחנים ונבחני עברית בלבד

טבלה מס. 2

בחיינה ראשונה	1984			1985		סה"כ
	1984	1985	1986	1985	1986	
בחיינה שנייה						
בחיינה שלישית						
1985	8 (5)	28 (4)	-	-	-	36 (9)
1986	4 (2)	570 (316)	13 (10)	49 (1)	28 (20)	664 (349)
1987	-	10 (10)	4 (4)	-	6 (6)	20 (20)
סה"כ בחיינה שנייה	12 (7)	608 (330)	17 (14)	49 (1)	34 (26)	720 (378)

שכיחות ואחוז נבחנים "כפולים" מתוך
 סה"כ הנבחנים בכל שנתוני הבחינה
 (כל הנבחנים ונבחני עברית בלבד)

	1984	1985	1986	1987*
סה"כ הנבחנים בשנתון	29,811	31,633	34,316	1,955
סה"כ נבחני עברית בשנתון	24,135	25,897	27,895	1,955
שכיחות נבחנים "כפולים" (כל הנבחנים)	59 -	4,293 (13.6%)	6,049 (17.6%)	255 (13.0%)
שכיחות נבחנים "כפולים" (נבחני עברית בלבד)	50	3,003	4,804	255
אחוז הנבחנים בעברית מתוך סה"כ הנבחנים	-	(9.5%)	(14.0%)	(13.0%)
אחוז הנבחנים בעברית מתוך סה"כ נבחני עברית	-	(11.6%)	(17.2%)	(13.0%)

* נבחני 1987 כוללים רק את הנבחנים אשר נבחנו בדצמבר 1986 .

שכיחות ואחוז נבחנים "משולשים" מתוך
סה"כ הנבחנים בכל שנתוני הבחינה
 (כל הנבחנים ונבחני עברית בלבד)

	1984	1985	1986	1987*
סה"כ נבחני עברית בשנתון	24,135	25,897	27,895	1,955
שכיחות נבחנים "משולשים" (כל הנבחנים)	-	36	664 (1.93%)	20 (1.02%)
שכיחות נבחנים "משולשים" (נבחני עברית בלבד)	-	9	349 (1.02%)	20 (1.02%)
אחוז הנבחנים בעברית מתוך סה"כ הנבחנים			(1.02%)	(1.02%)
אחוז הנבחנים בעברית מתוך סה"כ נבחני עברית			(1.25%)	(1.02%)

* נבחני 1987 כוללים רק את הנבחנים אשר נבחנו בדצמבר 1986.

פער הזמן (בחודשים) בין
בחינה פסיכומטרית ראשונה לשנייה

פער זמן (חודשים)	שכיחות	שכיחות יחסית
1-2	128	1.2
3-4	38	0.4
5-6	193	1.8
7-8	385	3.6
9-10	1719	16.1
11-12	4253	39.9
13-14	1566	14.7
15-16	517	4.8
17-18	220	2.1
19-20	220	2.1
21-22	341	3.2
23-24	575	5.4
25-26	277	2.6
27-28	107	1.0
29-30	42	0.4
31-32	51	0.5
33-34	19	0.2
35-36	5	0.04

3.2 ציון השיפור בין בחינות חוזרות

טרם נדון בתוצאות המתייחסות לציון השיפור, נקדים ונציג את מבנה הבחינה הפסיכומטרית ואת סולמות הציונים הנהוגים. הבחינה הפסיכומטרית מורכבת מ-5 מבחנים: ידע כללי, צורות, הבנה, מתמטיקה ואנגלית. הציון הסופי בבחינה (CG) מורכב משיקלול שווה של הציונים בחמשת המבחנים. הציונים מדווחים בציוני תקן לפי ההתפלגות הבאה:

הציון הסופי בבחינה הפסיכומטרית: ממוצע = 500 וסטיית התקן = 100

הציונים במבחנים השונים: ממוצע = 100 וסטיית התקן = 20

נתונים אלה עשויים להקל על הקורא את ההתייחסות לציון השיפור.

כל הממצאים אשר יוצגו בפרק זה מתייחסים לנבחני עברית בלבד, היינו, נבחנים אשר גבחנו בכל הבחינות בשפה העברית. ממצאים המתייחסים לנבחני שפות זרות יובאו בהמשך הדו"ח.

בבדיקה ראשונה של התפלגות ציון השיפור (ההפרש) לנבחנים "כפולים", נמצא ממוצע

$M=32.54$ וסטיית התקן $s=40.06$. כמו כן, נמצא מספר רב של ציוני שיפור החורגים

באופן בולט ממוצע האוכלוסיה. הציונים הקיצוניים בהתפלגות ציון השיפור הגיעו עד

כדי סטייה של $+9.8$ סטיות תקן לציון הפרש חיובי (שהן 425 נק') ו- -6.2 סטיות תקן

לציון הפרש שלילי (שהן -280 נק'). הסיכוי שציונים חריגים אלה אכן שייכים להתפלגות

ציוני ההפרש הוא מזערי. קיים חשש כי ציוני הפרש חריגים הם תוצאה של מעשי רמיה באחת

הבחינות, או, להבדיל, ביצוע כושל עקב מחלה. לאור הנחה זו, הוחלט לקבוע נקודת תוך,

אשר ציוני הפרש החורגים ממנה יוגדרו כציונים חריגים ויושמטו מהעיבודים. נקודת תוך

זו, נקבעה ל- ± 3.5 סטיות תקן סביב ממוצע ציון ההפרש. קריטריון זה תחם את ציוני

ההפרש "לגיטימיים" בין $+173$ ל- -108 נקודות וגרם להשמטה של 74 נבחנים מתוך $10,656$

הנבחנים ה"כפולים", ו- 31 נבחנים מתוך 720 הנבחנים ה"משולשים". כל העיבודים

המתייחסים לציוני הפרש בוצעו על נבחנים בעלי ציונים "לגיטימיים" בלבד.

