

---

# השפעת תוספת זמן על הציונים ועל מדרג הנבחנים בשלושת תחומי מכפ"ל

תמי קנת-כהן  
שמואל ברונר  
מרים איתן

אפריל 2008



**דוח מרכז 344**  
**ISBN:965-502-139-4**

## תוכן

2	תמצית
2	מבוא
4	שיטה
9	תוצאות
12	סיכום והערות
15	רשימת מקורות
17	נספחים
24	תמצית באנגלית

## תמצית

מטרת המחקר היתה לבדוק את ההשפעה של ארבע רמות של תוספת זמן בשלושת תחומי מכפ"ל על ההישגים ועל מדרג הנבחנים. במחקר השתתפו 516 נבדקים שנבחנו (בעברית ובתנאים רגילים) בבחינה ניסיונית שנערכה כחודש לפני מועד הבחינה הפסיכומטרית שאליה נרשמו. כל הנבחנים נבחנו בבחינה בת ששה פרקים, שניים בכל תחום. בכל אחד מן התחומים ניתן פרק אחד במסגרת הזמן הרגילה (25 דקות) ופרק אחר בתוספת זמן. הנבדקים הוקצו מקרית לאחד מארבעה תנאים של גודל תוספת הזמן – 5, 8, 10 או 13 דקות. ממצאי המחקר מעידים שתוספת זמן משפרת במעט את הביצוע בבחינה, כאשר מידת השיפור תלויה בגודל התוספת ובתחום מכפ"ל. אשר לגודל התוספת, נמצא שמידת השיפור בתוספת זמן של 8 דקות (4.7 נקודות בסולם של 50-150) גדולה באופן מובהק ממידת השיפור בתוספת זמן של 5 דקות (1.5 נקודות), ולא נמצא הבדל מובהק בין הרמות העוקבות האחרות של תוספת הזמן. בהתייחס לתחום מכפ"ל, נמצא שיש הבדל מובהק בין מידת השיפור בתחום הכמותי (5.3 נקודות) לבין מידת השיפור באנגלית (3.5 נקודות). מידת השיפור בתחום המילולי היתה נמוכה במעט מזו שבתחום הכמותי. הממצאים ביחס להשפעת גודל התוספת על הציון הכללי במכפ"ל תואמים את מה שנאמר לעיל ביחס לתחומים: מידת השיפור בתוספת זמן של 8 דקות (28 נקודות בסולם של 200-800) היתה גדולה באופן מובהק ממידת השיפור בתוספת זמן של 5 דקות (8 נקודות), ולא נמצא הבדל מובהק בין הרמות העוקבות האחרות של תוספת הזמן. השוואה בין דירוג הנבחנים בפרקים עם תוספת זמן לבין דירוגם בפרקים ללא תוספת זמן מגלה שבדרך כלל תוספת זמן אינה משנה את דירוג הנבחנים. עם זאת, נמצאו אינדיקציות לחוסר יציבות במדרג בתוספות הזמן הגדולות. הדיון בממצאי המחקר כולל התייחסות לממצאים ממחקרים מקבילים בעולם וכן המלצות בנוגע למערך המתאים לבחינת שאלות המחקר במחקרים בעתיד.

## מבוא

הישגים במבחני קבלה לאוניברסיטה אינם מיועדים בדרך-כלל להיות מושפעים ממהירות הביצוע של נבחנים. על-פי המדריך הטכני של ה-SAT, המהירות שבה נבחנים יכולים לענות על שאלות צריכה למלא תפקיד מזערי, אם בכלל, בקביעת הציונים (Donlon, 1984). באופן דומה, המדריך הטכני של ה-GRE (Briel, O'Neill, & Scheuneman, 1993) מצהיר שמבחני ה-GRE אינם אמורים להיות מואצים. למרות זאת, במבחנים הללו, בדומה למתקיים במכפ"ל, יש מגבלות זמן שנכפות בעיקר בשל אילוצים אדמיניסטרטיביים.

שתי הגדרות הוצעו באופן מסורתי (Lu & Sireci, 2007) למונח "מואצות" (speededness), אשר נבדלו זו מזו באימוץ פרספקטיבה של "אוריינטציה לציון" (score oriented) או של "אוריינטציה להוגנות" ("fairness oriented"). פרספקטיבה של אוריינטציה לציון מודדת את מידת המואצות של מבחן על-פי מידת ההשפעה שיש למגבלת הזמן על ציוני הנבחנים. אם ציוני הנבחנים במגבלת הזמן הקיימת במבחן נמוכים מאלה שהיו מתקבלים לו היה עומד לרשותם זמן בלתי מוגבל, המבחן נחשב למואץ (Bejar, 1985). פרספקטיבה של אוריינטציה להוגנות מגדירה מואצות כמידה שבה נבחנים מסוימים יוצאים נפסדים ממגבלת הזמן במבחן, יחסית לנבחנים אחרים (Schnipke, 1995). אם הדירוגים היחסיים של נבחנים בהתפלגות הציונים משתנים עם הכנסת מגבלת זמן, המבחן נחשב למואץ. הגדרת מואצות מתוך פרספקטיבה של אוריינטציה להוגנות רלוונטית יותר למבחני נורמה, כדוגמת מכפ"ל, שהם המבחנים האופייניים בהקשר של קבלה לאוניברסיטה. באופן כללי, המושג של מואצות מתוך אוריינטציה לציון הוא מושג רחב ומחמיר מזה של מואצות מתוך אוריינטציה להוגנות (מבחן יכול להיחשב מואץ מפרספקטיבה של אוריינטציה לציון ולא מואץ מפרספקטיבה של אוריינטציה להוגנות, אך לא להיפך).

מחקרים עדכניים בנושא של השפעת הוספת זמן בחינה על ציוני הנבחנים נערכו ביחס ל-SAT (Bridgeman, Trapani & Curley, 2004), ל-SAT המחודש (Bridgeman, 2004) ול-GRE (Bridgeman, Cline & Hessinger, 2004). בשני מחקרי ה-SAT נמצא שתוספת זמן של 50% הניבה שיפור מסויים (של 10-20 נקודות בסולם של 200-800) בחלק הכמותי ושיפור קטן יותר (של פחות מ-10 נקודות) בחלק המילולי. במחקר ה-GRE התקבל שיפור קטן (של 7 נקודות בסולם של 200-800) הן בחלק הכמותי והן בחלק המילולי. ממצאים אלה, ברוח הגדרת המואצות מתוך פרספקטיבה של אוריינטציה לציון, לצד ממצאים משלימים (שלא פורטו לעיל) ברוח הגדרת מואצות מתוך פרספקטיבה של אוריינטציה להוגנות (בהתבסס על השוואה בין המתאמים של הציונים עם/בלי תוספת זמן עם קריטריון חיצוני: הציון בבחינה בפועל או ציוני ביי"ס באנגלית ובמתמטיקה) הובילו למסקנה שאין אינדיקציה למואצות במבחנים שנחקרו.

המחקר הנוכחי ביקש לבדוק את מידת המואצות של מכפ"ל. באופן מפורט, מטרת המחקר הייתה לבדוק את ההשפעה של ארבע רמות של תוספת זמן (שנעו בין 20% ל-50% מהזמן המוקצב לפרק) בשלושת תחומי מכפ"ל על ההישגים (פרספקטיבה של אוריינטציה לציון) ועל מדרג הנבחנים (פרספקטיבה של אוריינטציה להוגנות). שאלות המחקר לפיכך היו:

1. האם תוספת זמן משפרת את הציון, ואם כן – האם יש קשר בין גודל התוספת לבין מידת השיפור?

באופן ספציפי, נעשה ניסיון לאתר "נקודת רוויה" – רמת תוספת זמן שבה השיפור ממוצה, ותוספת זמן מעבר לה לא תשפר את הביצוע בבחינה.

2. האם תוספת זמן משנה את מדרג הנבחנים, ואם כן – האם יש קשר בין גודל התוספת לבין מידת השינוי במדרג?