בבדיקת ציון השיפור הממוצע בין בחינות חוזרות, נמצא כי קיים שיפור ממוצע בציון

הפסיכומטרי הסופי ובכל המבחנים בין בחינה ראשונה לשנייה. ציון השיפור נע בין 5.96

ל- 3.81 נקודות למבחנים ומגיע ל- 32.49 נקודות בציון הפסיכומטרי הסופי, שהן כשליש

סטיות תקן בהתפלגות הציון הפסיכומטרי.

השיפור הגדול ביותר מושג במבחן מתמטיקה (5.96 נק') והקטן ביותר במבחן ידע כללי

(3.81 נק'). קיים שיפור בציון הפסיכומטרי גם בין בחינה שניה ושלישית. עם זאת,

שיפור זה קטן מהשיפור בין בחינה ראשונה ושנייה. השיפור הממוצע בציונים בין

בחינה שניה ושלישית נע בין 1.49 נק' (למבחן הבנה) ל- 5.05 נק' (למבחן מתמטיקה

וצורות); השיפור לציון הפסיכומטרי הסופי הוא 23.82 נק'.

סה"כ השיפור הממוצע בציון הפסיכומטרי בין בחינה ראשונה ושלישית הוא 49.44 נק', שהן

כמחצית סטיית התקן של התפלגות הציון הפסיכומטרי.

ממוצע הציונים בבחינה הראשונה בקרב נבחנים חוזרים נמוך במידה ניכרת מהממוצע באוכלוסיה (ראה טבלה מס. 7). ממוצע זה נמוך במיוחד בקרב נבחנים החוזרים על הבחינה שלוש פעמים. נבחנים אלה מתקשים להגיע לממוצע האוכלוסיה ($\mu = 500$) אפילו בבחינה השלישית. פיזור הציונים של הנבחנים החוזרים בכל הבחינות הוא קטן מהפיזור באוכלוסיה ($\sigma = 100$) והוא נע בין 78.9 ל-84.1.

קיים קשר שלילי מובהק בין ציון השיפור לבין הציון בבחינה הראשונה. $r = -0.195$, $P < .01$, ככל שהציון בבחינה הראשונה נמוך יותר, כן גדל השיפור בציון הפסיכומטרי בין בחינה ראשונה לשניה. מעל ציון 700 השיפור הממוצע הופך להיות שלילי (ראה טבלה מס. 8). המתאמים בין ציון השיפור במבחן ספציפי לבין הציון באותו מבחן בבחינה הראשונה גבוהים אף יותר מהמתאם בין ציון השיפור לבין הציון בבחינה הראשונה, והם נעים בין $r = -0.265$ למבחן אנגלית עד $r = -0.432$ למבחן צורות, היינו, מבחן צורות הוא המבחן הקשה ביותר לשיפור, בתנאי שהציון שהושג בו בבחינה הראשונה הוא גבוה יחסית (ראה טבלה מס. 9).

ממצאים אלה מורים כי הסיכויים לשיפור הציון בבחינה השניה הולכים וקטנים ככל שהציון בבחינה הראשונה גבוה יותר. תופעה זו באה לידי ביטוי גם במתאם השלילי המתקבל בין ציון השיפור לבין הציון בבחינה הראשונה $r = -0.195$ (ראה טבלה מס. 10), ברם, אין להסיק מכך מסקנות מרחיקות לכת, שכן המתאם בין ציון השיפור לבין הציון בבחינה הראשונה אינו משקף יותר מאשר את אפקט הרגרסיה אל הממוצע. למעשה, גודלו של מתאם זה תלוי אך ורק במהימנות "מבחן חוזר" (המתאם בין הבחינה הראשונה והשניה) וביחס השונויות של ציוני שתי הבחינות $\sigma_{x_1}/\sigma_{x_2}$; מתאם זה ילך ויגדל בכיוון השלילי ככל שהמהימנות תהיה גבוהה יותר ויחס השונויות גדול יותר.

כאשר מהימנות המבחן קטנה מ-1.00, ניתן להניח כי קיימת טעות מדידה. טעות זו תהיה קרוב לוודאי שלילית (הערכת חסר) לגבי נבחנים בעלי ציונים נמוכים וחיובית (הערכת יתר) לגבי נבחנים בעלי ציונים גבוהים; מכאן, שנבחנים בעלי ציונים נמוכים הם בעלי סיכוי טוב יותר לשפר את ציונם בבחינה חוזרת מאשר נבחנים בעלי ציונים גבוהים, וזהו בדיוק הסיכוי הנצפה לעיל.

לתופעה זו של רגרסיה לממוצע יש השלכות גם לגבי אומדן השיפור הממוצע באוכלוסיה כולה. היות וממוצע הציונים בבחינה הראשונה של הנבחנים החוזרים נמוך מממוצע הציונים באוכלוסיה כולה, ניתן להניח כי לו ניגשו כל הנבחנים לבחינה חוזרת, היה מתקבל ממוצע שיפור קטן מהממוצע הנצפה לעיל ($\mu = 32.49$). שאלה זו אכן נבדקה ונמצא כי אם כל הנבחנים בשנתון היו ניגשים לבחינה חוזרת, ממוצע ציון השיפור היה יורד ל-29.85 נקודות.

ממוצעים וסטיות תקן של ציון השיפור
 בין בחינות חוזרות
 (נבחני עברית בלבד)

ציון פסיכומטרי	ממוצע ס.ת.	נבחנים "כפולים" (n = 8071)		נבחנים "משולשים" (n = 374)	
		בחינה ראשונה ושנייה	בחינה ראשונה ושנייה	בחינה שנייה ושלישית	בחינה ראשונה ושלישית
		32.49 38.16	25.62 40.57	23.82 38.68	49.44 43.22
ידע כללי	ממוצע ס.ת.	3.81 9.88	4.64 15.36	2.45 10.29	6.21 10.45
צורות	ממוצע ס.ת.	5.87 14.82	4.64 15.36	5.05 14.84	9.69 16.61
הבנה	ממוצע ס.ת.	4.39 13.65	4.95 14.91	1.49 12.44	6.44 13.90
מתמטיקה	ממוצע ס.ת.	5.96 11.79	4.15 12.18	5.05 11.70	9.19 12.85
אנגלית	ממוצע ס.ת.	5.78 10.82	2.83 10.76	4.90 10.55	7.73 10.95

ממוצע וסטיות תקן של הציון הפסיכומטרי
 בבחינה הראשונה השנייה והשלישית
 (נבחני עברית בלבד)

	נבחנים "כפולים" (n = 8071)		נבחנים "משולשים" (n = 374)		
	בחינה ראשונה	בחינה שנייה	בחינה ראשונה	בחינה שנייה	בחינה שלישית
ציון פסיכומטרי	478.8 82.8	511.3 84.1	449.2 78.9	474.8 81.1	498.6 80.7
ידע כללי	95.4 17.3	99.2 17.1	90.3 16.3	94.0 16.5	96.8 16.9
צורות	98.5 17.8	104.4 17.6	94.1 17.6	98.7 17.4	103.8 16.0
הבנה	97.7 17.4	102.1 17.3	91.8 16.3	96.8 16.6	98.2 16.0
מתמטיקה	97.4 17.8	103.4 18.3	93.5 17.7	97.6 17.2	102.7 17.4
אנגלית	94.2 17.4	100.0 17.8	90.1 15.2	92.9 17.1	97.8 16.9

טבלה מס. 8 פיזור ציון השיפור (הפרש) בין הבחינה הפסיכומטרית ראשונה ושניה לפי קטגוריות של השגים בבחינה הפסיכומטרית הראשונה (CG1), (נבחני עברית בלבד)

קטגוריות CG1	N	ממוצע ציון השיפור	ס.ת. ממוצעת של ציון השיפור	פער זמן ממוצע (בחודשים)
200 - 274	15	57.9	42.4	12.9
275 - 299	58	49.4	43.5	13.8
300 - 324	120	43.9	39.3	14.4
325 - 349	246	42.4	38.2	13.0
350 - 374	411	41.7	40.5	13.1
375 - 399	567	38.6	38.8	13.8
400 - 424	682	35.7	38.3	13.5
425 - 449	794	37.0	37.9	13.4
450 - 474	900	37.9	38.8	13.6
475 - 499	804	33.8	37.6	13.8
500 - 524	799	30.8	38.9	13.5
525 - 549	747	30.6	36.1	13.8
550 - 574	598	25.7	34.4	14.0
575 - 599	450	20.3	34.2	14.2
600 - 624	300	15.2	33.2	14.3
625 - 649	164	14.1	29.7	15.9
650 - 674	72	13.2	29.2	15.7
675 - 699	33	11.3	23.9	14.8
700 - 774	17	-9.5	24.7	19.0

מתאמים בין ציון השיפור בין בחינה ראשונה ושניה
 לבין הציון בבחינה הפסיכומטרית הראשונה*
 (המתאמים חושבו לציון הסופי ובתוך כל מבחן בנפרד)

	ציון פסיכומטרי	ידע כללי	צורות	הבנה	מתמטיקה	אנגלית
ציון השיפור במבחן	-.195	-.310	-.432	-.399	-.293	-.265

* המתאם בתוך כל מבחן חושב בין הציון ב'מבחן' הספציפי בבחינה הפסיכומטרית הראשונה לבין ציון השיפור באותו 'מבחן' בין הבחינה הפסיכומטרית הראשונה לשניה.

.001 < p לכל המתאמים.