שאלות המחקר ייבדקו תוך השוואה בין תחומי מכפ"ל, וברוח זו יוצגו גם הממצאים. ממצאים הנוגעים לציון הכללי במכפ"ל יוצגו בנספח.

## שיטה

### נבדקים

במחקר השתתפו 516 נבדקים שנענו בחיוב לפנייה של המרכז הארצי לבחינות ולהערכה להשתתף בבחינה ניסיונית. הפניה נעשתה לכ-1,500 נרשמים לבחינה הפסיכומטרית אשר נדגמו אקראית מתוך הנרשמים לבחינה בירושלים או בתל-אביב באחד מן המועדים הבאים: דצמבר 2005, אפריל 2006 או דצמבר 2006. ל-1,500 הנרשמים נשלח מכתב ובו הצעה להשתתף בבחינה ניסיונית של המרכז הארצי לבחינות ולהערכה, הדומה מאוד בצורתה לבחינה הפסיכומטרית. נרשמים אלה הוזמנו להשתתף בבחינה כסוג של אימון לקראת הבחינה הפסיכומטרית, ונאמר להם כי בסופה יוכלו לקבל ציון שיהווה עבורם אומדן כללי לציון בבחינה. הובהר להם שאין להשתתפותם בניסוי כל השפעה על השתתפותם או ציונם בבחינה הפסיכומטרית עצמה, וכי הציון הוא הערכה בלבד ואינו מנבא במדויק או מחליף את ציון הבחינה. הבחינה הניסיונית התקיימה כחודש לפני מועד הבחינה הפסיכומטרית שאליה נרשמו.

### כלי המחקר

כלי המחקר היה נוסח מסויים (בעברית) של הבחינה הפסיכומטרית. ציוני התחומים והציון הסופי חושבו על-בסיס פרק יחיד בכל תחום, תוך התייחסות לציון בפרק זה כאילו היה ממוצע של הציונים בשני פרקי התחום.

## מעריך המחקר

כל הנבדקים נבחנו במבחן בן ששה פרקים, שניים בכל תחום. סדר הפרקים – אנגלית; מילולי; כמותי; אנגלית; מילולי; כמותי – היה זהה לכולם. בכל אחד מן התחומים ניתן פרק אחד במסגרת הזמן הרגילה – 25 דקות, ופרק אחר בתוספת זמן. הנבדקים הוקצו מקרית, בעת שהוזמנו לניסוי, לאחד מארבעה תנאים של גודל תוספת הזמן – 5, 8, 10 או 13 דקות. ההבדלים במידת ההיענות בקרב המוזמנים לתנאי הניסוי השונים (בהתייחס לתנאים של תוספת זמן של 5, 8 או 13 דקות) מסבירים את הפערים במספר הנבדקים בפועל. לתנאי של תוספת זמן של 10 דקות הוזמנה קבוצה גדולה יותר, כיוון שקבוצה זו שימשה כקבוצת ביקורת במחקר אחר שנערך במקביל (אנגל, 2008). בכל אחד מארבעת התנאים של גודל תוספת הזמן בוצעה הקצאה מקרית לאחד משני תנאים של עיתוי מתן התוספת – בפרקים 1, 3 ו-5 (הפרקים האי-זוגיים) או בפרקים 2, 4 ו-6 (הפרקים הזוגיים).

- התיאור שלעיל מגדיר מערך מחקר שכולל ארבעה משתנים בלתי-תלויים ("גורמים")<sup>1</sup>:
1. מתן תוספת זמן (גורם תוך-נבדקי): עם או בלי תוספת זמן.
  2. תחום מכפ"ל (גורם תוך-נבדקי): חשיבה מילולית (מילולי), חשיבה כמותית (כמותי), אנגלית (אנגלית).
  3. גודל התוספת (גורם בין-נבדקי): 5, 8, 10 או 13 דקות.
  4. עיתוי מתן התוספת (גורם בין-נבדקי): בפרקים אי-זוגיים או בפרקים זוגיים.

בנספח 1 בלוח 4 מוצגים מספר התצפיות, וכן הממוצעים וסטיות-התקן של המשתנה התלוי, במערך שתואר לעיל.

במקום המשתנה "עיתוי מתן התוספת" (האם התוספת ניתנה בפרקים האי-זוגיים או הזוגיים במבחן) הוגדר המשתנה "סדר מתן התוספת" (האם התוספת ניתנה בפרק הראשון או השני בתחום). פירוט הבדיקות והממצאים ביחס למשתנה זה מופיע בנספח 2<sup>2</sup>.

## מהלך המחקר

<sup>1</sup> המשתנה התלוי הוא הציון.

<sup>2</sup> מאחר שלא נמצא לסדר מתן התוספת אפקט עקבי מעבר לתחומים (ראו נספח 2), ובהינתן שהפעלת תנאים שונים של סדר מתן התוספת נהגתה מלכתחילה רק במטרה ליצור איזון בהפעלת המניפולציה העיקרית (הוספת פרקי זמן שונים) ולא כדי לבדוק שאלת מחקר כלשהי, בחרנו מכאן ואילך להשמיט את גורם הסדר מניתוח התוצאות (או, במלים אחרות, לנתח את האפקטים של יתר הגורמים מעבר לשני תנאי הסדר).

הנבדקים נבחנו על-פי נוהל הבחינה הרגיל. לאחר קריאת הוראות הסבר כלליות על מבנה הבחינה, קיבלו הנבדקים את חוברות הבחינה ותשובונים לסימון התשובות. בשלב זה נאמר לנבדקים שבבחינה זו יש שינוי בהקצאת הזמן ביחס לבחינה האמיתית. בראש כל פרק הכריז הבוחן כמה זמן מוקצב לפרק. כמו-כן, הזמן המוקצב לפרק הופיע בחוברת הבחינה.

### **בדיקת איכות הנתונים**

במטרה לוודא שנתוני המחקר תקפים (במובן שהם משקפים ביצוע בבחינה פסיכומטרית אמיתית) בוצעה סריקה אחר ציוני הבחינה הפסיכומטרית האמיתית של 516 משתתפי המחקר (במועדים דצמבר 2005 עד דצמבר 2006 ועד בכלל). נמצאו נתונים עבור 505 נבחנים. על בסיס נתונים אלה חושבו ממוצעים וסטיות תקן של הציונים בפרקי הבחינה האמיתית וכן מתאמים בין הציונים הללו לבין הציונים במחקר הנוכחי (ראו נספח 3). הערכים שהתקבלו מעידים שנתוני המחקר תקפים, ואינם מספקים אינדיקציה לכך שהסיטואציה הניסויית שבה נאספו מגבילה את התוקף החיצוני של מסקנות המחקר.

### **שלבים בניתוח הנתונים**

כל אחת משתי שאלות המחקר כללה שאלה מקדימה: האם יש אפקט לתוספת זמן (על הציונים – בשאלת המחקר הראשונה; על מדרג הנבחנים – בשאלת המחקר השנייה). תשובה חיובית על השאלה המקדימה היוותה תנאי להמשך הבדיקה, שהתמקדה, בכל אחת משתי השאלות, בחקירת אפקט תוספת הזמן כפונקציה של שני **גורמים מתנים** (moderators): גודל התוספת ותחום מכפ"ל. ברוח זו בוצעו העיבודים הסטטיסטיים, כפי שיפורט להלן.

*שאלת המחקר הראשונה: האם תוספת זמן משפרת את הציון, ואם כן – האם יש קשר בין גודל התוספת לבין מידת השיפור?*

1. בדיקה של השאלה המקדימה:

השאלה המקדימה היתה, כאמור, אם תוספת זמן (בגדלים שונים) משפרת את הציונים (בתחומים שונים). שאלה זו נבחנה (בכל אחד מן הצירופים של תחום מכפ"ל וגודל התוספת)



על-ידי בדיקת השערה חד-צדדית (שתוחלת הציון עם תוספת זמן גבוהה מתוחלת הציון ללא תוספת זמן) במדגמים מזווגים בעזרת התפלגות  $t^3$ .