3.3 מהימנות מבחן חוזר

את הדרכים המקובלות לבדיקת מהימנות מבחן היא בדיקת המתאם בין הציון בבחינה הראשונה לבין ציון בבחינה חוזרת. בדיקה זו אכן בוצעה כאן ונמצא כי מהימנות המבחן החוזר לציון הפסיכומטרי הסופי היא $r = .896$ בין בחינה ראשונה לשניה, $r = .885$ בין בחינה שניה לשלישית ו- $r = .854$ בין בחינה ראשונה לשלישית (ראה טבלה מס. 10). מהימנות מבחן חוזר למבחנים השונים נעה בין $.835$ ל- $.651$ בחינה ראשונה ושניה והיא נמוכה מעט בחינה ראשונה ושניה ושלישית וראשונה ושלישית. המבחן בעל מהימנות החזרה הגבוהה ביותר, באופן עקבי, הוא מבחן ידע כללי, בעוד שמהימנות מבחן חוזר הנמוכה ביותר שייכת למבחן צורות. בבדיקת הקשר בין הפרש הזמן בין בחינות חוזרות לבין ציון השיפור (ראה טבלה מס. 12), נמצא לרוב קשר בלתי מובהק, כאשר המגמה הכללית של המתאמים המובהקים המועטים מצביעה על קשר שלילי, היינו, הזמן פועל לרעתו של הנבחן: ככל שעובר זמן רב יותר בין הבחינות, ציון השיפור נמוך יותר. הדבר נכון בפרט במעבר בין בחינה ראשונה לשלישית.

טבלה מס. 10 מתאמים בין בחינות חוזרות (מהימנות מבחן חוזר)*

מתאמים בין בחינות חוזרות	נבחנים "כפולים" (n = 8071)	נבחנים "משולשים" (n = 375)		
	בחינה ראשונה ושניה	ראשונה ושניה	בחינה שניה ושלישית	ראשונה ושלישית
ציון פסיכומטרי	.896	.872	.885	.854
ידע כללי	.835	.817	.814	.803
צורות	.651	.619	.639	.553
הבנה	.690	.590	.710	.629
מתמטיקה	.787	.757	.771	.731
אנגלית	.811	.784	.808	.773

* כל המתאמים מובהקים $p < 0.001$.

טבלה מס. 11 ממוצעים וסטיות תקן של הפרש הזמן בין בחינות חוזרות (נבחני עברית בלבד)

		נבחנים "כפולים" (n = 8071)	נבחנים "משולשים" (n = 374)		
		בחינה ראשונה ושנייה	ראשונה ושנייה	בחינה שנייה ושלישית	ראשונה ושלישית
הפרש זמן (חדשים)	ממוצע ס.ת.	13.79 5.19	11.86 2.69	11.57 3.54	23.44 4.13

טבלה מס. 12 מתאמים בין הפרש זמן בין בחינות חוזרות לבין ציון השיפור (נבחני עברית בלבד)

מתאמים בין בחינות חוזרות	נבחנים "כפולים" (n = 8071)	נבחנים "משולשים" (n = 374)		
	בחינה ראשונה ושנייה	ראשונה ושנייה	בחינה שנייה ושלישית	ראשונה ושלישית
הפרש זמן - שיפור בציון הפסיכומטרי (CG)	-.019	-.044	-.078	-.104*
הפרש זמן - שיפור ידע כללי	-.031**	.040	-.007	-.007
הפרש זמן - שיפור במבחן צורות	-.028*	-.017	-.071	-.102*
הפרש זמן - שיפור הבנה	-.015	-.073	-.112*	-.084
הפרש זמן - שיפור במבחן מתמטיקה	-.009	-.068	-.033	-.106*
הפרש זמן - שיפור במבחן אנגלית	.015	.031	.052	.066

p < 0.05 *
p < 0.01 **

מרבית הנבחנים החוזרים על הבחינה הפסיכומטרית שנית, נבחנים בבחינה השניה באותה שפה בה נבחנו פעם ראשונה. רק 315 נבחנים מתוך 10,656 נבחנים "כפולים" החליפו את שפת הבחינה, מתוכם 136 (43.2%) החליפו את שפת הבחינה מעברית לשפה אחרת ו- 116 (36.8%) נבחנים החליפו את שפת הבחינה משפה אחרת לעברית. מתוך כלל הנבדקים שהחליפו שפת בחינה ($N = 315$), חילופי השפות הבולטים ביותר הם בין ערבית ועברית (46.7%) ($N = 147$) ובין נוסח משולב ושפות אחרות כולל עברית (34.6%) ($N = 109$). 47 מתוך 720 הנבחנים "המשולשים" החליפו את שפת הבחינה בין שלושת הבחינות (ראה טבלה 13).

האחוז הגבוה ביותר של נבחנים חוזרים ("כפולים") מבין נבחני השפות השונות (יחסית לשכיחותם באוכלוסית הנבחנים כולה), שייך לנבחנים בשפה הערבית 18.6% (אחוז זה מתייחס רק לנבחנים שנבחנו בשפה הערבית בשתי הבחינות). אחוז הנבחנים ה"כפולים" בקרב נבחני עברית הוא 10.1%. אחוז הנבחנים ה"כפולים" בקרב נבחני שפות אחרות נמוך במידה ניכרת ונע בין 6.1% - ל-3.7%. גם בקרב הנבחנים ה"משולשים", אחוז הנבחנים בשפה הערבית הוא הגבוה ביותר - 2.8%, והוא גדול פי שש לערך, מהיחס המקביל בקרב נבחני עברית (ראה טבלה מס. 14).

בהשוואת ציון השיפור לנבחני שפות שונות ניתן להבחין במספר מגמות (ראה טבלה מס. 15). השיפור הממוצע הגבוה ביותר הושג בקבוצת הנבחנים שנבחנו בבחינה הראשונה בעברית ובבחינה השניה בשפה זרה (לרוב שפת אם). זאת להוציא את הנבחנים שעברו מעברית לספרדית ($N = 5$). שיפור ממוצע הושג בכל קבוצות הנבחנים שנבחנו באותה שפה בשתי הבחינות, כאשר השיפור הממוצע נע בין 11.4 נק' (ערבית) ל-38.1 נק' (ספרדית).

בקבוצת הנבחנים שעברו משפה זרה לעברית אין מגמת שיפור אחידה. כמו כן, מרבית המידגמים קטנים מכדי להסיק מסקנות. אחת התופעות הבולטות יותר, היא היתרון שיש לנבחני ערבית כאשר הם נבחנים בשפתם. ציון השיפור הממוצע בתוך השפה הוא 11.4 נק'. המעבר מבחינה בשפה העברית לבחינה בשפה הערבית מקנה לנבחנים שיפור ממוצע של 56.8 נק', בעוד שמעבר הפוך בין השפות מקנה להם ציון שיפור שלילי של -13.4 נק'.

טבלה מס. 13 פיזור נבחנים "כפולים" בבחינה ראשונה ושניה לפי שפת הבחינה**
 (כל הנבחנים N = 10656).

משולב	רוסית	ערבית	ספרדית	צרפתית	אנגלית	עברית	בחינה II בחינה I
36 (0.3%)	15 (0.1%)	73 (0.7%)	5 *	1 *	6 *	8097 (76%)	עברית
20 (0.2%)					68 (0.6%)	8 *	אנגלית
11 *				85 (0.8%)			צרפתית
			100 (0.9%)			4 *	ספרדית
		1899 (17.8%)			1 *	74 (0.7%)	ערבית
	24 (0.2%)					8 *	רוסית
76 (0.7%)	3 *	19 (0.2%)	2 *	3 *	2 *	20 (0.2%)	משולב

* פחות מ-0.1%

** אחוז הנבחנים מחושב מתוך סה"כ הנבחנים (N = 10,656).

טבלה מס. 14 שכיחות יחסית של נבחנים חוזרים לפי שפת הבחינה*

	עברית	ערבית	אנגלית	צרפתית	ספרדית	רוסית	משולב	סה"כ**
סה"כ נבחנים בשנים 1983-86	79882	10185	1826	1421	2615	436	1347	97715
שכיחות נבחנים "כפולים"	8097	1899	68	86	100	24	76	10656
	10.1%	18.6%	3.7%	6.1%	3.8%	5.5%	5.6%	10.9%
שכיחות נבחנים "משולשים"	378	285	2	7	—	1		720
	0.47%	2.8%	0.11%	0.49%		0.23%		0.73%

* השכיחות מתייחסות לנבחנים שנבחנו באותה שפת בחינה בכל הבחינות החוזרות.

** הסה"כ כולל את כל הנבחנים החוזרים, לרבות נבחנים שהחליפו את שפת הבחינה.

טבלה מס. 15 ממוצע בחינה ראשונה ציון השיפור ופער הזמן עבור נבחנים "כפולים" שנבחנו בשפות שונות (כל הנבחנים N = 10656).

פער הזמן	ציון השיפור	ציון בחינה ראשונה	N	שפת הבחינה השנייה	שפת הבחינה הראשונה
13.8	32.5	478.8	8071	עברית	עברית
12.2	56.8	383.0	71	ערבית	עברית
13.2	41.7	530.7	6	אנגלית	עברית
10.2	-1.6	414.2	5	ספרדית	עברית
8.9	95.7	443.3	15	רוסית	עברית
13.5	38.7	464.1	36	משולב	עברית
12.8	11.4	414.7	1865	ערבית	ערבית
12.0	22.2	484.1	66	אנגלית	אנגלית
11.4	37.3	470.6	85	צרפתית	צרפתית
12.3	38.1	455.4	99	ספרדית	ספרדית
12.1	32.4	456.8	24	רוסית	רוסית
12.2	37.9	445.5	76	משולב	משולב
17.3	-13.4	391.9	70	עברית	ערבית
13.6	71.3	491.0	8	עברית	אנגלית
8.9	15.0	440.7	4	עברית	ספרדית
16.7	16.0	491.7	8	עברית	רוסית
17.6	19.1	440.2	20	עברית	משולב

בהשוואה בין המינים נערכו השוואות משני הסוגים. ההשוואה הראשונה עוסקת בהבדלים בין פרופורצית הגברים ופרופורצית הנשים בקרב אוכלוסיית הנבחנים החוזרים, דהיינו, האם יש הבדל בין גברים ונשים בנטיה לחזור על הבחינה הפסיכומטרית (ראה טבלה מס. 16). ההשוואה השנייה עוסקת בהבדלים בין גברים ונשים בציון השיפור, בציון הפסיכומטרי הסופי ובמבחנים השונים (ראה טבלאות 17 ו-18).