2. וידוא שהרמה ה"התחלתית"<sup>4</sup> (ציון בלי תוספת זמן) דומה בערכים השונים של הגורמים המתנים:

במטרה לבדוק שקשר שעשוי להימצא בין האפקט של תוספת הזמן (שהוא, הפער בין ציון עם תוספת לבין ציון בלי תוספת) לבין גורם מתנה כלשהו אינו תוצר של הבדלים ב"רמה ההתחלתית" (כלומר בציון בלי תוספת זמן) בין הערכים השונים של אותו גורם מתנה<sup>5</sup>, בוצע ניתוח שונות דו-גורמי עם מדידות חוזרות (בתחומים מילולי, כמותי ואנגלית) בין ארבע רמות של גודל תוספת הזמן על הציון ללא תוספת זמן. לא נמצא אפקט מובהק לגודל תוספת הזמן או לתחום מכפ"ל, אך האינטראקציה בין שני המשתנים נמצאה מובהקת סטטיסטית:  $F[6,1024] = 2.80, p < 0.05$ . תוצאות הניתוח מוצגות בנספח 4.

3. ניכוי הרמה ההתחלתית:

לאור העובדה שנמצא במדגם קשר בין הגורמים המתנים לבין הרמה ההתחלתית, אזי כדי לחקור את ההפרש שבין הציון עם תוספת זמן לבין הציון ללא תוספת זמן כפונקציה של הגורמים המתנים, יש לנכות את השפעתו של הציון ללא תוספת זמן על ההפרש הנ"ל, שיכונה להלן ה"הפרש הגולמי". כדי לנכות את השפעתה של הרמה ההתחלתית על ההפרש הגולמי חושב (עבור כל תחום מכפ"ל בנפרד – מעבר לערכים השונים של גודל התוספת) קו הרגרסיה לניבוי ההפרש הגולמי לפי הציון ללא תוספת זמן. ציון הטעות (השארית) מן הרגרסיה הזו<sup>6</sup> – שהוא המשתנה התלוי החדש – סולם כך שהממוצע וסטיית התקן שלו יהיו זהים לאלו של ההפרש הגולמי בכל אחד משלושת תחומי מכפ"ל. ציון זה יכונה להלן "הפרש מנוכה". פעולת הסילום שכרוכה בהגדרת ההפרש המנוכה נועדה לאפשר זיהוי אפקט של תחום מכפ"ל, שכן ללא סילום היה ממוצע ציוני הטעות זהה (שווה ל-0) בשלושת התחומים<sup>7</sup>. חשוב להדגיש שהסילום שבוצע משמר את ממוצעי הפרשים הגולמיים (שלא נוכתה מהם השפעת הרמה ההתחלתית) בין הציון עם תוספת זמן לבין הציון ללא תוספת זמן בתחומי מכפ"ל. לא ניתן

<sup>3</sup> בהינתן שמדובר בשתיים-עשרה השוואות (שחלקן תלויות) יש לקחת בחשבון שהסיכוי של טעות אחת לפחות במערכת ההשוואות גבוהה מהסיכוי לטעות בכל השוואה יחידה.

<sup>4</sup> "התחלתית" (במרכאות) כיוון שבמחצית מן המקרים נקבעה הרמה ההתחלתית כציון בפרק השני בתחום. <sup>5</sup> אפריורית אין סיבה לצפות שיהיו הבדלים בציון בלי תוספת זמן בין ערכים שונים של הגורמים המתנים: בהתייחס לגורם גודל התוספת נזכיר שבוצעה הקצאה מקרית של נבדקים לתנאי הניסוי השונים, ובהתייחס לגורם תחום מכפ"ל נציין שממוצעי הציונים בשלושת תחומי מכפ"ל בקרב הנבחים בעברית הם דומים (פבלוב ואורן, 2002).

<sup>6</sup> ציון הטעות הזה זהה לציון הטעות המתקבל מרגרסיה של הציון עם תוספת זמן על הציון ללא תוספת זמן. <sup>7</sup> מה שלא מונע את בדיקת האפקט של גודל התוספת ושל האינטראקציה בין התחום לגודל התוספת על ציוני הטעות.

לנכות, במערך המחקר הנוכחי, את השפעת הרמה ההתחלתית על ההפרשים האלה. יש, אם כן, חוסר עקביות בין ניתוח האפקט של גודל תוספת הזמן לבין ניתוח האפקט של תחום מכפ"ל: בעוד שניתוח האפקט של תחום מכפ"ל מסתמך על הסילום של ציוני הטעות, ניתוח האפקט של גודל התוספת (ושל האינטראקציה בין גודל התוספת לבין תחום) איננו מושפע מסילום זה.

4. ניתוח אפקט תוספת הזמן על ההפרש המנוכה כפונקציה של הגורמים המתנים: המשתנה התלוי הוא, כאמור, הפרש מנוכה (מן הציון ללא תוספת זמן) בין הציון עם תוספת זמן לציון ללא תוספת, שמאפייני התפלגותו (ממוצע וסטיית-תקן) בכל תחום זהים לאלו של ההפרש הגולמי. בניסיון להסביר הפרש זה יבוצע ניתוח שונות דו-גורמי עם גודל תוספת הזמן (גורם בין-נבדקי) ותחום מכפ"ל (גורם תוך-נבדקי) כמשתנים בלתי תלויים.

*שאלת המחקר השניה: האם תוספת זמן משנה את מדרג הנבחנים, ואם כן – האם יש קשר בין גודל התוספת לבין מידת השינוי במדרג?*

1. בדיקה של השאלה המקדימה:

השאלה המקדימה היתה, כאמור, האם תוספת זמן (בגדלים שונים) משנה את מדרג הנבחנים (בתחומים שונים) בהשוואה לדירוגם במצב ללא תוספת זמן. במטרה לענות על שאלה זו חושבו מתאמי פירסון בין ציוני הנבחנים עם ובלי תוספת זמן בכל אחד מ-12 הצירופים של גודל תוספת הזמן ותחום מכפ"ל. מתאמים אלה הושאו (Cohen & Cohen, 1983, pp. 53-55), בכל תחום, למתאמים המתקבלים בין שני פרקים בתחום בבחינה רגילה. האחרונים התבססו על הציונים בבחינה האמיתית של משתתפי המחקר.

2. ניתוח אפקט תוספת הזמן על השינוי במדרג כפונקציה של הגורמים המתנים: לצורך זה יבוצעו בכל רמה של תוספת זמן (שלוש) השוואות בין התחומים (Raghunathan, Rosenthal & Rubin, 1996) ובכל תחום – (שש) השוואות בין רמות תוספת הזמן (Cohen & Cohen, 1983, pp. 53-55). על-סמך מספר ההשוואות המובהקות וזהותן ננסה להתרשם בדבר דפוס אפשרי של השתנות המתאמים.

## תוצאות<sup>8</sup>

שאלת המחקר הראשונה

האם תוספת זמן (בגדלים שונים) משפרת את הציון (בתחומים שונים)?

בלוח 1 מוצגים הממוצעים וסטיות התקן של הציון עם תוספת זמן והציון ללא תוספת זמן וכן תוצאות מבחני t שבוצעו להשוואה בין שני הציונים (על-ידי בדיקת ההשערה שתוחלת הציון עם תוספת זמן גבוהה מתוחלת הציון ללא תוספת זמן) בצירופים השונים של תחום מכפ"ל וגודל תוספת הזמן.

לוח 1: ממוצעים (וסטיות תקן) של הציון ללא תוספת זמן והציון עם תוספת זמן ותוצאות מבחני t

תחום מכפ"ל									n	גודל התוספת
אנגלית			כמותי			מילולי				
t	עם תוספת	ללא תוספת	t	עם תוספת	ללא תוספת	t	עם תוספת	ללא תוספת		
0.23	123.2 (20.6)	122.8 (21.2)	0.99	120.8 (18.2)	119.5 (18.5)	1.83*	120.9 (20.3)	118.5 (19.7)	64	5
3.21**	125.3 (21.7)	121.9 (20.6)	5.27***	124.6 (19.0)	118.7 (18.3)	3.41**	121.7 (19.1)	117.3 (20.8)	87	8
6.98***	121.4 (22.5)	116.8 (23.4)	6.86***	123.5 (18.5)	118.2 (18.6)	7.27***	122.5 (20.0)	117.2 (20.1)	270	10
2.35*	120.1 (24.2)	117.4 (22.3)	6.93***	124.2 (17.4)	116.6 (18.6)	3.89**	126.7 (15.5)	121.0 (17.9)	95	13
			$p < .0001$ ***			$p < .001$ **			$p < .05$ *	

כמעט בכל הצירופים של גודל התוספת ותחום מכפ"ל נמצא שהציון עם תוספת זמן גבוה באופן מובהק מן הציון ללא תוספת. החריגים היחידים מצויים בתוספת זמן של 5 דקות בתחום הכמותי ובאנגלית, שם השינוי בציון אינו מובהק מבחינה סטטיסטית, אך גם במקרים אלה מתקיימים הבדלים בכיוון המצופה. הממצאים שלעיל מספקים אינדיקציה לכך שתוספת זמן, בגדלים שונים, משפרת את הציון בתחומים שונים. חשוב להזכיר – ואפשר להתרשם מכך גם מעיון בערכי הממוצעים – שיש הבדלים בין הצירופים השונים של גודל התוספת ותחום מכפ"ל ברמה ההתחלתית. לפיכך השאלה – האם מידת השיפור שונה בין גדלים שונים של תוספת זמן, בין תחומים שונים ו/או בין צירופים שונים של שני הגורמים הללו –

<sup>8</sup> ממצאים מקבילים ביחס לציון הכללי במכפ"ל מוצגים בנספח 5.

תיבדק, כפי שצויין בפרק השיטה, ביחס להפרש בין ציון עם תוספת זמן לציון ללא תוספת זמן מנוכה מן הציון ללא תוספת זמן.

האפקט של גודל תוספת הזמן ותחום מכפ"ל על ההפרש המנוכה בציון

נערך ניתוח שונות דו-גורמי עם מדידות חוזרות (בשלושת תחומי מכפ"ל) על ההפרש המנוכה בציון בארבע הקבוצות של גודל תוספת הזמן. התוצאות מוצגות בלוח 2.

לוח 2: ממוצעים וסטיות תקן ומבחן שונות על ההפרש המנוכה בציון, לפי גודל תוספת הזמן, בשלושת תחומי מכפ"ל

גודל תחום F(6,1024)	תחום F(2,1024)	גודל F(3,512)	תחום						n	גודל התוספת
			אנגלית		כמותי		מילולי			
			ס"ת	ממוצע	ס"ת	ממוצע	ס"ת	ממוצע		
1.05	3.71*	4.92**	11.1	0.8	10.5	1.3	11.5	2.4	64	5
			10.2	3.8	11.0	6.1	11.6	4.2	87	8
			10.7	4.4	12.6	5.3	12.4	5.2	270	10
			11.8	2.6	10.5	7.4	13.4	6.5	95	13

$p < .01$  \*\*  $p < .05$  \*

עיון בממצאי הלוח מעלה שהשינוי בציון כתוצאה מהוספת זמן מושפע מגודל התוספת. ממוצע (מעבר לשלושת התחומים)<sup>9</sup> ההפרש המנוכה בציון עבור ארבעת התנאים של גודל התוספת היה 1.5, 4.7, 5.0 ו-5.5 עבור תוספת זמן של 5, 8, 10 ו-13 דקות בהתאמה. שלוש השערות מתוכננות נוסחו ביחס למקור ההבדלים בין ארבעת התנאים של גודל התוספת: (1) תוספת זמן של 8 דקות משפרת את הציון יותר מתוספת זמן של 5 דקות; (2) תוספת זמן של 10 דקות משפרת את הציון יותר מתוספת זמן של 8 דקות; ו-(3) תוספת זמן של 13 דקות משפרת את הציון יותר מתוספת זמן של 10 דקות. רק ההבדל בין התנאי של תוספת זמן של 5 דקות לבין התנאי של תוספת זמן של 8 דקות נמצא מובהק סטטיסטית ( $F[1,512] = 7.61, p < .01$ ).

עוד עולה מן הלוח שהשינוי בציון כתוצאה מהוספת זמן מושפע מתחום מכפ"ל. ממוצע (מעבר לארבעת התנאים של גודל התוספת) ההפרש המנוכה בציון עבור שלושת תחומי מכפ"ל היה 4.9, 5.3 ו-3.5 עבור מילולי, כמותי ואנגלית בהתאמה.

<sup>9</sup> חושב בשני שלבים: ראשית, חושב לכל אדם ממוצע פשוט של ההפרשים המנוכים בציוניו בשלושת התחומים. שנית, בכל אחד מארבעת התנאים של גודל התוספת חושב ממוצע, מעבר לאנשים, של הממוצע של השלב הקודם.

מבחן post-hoc של Tukey העיד שרק ההבדל בין ההפרש הממונה בציון באנגלית לבין ההפרש הממונה בציון בתחום הכמותי הוא מובהק (ההפרש המינימלי בין שני ממוצעים שיכול להיחשב כמובהק<sup>10</sup> עבור  $\alpha = 0.05$  ו-1546 ד"ח הוא 1.71). כפי שעוד ניתן להיווכח מלוח 2, לא נמצאה אינטראקציה בין גודל התוספת לבין תחום מכפ"ל.

### שאלת המחקר השנייה

#### האם תוספת זמן משנה את מדרג הנבחנים?

בלוח 3 מוצגים מתאמי פירסון בין ציוני הנבחנים ללא תוספת זמן לבין ציוניהם עם תוספת זמן לפי תחום מכפ"ל וגודל תוספת הזמן.

לוח 3: מתאמי פירסון בין ציוני הנבחנים עם ובלי תוספת זמן, לפי גודל תוספת הזמן, בשלושת תחומי מכפ"ל

תחום			n	גודל התוספת
אנגלית	כמותי	מילולי		
0.86	0.84	0.86	64	5
0.89	0.84	0.82	87	8
0.89	0.77	0.82	270	10
0.88	0.82	0.64	95	13

המתאמים בין הציונים בפרק ללא תוספת זמן לבין הציונים בפרק עם תוספת זמן במחקר הנוכחי הושאו למתאמים בין שני פרקים בבחינה רגילה. מתאמים כאלה – בין הפרק הראשון לפרק השני בכל תחום – חושבו בבחינה האמיתית של 505 ממשותפי המחקר, שציוניהם האמיתיים אותרו. הערכים שהתקבלו היו: 0.81, 0.83 ו-0.90 עבור התחומים מילולי, כמותי ואנגלית בהתאמה. בוצעו 12 מבחנים סטטיסטיים (Cohen & Cohen, 1983, pp. 53-55), ארבעה לכל תחום, לבדיקת מובהקות ההבדל בין המתאמים שהתקבלו במחקר לבין המתאם, שהוצג לעיל, בין שני הפרקים בתחום בבחינה האמיתית. ב-10 מתוך 12 ההשוואות לא נמצא הבדל מובהק סטטיסטית בין המתאם שהתקבל במחקר לבין ערכו של המתאם בין שני פרקים בבחינה האמיתית<sup>11</sup>. ניתן לסכם שלא נמצאה ירידה שיטתית בגובה

<sup>10</sup> הפרש זה, המכונה HSD (Honestly Significant Difference), חושב על-פי הנוסחה שמשמשת עבור גורם בין-נבדקי, שמציבה תנאי מחמיר לקבלת תוצאה מובהקת תחת מערך תוך-נבדקי.  
<sup>11</sup> הבדל מובהק בכיוון הצפוי התקבל בתחום המילולי בתוספת זמן של 13 דקות ( $z=-3.49, p<.01$ ) ובתחום הכמותי בתוספת זמן של 10 דקות ( $z=-2.84, p<.01$ ).

המתאם בין שני פרקים בעקבות הוספת זמן לאחד מהם. לאור ממצא זה לא בוצעו השוואות נוספות בין המתאמים שבלוח 3.