בהשוואה בין המינים בנוגע לפרופורצית הנבחנים החוזרים נמצא, כי אחוז הגברים בקרב הנבחנים ה"כפולים" גבוה באופן מובהק מאחוז הנשים בקבוצה זו ($p < 0.01$) $\chi^2=76.7$, וזאת בהשוואה ליחס בין נשים וגברים באוכלוסיית הנבחנים כולה. כמו כן, נמצא אחוז גבוה יותר של גברים בקרב קבוצת הנבחנים ה"משולשים" וזאת הן ביחס לאחוז הגברים באוכלוסיית הנבחנים כולה, הן ביחס לאחוז הגברים בקרב הנבחנים ה"כפולים"

($p < 0.05$) $\chi^2=28.05$. בהיעדר נתונים על התפלגות המינים בקרב נבחני עברית בלבד, לא ניתן לבדוק אם מגמה זו קיימת גם בתוך קבוצת נבחני עברית.

בהשוואה בין גברים ונשים בנושא ציון השיפור בין בחינות חוזרות, לא נמצאו הבדלים מובהקים. מעניין לציין כי ציון השיפור כמעט זהה בשתי הקבוצות, למרות שהציון הפסיכומטרי הממוצע בקבוצת הנשים נמוך במידה ניכרת מהציון המקביל בקבוצת הגברים ולפיכך צפוי היה שנשים תשפרנה את ציונן מעט יותר. בהשוואת ציון השיפור במבחנים השונים נמצא כי גברים משיגים ציון שיפור גבוה יותר במבחן במתמטיקה וזאת למרות שציונם במבחן זה בבחינה הראשונה גבוה במידה ניכרת מציונן של הנשים.

	נבחנים "משולשים"		נבחנים "כפולים"		כל הנבחנים בשנתונים 1984-85
	נבחני עברית	כל הנבחנים	נבחני עברית	כל הנבחנים	
זכר	215 (57.2%)	481 (67.1%)	4,219 (53.9%)	5,877 (57.3%)	29,904 (52.96%)
נקבה	161 (42.8%)	236 (32.9%)	3,611 (46.1%)	4,384 (42.7%)	26,560 (47.04%)
סה"כ	376	717	7,830	10,261	56,464

טבלה מס. 17 השוואה בין גברים ונשים על ממוצע ציון השיפור בבחינות חוזרות - נבחנים "כפולים" (נבחני עברית בלבד)

מבחן t	נקבה	זכר	משתנים
	3,606	4,202	N
t=17.4***	462 (83.5)	494 (78.5)	ממוצע ממוצע בחינה ראשונה ס.ת.
t=17.2***	494 (86.0)	526 (83.5)	ממוצע ממוצע בחינה שניה ס.ת.
t=0.35	32.4 (37.9)	32.7 (38.3)	ממוצע ציון השיפור בציון הפסיכומטרי ס.ת.
t=0.52	3.7 (9.9)	3.8 (9.8)	ממוצע ציון השיפור בידע כללי ס.ת.
t=-.85	6.1 (14.7)	5.8 (14.9)	ממוצע ציון השיפור במבחן צורות ס.ת.
t=-.83	4.5 (11.7)	4.3 (13.7)	ממוצע ציון השיפור במבחן הבנה ס.ת.
t=4.22 ***	5.4 (11.7)	6.5 (11.8)	ממוצע ציון השיפור במבחן מתמטיקה ס.ת.
t=-1.94	6.1 (10.6)	5.6 (11.0)	ממוצע ציון השיפור במבחן אנגלית ס.ת.
t=1.2	13.7 (4.9)	13.9 (5.3)	ממוצע הפרש זמן בין בחינות ס.ת.

p < .001 ***

השוואה בין גברים ונשים על ציון השיפור בין בחינות חוזרות - נבחנים "משולשים" (נבחני עברית בלבד)

מבחן t	נקבה	זכר	משתנים
	160	212	N
t=4.44***	428.5 (80.0)	464.4 (75.0)	ממוצע בחינה ראשונה
t=3.16**	458.9 (87.6)	486.0 (73.7)	ממוצע בחינה שניה
t=4.12***	478.5 (84.9)	513.4 (74.4)	ממוצע בחינה שלישית
t=-2.12*	30.4 (38.7)	21.6 (40.9)	שיפור בציון הפסיכומטרי בין בחינה ראשונה ושניה
t=1.91	19.7 (38.5)	27.3 (38.1)	שיפור בציון הפסיכומטרי בין בחינה שניה ושלישית
t=-0.26	50.1 (39.5)	48.9 (46.1)	שיפור בציון הפסיכומטרי בין בחינה ראשונה ושלישית