## סיכום והערות

### ממצאים עיקריים

ממצאי המחקר מעידים שתוספת זמן משפרת במעט את הביצוע בבחינה: בכל הרמות של גודל התוספת בשלושת תחומי מכפ"ל נמצא שהציון הממוצע עם תוספת זמן גבוה מן הציון ללא תוספת. אפקט זה נמצא מובהק במרבית הצירופים של גודל התוספת ותחום מכפ"ל.

בחקירת האפקט של גודל תוספת הזמן ותחום מכפ"ל על מידת השיפור נמצא אפקט מובהק לכל אחד משני הגורמים (אך לא לאינטראקציה ביניהם). בהתייחס לגורם של גודל תוספת הזמן נוסחו מראש השערות ספציפיות במטרה לזהות את גודל תוספת הזמן שבה השיפור ממוצה. נמצא שמידת השיפור בתוספת זמן של 8 דקות גדולה באופן מובהק ממידת השיפור בתוספת זמן של 5 דקות (הפרש מנוכה ממוצע של 4.7 ו-1.5 נקודות בהתאמה, בסולם של 50-150), ולא נמצא הבדל מובהק בין הרמות העוקבות האחרות של תוספת הזמן. ממצאים אלה מרמזים על כך שתוספת זמן של כ-8 דקות מאפשרת מיצוי של השיפור שניתן להשיג באמצעות תוספת זמן. בהתייחס לגורם של תחום מכפ"ל, נמצא שיש הבדל מובהק בין מידת השיפור בתחום הכמותי לבין מידת השיפור באנגלית (הפרש מנוכה ממוצע של 5.3 ו-3.5 נקודות בהתאמה). ממצא זה מספק אינדיקציה לגבי הדיפרנציאציה בין תחומי הבחינה השונים במידת המואצות.

הממצאים ביחס לציון הכללי במכפ"ל תואמים את מה שנאמר לעיל ביחס לתחומים: תוספת זמן, בכל הרמות של גודל התוספת, משפרת את הציון באופן מובהק. מידת השיפור נמצאה תלויה בגודל התוספת, כאשר השיפור בתוספת זמן של 8 דקות גדול באופן מובהק מן השיפור בתוספת זמן של 5 דקות (הפרש ממוצע של 28 ו-8 נקודות בהתאמה, בסולם של 200-800) ולא נמצא הבדל מובהק בין הרמות העוקבות האחרות של תוספת הזמן.

השוואה בין דירוג הנבחנים בתנאי עם תוספת זמן לבין דירוגם בתנאי ללא תוספת זמן (באמצעות השוואת המתאם בין ציוניהם בפרק עם תוספת זמן לציוניהם בפרק בלי תוספת זמן למתאם הצפוי בין שני פרקים בבחינה רגילה) מגלה שבדרך כלל תוספת זמן אינה משנה את דירוג הנבחנים, מעבר לתנודות מקריות (שמקורן במהימנות בלתי מושלמת), וזאת

בהתייחס לכל אחד משלושת התחומים. יחד עם זאת, יש אינדיקציות לחוסר יציבות במדרג כשתוספת הזמן גדלה: שני המקרים שבהם נמצאה ירידה במתאם בין הציונים בפרק עם תוספת זמן לציונים בפרק בלי תוספת זמן היו בתוספות הזמן הגדולות (10 ו-13 דקות).

אינדיקציות אלה נתמכות בממצאים ביחס לציון הכללי במכפ"ל, שם נצפתה, בתוספות הזמן הגדולות, ירידה במתאם בין הציונים בשני חצאי הבחינה (החצי הכולל את הפרקים עם תוספת זמן והחצי הכולל את הפרקים ללא תוספת זמן) בהשוואה למתאם הצפוי בין שני חצאים בבחינה רגילה. כמו-כן נמצא, בהתייחס לתוספות הזמן הגדולות, שככל שתוספת הזמן גדולה יותר, חוסר היציבות במדרג גדול יותר.

ממצאי המחקר הנוכחי מתיישבים באופן חלקי עם ממצאים שהתקבלו במחקרים מקבילים בעולם (Bridgeman, 2004; Bridgeman, Cline & Hessinger, 2004; Bridgeman, 2004; Trapani & Curley, 2004). הלימה בין המחקר הנוכחי למחקרי ה-SAT וה-GRE מתקיימת בהתייחס לממצא השיפור בתחום הכמותי כתוצאה מהוספת זמן גדול מהשיפור בתחומים המילוליים. עם זאת, מידת השיפור בציון שנמצאה במחקר הנוכחי הייתה גבוהה במעט מזו שנצפתה במחקרים שלעיל. כך, למשל, תוספת זמן של פי 1.5 הניבה בתחום הכמותי במחקר הנוכחי שיפור ממוצע של כשליש סטיית-תקן בהשוואה לשיפור של כעשירית עד חמישית סטיית תקן במחקרים שלעיל. בין הגורמים שעשויים להסביר את ההבדל בין המחקרים ניתן למנות הבדלים משמעותיים בין המתודולוגיה של שלושת המחקרים שלעיל (שהתבססו על מערך שבו הן מתן התוספת והן תחום המבחן תופעלו כגורמים בין-נבדקיים) לזו של המחקר הנוכחי (ראו הרחבה בעניין זה להלן). לא נצפו הבדלים עקביים בין המחקר הנוכחי למחקרי ה-SAT וה-GRE בהתייחס להשפעת תוספת זמן על מדרג הנבחנים.

## הערות והמלצות

1. המגבלות המרכזיות בנייתוח הנתונים ובהסקת מסקנות מהם נובעות מן העובדה ששני משתנים בלתי-תלויים – "עם או בלי תוספת זמן" ו"תחום מכפ"ל" – תופעלו כגורמים תוך-נבדקיים, כלומר, במערך של מדידות חוזרות. בהתייחס למשתנה "עם או בלי תוספת זמן", ההשלכה העיקרית של תפעולו כגורם תוך-נבדקי (מעבר לתפעול הבעייתי של המשתנה "סדר מתן התוספת") היא שעצם **המעבר בין התנאים** (בלי תוספת, עם תוספת) עלול ליצור אפקט מלאכותי (ארטיפקט) של המניפולציה המחקרית (כגון ציפיות מטעות), שלא ניתן להפרידו מן האפקט שמעניין אותנו (זה של זמני היבחנות שונים). בהתייחס למשתנה "תחום מכפ"ל", ההשלכה העיקרית של תפעולו כגורם תוך-נבדקי היא שהשוואה בין התחומים במידת השיפור עלולה להיות מושפעת מהבדלים בין התחומים ברמה ההתחלתית, הבדלים

שלא ניתן לנכותם כאשר המשתנה מתופעל כגורם תוך-נבדקי. נראה שהמערך האופטימלי לבדיקת שאלות המחקר הוא מערך רנדומלי שלם, שבו כל הגורמים מתופעלים כגורמים בין-נבדקיים, כפי שבוצע במחקרים שצוטטו לעיל<sup>12</sup>. שימוש במערך כזה כרוך, כמובן, באיבוד היתרונות של מערך מדידות חוזרות (שהם חיסכון בנבדקים והקטנת שונות הטעות).

2. בהתייחס לאפקט של סדר מתן התוספת, כפי שצוין בגוף הדוח, האפקט של סדר מתן התוספת לא תופעל במערך הנוכחי באופן שניתן לנתח אותו בניתוח סטטיסטי שכולל את הגורם "תחום מכפ"ל" (וגם לא ניתן היה לחקור את האפקט שלו ביחס לציון הכללי). באופן כללי, גם במצב שבו יש עירוב של שתי השיטות (כלומר, חלק מן הגורמים הם בין-נבדקיים וחלק הם תוך-נבדקיים) יש לוודא שמתקיים מערך פקטוריאלי שמאפשר ניתוח מלא של כל האפקטים<sup>13</sup>.

3. כפי שצוין בפרק השיטה, סדר הפרקים במבחן היה אחיד לכל הנבדקים. משמעות הדבר היא שיתכן ערבוב בין האפקט של סדר לבין האפקט של היבחות בפרק מסוים (תוספת בפרק ראשון בתחום היתה תמיד בפרק מסוים אחד ותוספת בפרק השני בתחום היתה תמיד בפרק האחר). מצב זה היה נמנע, כמובן, לו מתן או אי-מתן תוספת היה מתופעל כגורם בין-נבדקי.

4. ראוי לציון קושי אינהרנטי בחקירת האפקט של תוספת זמן במתכונת המחקר הנוכחי, שמקורו בעובדה שהנבחנים מגיעים עם מוכנות, ברמה כזו או אחרת, למבחן בתנאים הקיימים. ניתן להניח שהוספת זמן בעת המבחן עצמו (גם אם הוכרז עליה בתחילת הפרק) אינה משחזרת באופן מלא את המצב שאותו נכון היה לדמות כדי לבדוק את השפעת תוספת הזמן. ההשלכות של חלקיות השחזור לא נבחנו במחקר זה.

5. נמצאה עדות לכך שתוספת זמן של כ-8 דקות מאפשרת מיצוי של השיפור שניתן להשיג באמצעות תוספת זמן. מאחר שערך זה מצוי היטב בתוך טווח ערכי המשתנה הבלתי-תלוי "גודל תוספת הזמן" שתופעלו במחקר הנוכחי, אין סיבה, לכאורה, להמליץ על הרחבת טווח הערכים של משתנה זה במחקר עתידי. חשוב, עם זאת, להפנות את תשומת הלב לכך שביחס לשניים מן התחומים (מילולי וכמותי), וכן ביחס לציון הכללי, המשיך השיפור לגדול גם במעבר מתוספת זמן של 10 דקות לתוספת זמן של 13 דקות. לאור זאת, ולו גם רק כדי

<sup>12</sup> מערך כזה יכול שני גורמים: גודל תוספת הזמן (כאשר קבוצה אחת תהיה קבוצת ביקורת, שלא תקבל תוספת זמן) ותחום מכפ"ל.

<sup>13</sup> בהינתן סדר הפרקים במבחן במחקר הנוכחי (אנגלית; מילולי; כמותי; אנגלית; מילולי; כמותי) אזי תפעול שבו התוספת היתה ניתנת בשלושת הפרקים הראשונים או בשלושת הפרקים האחרונים (ולא בזוגיים או באי-זוגיים) היה מאפשר בדיקה כזו.



לבסס את מסקנות המחקר הנוכחי, מומלץ להרחיב במחקר עתידי את טווח הערכים של המשתנה "גודל תוספת הזמן".

6. ניתן היה לצפות שבמעבר בין תוספת זמן של 5 דקות לתוספת זמן של 8 דקות תהיה מידת השיפור קטנה מזו שבמעבר בין מצב ללא תוספת זמן למצב עם תוספת זמן של 5 דקות, הן משום שהפרש הזמנים בין שני המצבים (3 דקות) במעבר הראשון קטן מזה שבמעבר השני (5 דקות) והן בשל התופעה הטיפוסית שניתן לכנותה בהקשר הנוכחי "שיפור שולי פוחת". הממצא, המנוגד לאינטואיציה, שהתקבל במחקר הנוכחי ראוי לבדיקה ולחשיבה נוספות.

7. חשוב לסייג את יכולת ההכללה של ממצאי המחקר לגבי אוכלוסיות שלא השתתפו במחקר (נבחנים בעלי לקויות שונות, נבחנים בשפות זרות). ראוי לבדוק מהי ההשפעה של תוספת זמן בבחינה על נבחנים כאלה, ובהתאם לכך, מהן ההשלכות של תוספת זמן על הדירוג של אוכלוסיות שונות של נבחנים.

## רשימת מקורות

אנגל, א' (2008). *השפעת תוספת זמן בבחינה הפסיכומטרית על נבחנים עם ADHD*. עבודת גמר לתואר מוסמך, ירושלים: האוניברסיטה העברית בירושלים.

פבלוב, י' ואורן, כ' (2002). *דוח סטטיסטי רב-שנתי לשנות מבחן 1991-2000*. ירושלים: מרכז ארצי לבחינות ולהערכה.

Bejar, I. I. (1985). *Test speededness under number-right scoring: An analysis of the Test of English as a Foreign Language* (Report No. ETS-RR-85-11). Princeton, NJ: Educational Testing Service.

Bridgeman, B. (2004). *Speededness as a threat to construct validity*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Bridgeman, B., Cline, F., & Hessinger, J. (2004). Effect of extra time on verbal and quantitative GRE scores. *Applied Measurement in Education*, 17, 25-37.

- Bridgeman, B., Trapani, C., & Curley, E. (2004). Impact of fewer questions per section on SAT I scores. *Journal of Educational Measurement, 41*, 291-310.
- Briel, J. B., O'Neill, K. A., & Scheuneman, J. D. (1993). *GRE technical manual*. Princeton, NJ: Educational Testing Service.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2<sup>nd</sup> ed.). Hillsdale, NJ: Erlbaum.
- Donlon, T. F. (Ed.). (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.
- Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice, 26*, 29-37.
- Raghunathan, T. E., Rosenthal, R., & Rubin, D. B. (1996). Comparing correlated but nonoverlapping correlations, *Psychological Methods, 1*, 178-183.
- Schnipke, D. L. (1995). *Assessing speededness in computer-based tests using item response times*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.

# נספח 1

## מספר תצפיות, ממוצעים וסטיות של הציונים במערך הראשוני של המחקר

לוח 4: ממוצעים (וסטיות תקן) של הציונים במערך הראשוני של המחקר

תחום מכפ"ל								n	עיתוי מתן התוספת	גודל התוספת
כללי		אנגלית		כמותי		מילולי				
עם תוספת	ללא תוספת	עם תוספת	ללא תוספת	עם תוספת	ללא תוספת	עם תוספת	ללא תוספת			
616.2 (87.5)	610.0 (86.7)	122.6 (21.3)	123.7 (20.2)	119.3 (17.7)	118.5 (17.8)	120.4 (18.9)	117.7 (17.4)	32	פרקים 5,3,1	5
626.7 (105.6)	616.0 (108.1)	123.8 (20.3)	122.0 (22.4)	122.3 (18.9)	120.4 (19.5)	121.4 (21.9)	119.3 (22.1)	32	פרקים 6,4,2	
632.7 (97.9)	610.7 (100.5)	126.9 (20.2)	119.88 (21.5)	125.0 (18.5)	119.1 (18.6)	119.5 (19.9)	119.5 (22.1)	42	פרקים 5,3,1	8
637.1 (99.2)	603.8 (94.9)	123.8 (23.1)	123.7 (19.7)	124.2 (19.7)	118.3 (18.2)	123.7 (18.3)	115.3 (19.6)	45	פרקים 6,4,2	
617.9 (107.1)	592.1 (99.3)	121.5 (23.6)	116.1 (24.1)	121.8 (19.5)	117.1 (18.3)	119.1 (21.6)	115.2 (20.6)	135	פרקים 5,3,1	10
641.3 (88.0)	607.9 (99.1)	121.3 (21.5)	117.5 (22.8)	125.2 (17.3)	119.3 (18.9)	125.9 (17.7)	119.1 (19.5)	135	פרקים 6,4,2	
631.8 (75.8)	608.1 (79.5)	119.0 (24.5)	118.6 (21.0)	124.3 (16.9)	117.0 (17.4)	124.0 (15.0)	121.6 (16.1)	48	פרקים 5,3,1	13
646.7 (84.3)	602.6 (101.3)	121.3 (24.1)	116.1 (23.6)	124.1 (18.2)	116.2 (19.9)	129.6 (15.5)	120.5 (19.9)	47	פרקים 6,4,2	

## נספח 2

### החלפת המשתנה "עיתוי מתן התוספת" במשתנה "סדר מתן התוספת"

ברור בעליל שבמערך המחקר שתואר לעיל, שבו בתוך נבדק מופעלים שני התנאים - מתן תוספת זמן ואי-מתן תוספת זמן - מתעוררת שאלה בדבר אפקט אפשרי לסדר מתן התוספת, כלומר לעובדה שהתנאי של מתן תוספת זמן **קדם** לתנאי של אי-מתן תוספת, או להיפך. ומאחר ושאלות המחקר מותנות במשתנה הבלתי-תלוי "תחום מכפ"ל", אזי בדיקת האפקט של סדר מתן התוספת צריך להתקיים בתוך תחום מכפ"ל. במלים אחרות: המשתנה הרלוונטי הוא המשתנה "סדר מתן התוספת" (שמוגדר בתוך תחום, שכן ערכו מציין באיזה פרק **מבין שני פרקי התחום** – הראשון או השני – ניתנה התוספת), שערכיו עבור כל נבדק צריכים להיגזר מן המשתנה "עיתוי מתן התוספת".