3.6 תרומתם של משתני רקע אחרים לציון השיפור - השכלה וגיל

אחת הסיבות האפשריות העשויות להסביר שיפור בציון הפסיכומטרי היא עיסוק אינטנסיבי בפעילות למידה כלשהי. במקרה זה, הדרך היחידה בה ניתן להתחקות אחר פעילות מסוג זה היא לבדוק שינוי ברמת השכלה כפי שהנבחן מדווח עליה בטופס ההרשמה. עם זאת, יש לזכור כי יתכן מצב בו הנבחן עסק בלמידה בפרק הזמן שעבר בין שתי הבחינות ועובדה זו לא תבוא לידי ביטוי בשינוי רמת השכלה. על מנת לבדוק את הקשר בין שינוי ברמת השכלה לבין שיפור בציון הפסיכומטרי, חושב מתאם בין שני המשתנים. המתאם שהתקבל הוא $r=0.005$ ל- $N=3393$ והוא אינו מובהק. מתאם זה נבדק פעם נוספת רק בתוך קבוצת הנבחנים אשר שינו רמת השכלה. בבדיקה זו נמצא מתאם $r=0.056$ ל- $N=641$. גם מתאם זה אינו מובהק. דהיינו, אין קשר בין שיפור רמת ההשכלה כפי שהיא מופיעה בטופס ההרשמה לבין שיפור בציון הפסיכומטרי.

בבדיקת הקשר בין הגיל בזמן הבחינה השניה לבין ציון השיפור נמצא מתאם שלילי, נמוך אומנם, אבל מובהק $r=-.04$, $p < 0.001$, $N=7640$. דומה כי התרומה המרכזית למתאם שלילי זה נובעת מהמתאם השלילי בין גיל לבין ציון השיפור במבחן באנגלית $r=-.041$. המתאם בין גיל לבין ציון השיפור במבחנים האחרים אינו מובהק.

- במרוצת השנים 1984-1986 נבחנו בבחינה הפסיכומטרית 97,715 נבחנים, מתוכם 79,882 נבחנים נבחנו בשפה העברית והיתר בשפות זרות. מתוך סה"כ הנבחנים, נבחנו 10,656 איש בבחינה הפסיכומטרית לפחות פעמיים, 720 איש נבחנו לפחות שלוש פעמים ו-22 איש נבחנו ארבע פעמים.
- מתוך כלל הנבחנים בשנתון בחינה נתון, נבחנים כ- 17%-13% איש בפעם השניה וכ- 2%-1% נבחנים בפעם השלישית. אחוזים אלה עשויים לעלות מעט עם השנים, כאשר מתקר המעקב יקיף פרקי זמן ארוכים יותר.
- מרבית הנבחנים חוזרים בפעם השניה על הבחינה הפסיכומטרית תוך פרק זמן הנע בין 9-14 חודש, (71%). 95.3% מהנבחנים החוזרים שנית על הבחינה, ניגשים לבחינה השניה תוך פחות משנתיים. פער הזמן הממוצע בין בחינה ראשונה ושניה הוא 13.8 חודש.
- קיים שיפור בציון הפסיכומטרי בין בחינה ראשונה לשניה הן בציון הסופי, הן בכל המבחנים המרכיבים את הבחינה הפסיכומטרית. השיפור הממוצע נע בין 0.3 ל-0.2 סטיות תקן למבחנים השונים ומגיע ל-32.5 נק' לציון הפסיכומטרי הסופי, שיפור שהוא כשליש סטיית התקן בהתפלגות הציון הפסיכומטרי (500, 100). שיפור זה קרוב בגודלו לשיפור המושג בבחינות ה-SAT שהוא כ-0.25 סטיות תקן. השיפור הגדול ביותר מושג במבחן במתמטיקה (5.96 נק') והשיפור הקטן ביותר מושג במבחן ידע כללי (3.81 נק').
- קיים שיפור בציון הפסיכומטרי ובמבחנים השונים גם במעבר בין בחינה שניה לשלישית. שיפור זה נע בין 5.05 ל-1.49 נק' למבחנים השונים ומגיע ל-23.8 נק' לציון הסופי. סה"כ השיפור הממוצע בציון הפסיכומטרי בין בחינה ראשונה ושלישית הוא 49.4 נק' שהן כמחצית סטיית התקן של התפלגות הציון הפסיכומטרי.
- ממוצע הציונים בבחינה הראשונה בקרב נבחנים חוזרים הוא נמוך מהממוצע באוכלוסיה; הממוצע לנבחנים "כפולים" הוא 478.8 ולנבחנים "משולשים" 449.2. גם פיזור הציונים של נבחנים חוזרים בכל הבחינות נמוך מהפיזור באוכלוסיה והוא נע בין 84.1 ל-78.9 סטיות תקן.
- כצפוי, קיים קשר מובהק בין השיפור בציון הפסיכומטרי לבין הציון בבחינה הראשונה $r = -.195$. ככל שהציון בבחינה הפסיכומטרית נמוך יותר כן גדלה ההסתברות לשיפור בבחינה השניה. תופעה זו היא למעשה אפקט הרגרסיה אל הממוצע. מבין המבחנים השונים, נראה כי מבחן הצורות הוא המבחן הקשה ביותר לשיפור, כאשר הציון שהושג בו בבחינה הראשונה הוא גבוה יחסית ($r = -.432$).
- מהימנות המבחן החוזר לציון הפסיכומטרי הסופי היא $r = .896$ באומדן המבוסס על בחינה ראשונה ושניה, ו- $r = .854$ באומדן המבוסס על בחינה שניה ושלישית. מהימנות מבחן חוזר למבחנים השונים נעה בטווח $r = .835$ - $.651$ בהשוואה בין בחינה ראשונה לשניה. היא נמוכה מעט מערך זה בהשוואה בין בחינה שניה ושלישית. המבחן בעל המהימנות הגבוהה ביותר באופן עקבי, הוא מבחן ידע כללי, והמבחן בעל המהימנות הנמוכה ביותר הוא מבחן הצורות.