הבעיה היא, שלא ניתן לאפיין נבדק על-פי המשתנה "סדר מתן התוספת"<sup>14</sup>, כפי שנדרש ממשתנה בין-נבדקי, שכן כאשר התוספת ניתנה בפרקים האי-זוגיים אזי **עבור אותו נבדק** היא ניתנה בשניים מן התחומים (אנגלית וכמותי) בפרק הראשון מבין שני פרקי התחום, ואילו בתחום השלישי (מילולי) היא ניתנה בפרק השני מבין שני פרקי התחום, ולהיפך כאשר התוספת ניתנה בפרקים הזוגיים.

תפעול כזה של סדר מתן התוספת אינו מאפשר לבדוק את אפקט הסדר במערך פקטוריאלי עם המשתנה "תחום מכפ"ל" (וגם לא ביחס לציון הכללי). לאור האילוצים שתוארו, נבחן אפקט הסדר ביחס לכל תחום בנפרד.

### האפקט של סדר מתן התוספת (במודל הכולל את גודל התוספת) על ההפרש בציון

בכל אחד משלושת תחומי מכפ"ל בוצע ניתוח שונות דו-גורמי על ההפרש<sup>15</sup> בציון. המשתנים הבלתי-תלויים היו גודל תוספת הזמן וסדר מתן התוספת. התוצאות מוצגות בלוח 5. הדיון בממצאי הלוח יתמקד באפקט הסדר (ובאינטראקציה בינו לבין גודל התוספת) בלבד.

<sup>14</sup> זאת, בניגוד למשתנה המקורי "עיתוי מתן התוספת" (האם התוספת ניתנה בפרקים האי-זוגיים או הזוגיים), שהוא משתנה בין-נבדקי "תקני".

<sup>15</sup> בבדיקה מקדימה (ניתוח שונות דו-גורמי שבוצע בכל אחד משלושת תחומי מכפ"ל על הציון ללא תוספת זמן כפונקציה של גודל תוספת הזמן וסדר מתן התוספת) לא נמצא קשר מובהק סטטיסטית בין גודל התוספת או הסדר לבין ה"רמה ההתחלתית" (הציון ללא תוספת זמן). לכן האפקט של גודל התוספת והסדר על השינוי בציון נבדק ביחס להפרש (ולא להפרש המנוכה מן הציון ללא תוספת זמן) בין הציון עם תוספת לציון בלי תוספת.

לוח 5: ממוצעים וסטיות תקן ומבחני שונות על ההפרש בציון בשלושת תחומי מכפ"ל לפי

גודל תוספת הזמן וסדר מתן התוספת

גודלXסדר F(3,508)	סדר F(1,508)	גודל F(3,508)	סדר מתן התוספת						גודל	תחום
			בפרק שני			בפרק ראשון				
			ס"ת	ממוצע	n	ס"ת	ממוצע	n		
2.29	12.61*	1.23	11.0	2.7	32	10.6	2.2	32	5	מילולי
			12.4	0.0	42	10.1	8.5	45	8	
			12.2	3.9	135	11.9	6.8	135	10	
			13.4	2.4	48	14.5	9.1	47	13	
0.07	0.39	3.84**	11.2	1.8	32	9.6	0.8	32	5	כמותי
			10.8	6.0	45	10.3	5.9	42	8	
			12.9	5.9	135	12.4	4.6	135	10	
			10.3	7.9	47	11.3	7.4	48	13	
5.34**	0.03	2.99***	11.0	1.8	32	11.5	-1.1	32	5	אנגלית
			9.8	0.0	45	8.7	7.0	42	8	
			11.3	3.9	135	10.4	5.4	135	10	
			11.7	5.2	47	10.6	0.4	48	13	

$p < .05$  \*\*\*  $p < .01$  \*\*  $p < .001$  \*

עיון בממצאי הלוח מעלה שהאפקט של סדר מתן התוספת מובהק סטטיסטית רק בתחום המילולי (וכיוון הממצאים הוא שתוספת זמן בפרק הראשון מבין שני הפרקים בתחום, משפרת את הציון יותר מאשר תוספת דומה בפרק השני). בתחום הכמותי אין אפקט מובהק לגורם זה, ואילו באנגלית נמצאה אינטראקציה מובהקת בין סדר מתן התוספת לבין גודל התוספת: בחלק מתנאי גודל התוספת (8 דקות ו-10 דקות) נמצאה תופעה דומה בכיוונה למה שנמצא בתחום המילולי ובתנאים האחרים נמצאה תופעה הפוכה. ניתן להתרשם, אם-כן, שלסדר מתן התוספת אין אפקט עקבי מעבר לתחומים.

### נספח 3

#### השוואה בין ציוני המחקר הנוכחי לציוני הבחינה הפסיכומטרית האמיתית<sup>16</sup>

לוח 6א: ממוצעים (וסטיות תקן) של הציונים בתחומי הבחינה במחקר הנוכחי ובבחינה האמיתית

	בבחינה האמיתית		במחקר הנוכחי	
	פרק ראשון	פרק שני	פרק עם תוספת	פרק ללא תוספת
מילולי	118.5 (18.5)	118.5 (18.7)	123.2 (19.0)	118.3 (19.8)
כמותי	119.9 (18.2)	119.8 (18.0)	123.8 (18.2)	118.5 (18.5)
אנגלית	119.6 (21.4)	120.0 (22.5)	122.4 (22.4)	118.9 (22.4)

לוח 6ב: מתאמי פירסון בין הציונים במחקר הנוכחי ובבחינה האמיתית לפי תחומים

	בבחינה האמיתית		במחקר הנוכחי	
	פרק ראשון	פרק שני		
מילולי	0.81	0.80	0.80	פרק ללא תוספת
	0.80	0.79		פרק עם תוספת
	0.81			פרק ראשון
כמותי	0.79	0.80	0.79	פרק ללא תוספת
	0.80	0.80		פרק עם תוספת
	0.83			פרק ראשון
אנגלית	0.89	0.87	0.88	פרק ללא תוספת
	0.88	0.88		פרק עם תוספת
	0.90			פרק ראשון
מכפ"ל <sup>1</sup>	0.89	0.90	0.91	פרק ללא תוספת
	0.89	0.90		פרק עם תוספת
	0.93			פרק ראשון

<sup>1</sup> ציון מכפ"ל בבחינה האמיתית חושב כממוצע הציונים של שלושת הפרקים הראשונים ("פרק ראשון") או של שלושת הפרקים השניים ("פרק שני") בתחומים.

<sup>16</sup> הנתונים מבוססים על 505 נבחנים שנמצאו עבורם ציונים בבחינה האמיתית.

## נספח 4

האם הרמה ה"התחלתית" (ציון בלי תוספת זמן) דומה בצירופים שונים של תחום מכפ"ל וגודל תוספת הזמן?

ניתוח שונות דו-גורמי עם מדידות חוזרות (בתחומים מילולי, כמותי ואנגלית) בין ארבע רמות של גודל תוספת הזמן בוצע על הציון ללא תוספת זמן. התוצאות מוצגות בלוח 7.

לוח 7: ממוצעים וסטיות תקן ומבחן שונות על הציון ללא תוספת זמן, לפי גודל תוספת הזמן, בשלושת תחומי מכפ"ל

גודל תחום F(6,1024)	תחום F(2,1024)	גודל F(3,512)	תחום						n	גודל התוספת
			אנגלית		כמותי		מילולי			
			ממוצע	ס"ת	ממוצע	ס"ת	ממוצע	ס"ת		
2.80*	1.40	0.63	122.8	21.2	119.5	18.5	118.5	19.7	64	5
			121.9	20.6	118.7	18.3	117.3	20.8	87	8
			116.8	23.4	118.2	18.6	117.2	20.1	270	10
			117.4	22.3	116.6	18.6	121.0	17.9	95	13

\*  $p < .05$

ניתן להיווכח שלא נמצא אפקט מובהק לגודל תוספת הזמן או לתחום מכפ"ל על הציון ללא תוספת זמן, אך האינטראקציה בין שני המשתנים נמצאה מובהקת סטטיסטית. פירוש הדבר שאין הבדלים עקביים בין התחומים (מעבר לרמות השונות של גודל התוספת) או בין הרמות השונות של גודל התוספת (מעבר לתחומים) בציון ללא תוספת זמן, אך יש הבדלים בציון זה בין צירופים מסוימים של גודל התוספת ותחום מכפ"ל.

## נספח 5

### ממצאי המחקר ביחס לציון הכללי במכפ"ל

שאלת המחקר הראשונה

האם תוספת זמן (בגדלים שונים) משפרת את הציון הכללי במכפ"ל?

בלוח 8 מוצגים הממוצעים וסטיות התקן של הציון עם תוספת זמן והציון ללא תוספת זמן וכן תוצאות מבחני t שבוצעו להשוואה בין שני הציונים (על-ידי בדיקת ההשערה שתוחלת הציון עם תוספת זמן גבוהה מתוחלת הציון ללא תוספת זמן) ברמות שונות של גודל תוספת הזמן.

לוח 8: ממוצעים (וסטיות תקן) של הציון ללא תוספת זמן והציון עם תוספת זמן ותוצאות מבחני t עבור הציון הכללי במכפ"ל

t	עם תוספת	ללא תוספת	n	גודל התוספת
1.84*	621.4 (96.4)	613.0 (97.3)	64	5
7.17**	635.0 (98.0)	607.1 (97.1)	87	8
11.54**	629.6 (98.5)	600.0 (99.3)	270	10
7.23**	639.2 (80.0)	605.4 (90.5)	95	13

$p < .0001$  \*\*       $p < .05$  \*

ניתן להיווכח שתוספת זמן, בכל הגדלים שנבדקו, משפרת את הציון הכללי במכפ"ל. נשאלת השאלה אם יש הבדל בשיפור כפונקציה של גודל התוספת.



## האפקט של גודל תוספת הזמן על השיפור בציון הכללי במכפ"ל

בניתוח שונות חד-גורמי על ההפרש<sup>17</sup> בציון הכללי במכפ"ל נמצא שהאפקט של גודל תוספת הזמן מובהק סטטיסטית ( $F[3,512] = 5.52, p < .001$ ). שלוש השערות מתוכננות נוסחו ביחס למקור ההבדלים בין ארבעת התנאים של גודל התוספת: (1) תוספת זמן של 8 דקות משפרת את הציון יותר מתוספת זמן של 5 דקות; (2) תוספת זמן של 10 דקות משפרת את הציון יותר מתוספת זמן של 8 דקות; ו-(3) תוספת זמן של 13 דקות משפרת את הציון יותר מתוספת זמן של 10 דקות. רק ההבדל בין התנאי של תוספת זמן של 5 דקות לבין התנאי של תוספת זמן של 8 דקות נמצא מובהק סטטיסטית ( $F[1,512] = 8.15, p < .01$ ).

### *שאלת המחקר השניה*

#### האם תוספת זמן משנה את מדרג הנבחנים עפ"י הציון הכללי במכפ"ל?

מתאמי פירסון בין ציוני הנבחנים במכפ"ל עם ובלי תוספת זמן היו: 0.93, 0.93, 0.91 ו-0.86. עבור תוספת זמן של 5, 8, 10 ו-13 דקות בהתאמה. לשם השוואה, המתאם שמתקבל בין הציונים שחושבו על-פי שני חצאים של הבחינה הפסיכומטרית האמיתית שבה נבחנו 505 ממשותפי המחקר הוא 0.93. המתאם בין החצאים בתוספות הזמן הקטנות היה זהה, אם כן, למתאם שמתקבל בין שני חצאים בבחינה רגילה. לעומת זאת, המתאם בין החצאים בתוספות הזמן הגדולות היה נמוך באופן מובהק ( $z = -2.21, p < .05$ ) עבור תוספת זמן של 10 דקות ו- $z = -3.33, p < .01$  עבור תוספת זמן של 13 דקות) מן המתאם בין שני חצאים בבחינה רגילה. ניתן להתרשם, אם כן, שביחס לציון הכללי, בתוספות זמן גדולות יש חוסר יציבות במדרג. נשאלת השאלה אם מידת חוסר היציבות תלויה בגודל התוספת.

#### האפקט של גודל תוספת הזמן על השינוי במדרג עפ"י הציון הכללי במכפ"ל

נמצא הבדל מובהק ( $z = 1.75, p < .05$ ) באופן גבולי בין המתאם בין חצאים בתוספת זמן של 13 דקות למתאם בתוספת זמן של 10 דקות. ניתן לסכם, אם כן, שביחס לציון הכללי, בתוספות זמן גדולות, יש אינדיקציה לכך שככל שתוספת הזמן גדולה יותר, חוסר היציבות במדרג גדול יותר.

<sup>17</sup> בבדיקה מקדימה (ניתוח שונות חד-גורמי על הציון הכללי במכפ"ל ללא תוספת זמן כפונקציה של גודל התוספת) לא נמצא קשר מובהק סטטיסטית בין גודל תוספת הזמן לבין ה"רמה ההתחלתית" (הציון ללא תוספת זמן). לכן האפקט של גודל תוספת הזמן על השיפור בציון נבדק ביחס להפרש (ולא להפרש המנוכה) בין הציון עם תוספת לציון בלי תוספת.

# **The Effect of Extended Time on Scores and Examinee Rank Order for Three PET Domains**

**Tami Kennet-Cohen, Shmuel Bronner and Miriam Eitan**

## **Abstract**

The effects of extended time on performance in each of the three domains of the Psychometric Entrance Test (PET) were examined. The data were based on the scores of 516 participants, sampled from the pool of PET registrants (provided that they asked to be tested in Hebrew and requested no special accommodations). The participants took an experimental test consisting of six sections, two in each of the three domains. One of the two sections was allotted the standard administration time (25 minutes) and the other was allotted extended time (an addition of 5, 8, 10 or 13 minutes). The participants were randomly assigned to one of the four conditions of extended time. Results indicated that extending the time led to a small increase in domain scores, with the magnitude of the increase depending on both the degree of the extension and the domain of PET. It was found that extending the time by 8 minutes led to a significantly higher increase in scores compared to the condition where time was extended by 5 minutes (4.7 points versus 1.5 points respectively on a 50-150 scale). No significant differences were found between the other consecutive levels of time extension. A significant difference was found between improvement in the quantitative domain (5.3 points) and in the English domain (3.5 points). Improvement in the verbal domain was a bit less than in the quantitative domain. Results with respect to the effect of time extension on the total score on PET were similar to those reported above: extending the time by 8 minutes led to a significantly higher increase in scores compared to extending the time by 5 minutes (28 points versus 8 points respectively on a 200-800 scale), and no significant differences were found between the other consecutive levels of time extension.

A comparison between the relative rankings of examinees in sections with and without extended time revealed that, generally, extended time did not change the relative rank order of examinees. However, some indications of instability in rank order were found in the larger time increments.

The results are discussed in reference to findings from similar studies.

Comments regarding the methodological limitations of the study are presented and suggestions for future research are proposed.