- נמצא קשר שלילי מובהק אך חלש, בין פער הזמן בין הבחינות החוזרות לבין ציון השיפור במספר מבחנים, וזאת בניגוד לממצאים שהתקבלו על בחינת ה-SAT.
- בהשוואה בין נבחני שפות שונות נמצא כי מרבית הנבחנים בבחינה השנייה נבחנים באותה שפה בה נבחנו בבחינה הראשונה (97%). מקרב הנבחנים אשר החליפו את שפת הבחינה מרבית המעברים היו מעברית לערבית ולהיפך (46.7%), $N=147$, ומנוסח משולב לשפות אחרות כולל עברית ולהיפך, (34.6) $N=109$.
- אחוז הנבחנים העוברים מעברית לשפה אחרת גבוה מעט מאחוז הנבחנים העוברים משפה אחרת לעברית - 43.2%, לעומת 36.8% בהתאמה.
- האחוז הגבוה ביותר של נבחנים חוזרים מבין נבחני השפות השונות שייך לנבחנים בשפה הערבית. היחס בקבוצה זו הוא 18.6% לנבחנים "כפולים" ו-2.8% לנבחנים "משולשים". האחוז המקביל בקרב נבחני עברית הוא 10.1% לנבחנים "כפולים" ו-0.47% לנבחנים משולשים; אחוז זה קרוב למדי לאחוז הממוצע באוכלוסייה.
- בדרך כלל מושג ציון שיפור גבוה יותר במעבר מעברית לבחינה בשפה אחרת (לרוב שפת אם), יחסית לשיפור הממוצע בקרב נבחני עברית ויחסית לשיפור הממוצע בין שתי בחינות שנעשו באותה השפה. השיפור הממוצע בין שתי בחינות שנעשו באותה השפה נע בין 11.4 ל-38.1 נק' והוא בד"כ קרוב לממוצע השיפור בקרב נבחני עברית. בקבוצת הנבחנים שעברו משפה אחרת לעברית אין מגמת שיפור אחידה.
- בהשוואה בין המינים נמצא כי אחוז הגברים בקרב נבחנים חוזרים "כפולים" ו"משולשים" גבוה באופן מובהק מאחוז הנשים בשתי הקבוצות הנ"ל. וזאת בהשוואה ליחס בין נשים וגברים באוכלוסיית הנבחנים כולה.
- לא נמצא הבדל מובהק בציון השיפור בין גברים ונשים בקרב הנבחנים ה"כפולים" הן בציון הפסיכומטרי הסופי, הן במבחנים השונים, להוציא את מבחן המתמטיקה בו הגברים משיגים שיפור גבוה יותר באופן מובהק. גם בקרב הנבחנים ה"משולשים" לא נמצא הבדל בציון השיפור במעבר בין בחינה שניה לשלישית.
- לא נמצא קשר בין ציון השיפור לבין שינוי ברמת ההשכלה. עם זאת, אין להוציא את האפשרות שעיסוק אינטנסיבי בלמידה עשוי לשפר את הציון בבחינה הפסיכומטרית השנייה. יש להביא בחשבון כי המדד "שיפור ברמת השכלה" אינו בהכרח מייצג נאמנה את כמות הלמידה שהנבחן השקיע בין שתי הבחינות.
- נמצא קשר שלילי נמוך, אך מובהק, $r=-.04$, בין גיל לבין ציון השיפור. כלומר, קשה יותר לשפר את הציון בגילאים גבוהים יותר. נראה כי התרומה המרכזית לקשר שלילי זה נובעת מירידה בהישגים במבחן באנגלית, ככל שהגיל עולה.

מקורות

- Alderman, D.L. (1981a). Student self-selection and test repetition. Educational and Psychological Measurement, 41, 1073.
- Alderman, D.L. (1981b). Measurement error and SAT score change. College Board Report No. 81-9, ETS RR No. 81-39. New York: College Entrance Examination Board.
- Angoff, W.H. (1971). The college board admissions testing program: A technical report on research and development activities relating to the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board.
- ETS, (1982). ATP Guide for high schools and colleges, SAT and Achievement Tests. New York: College Entrance Examination Board.
- Kingston, N., & Turner, N. (1984). Analysis of score change patterns of examinees repeating the graduate record examinations general test. GRE Board Professional Report 83-85p. Princeton, New Jersey: ETS.
- Nevo, B. (1976). The effects of general practice, specific practice, and item familiarization on change in aptitude test scores. Measurement and Evaluation in Guidance, 9, 16-20.

