

השלכות השימוש במכפ"ל מרבי

תמר קנת-כהן
שמואל ברונר
כרמל אורן

אוקטובר 2010



דוח מרכז 368
ISBN:978-965-502-157-8

© All rights reserved
NITE
P.O.B. 26015 Jerusalem

© כל הזכויות שמורות
מרכז ארצי לבחינות ולהערכה
ת"ד 26015 ירושלים 91260

תוכן

2	תקציר
3	מבוא
5	שיטה
8	תוצאות
13	סיכום ודין
15	מקורות
18	נספחים
25	תמצית באנגלית

תקציר

מטרת המחקר הייתה לבחון את השלכות השימוש, לצורך קבלה ללימודים, בציון מכפ"ל (מבחן הכניסה הפסיכומטרי לאוניברסיטאות) שמחושב על פי הציונים הגבוהים ביותר בשלושת תחומי המבחן, גם אם ציונים אלו התקבלו מהיבחנות במועדים שונים (להלן "מכפ"ל מרבי"). השימוש בציון זה הושווה למדיניות הנוכחית, של חישוב ציון כללי בכל מועד היבחנות ושימוש בציון הכללי הגבוה ביותר ("מכפ"ל בפועל"). כדי לספק הקשר להשוואה בין שתי החלופות נבדקה גם האפשרות של שימוש בציון מכפ"ל אשר התקבל במועד ההיבחנות הראשון ("מכפ"ל ראשון").

ההשלכות של שינוי מדיניות הקבלה ממכפ"ל בפועל למכפ"ל מרבי התייחסו לכמה (נבחנים ונקודות בציון) ומי (על פי משתני רקע נבחרים) מרוויחים מן השינוי. נושאים אלה (מחקר א') נבדקו בקרב 67,806 נבחנים במכפ"ל בשנת 2008. ההשלכות שנבדקו נוגעות גם לשינוי הצפוי בתוקף הניבוי של מכפ"ל ומרכיביו וכן של ציון הסכם. נושא זה (מחקר ב') נבדק בקרב 57,632 רשומות של תלמידי שנה א' במחוזות תשס"ה, תשס"ו ותשס"ז באוניברסיטאות בישראל.

37% מן הנבחנים שנכללו במחקר א' נבחנו יותר מפעם אחת. מקרב הנבחנים החוזרים 58% היו משפרים את ציונם בשיטת מכפ"ל מרבי יחסית לציונם כיום. השיפור בציון במעבר ממכפ"ל בפועל למכפ"ל מרבי (בקרב המשפרים ציון) הוא כ-13 נקודות בממוצע. לשם השוואה, השיפור במעבר ממכפ"ל ראשון למכפ"ל בפועל בקבוצה זו הוא כ-42 נקודות בממוצע. ניתן להתרשם שהאפקט על הציון של מעבר ממכפ"ל בפועל למכפ"ל מרבי איננו מבוטל. מבין תחומי מכפ"ל, אחוז המשפרים וגודל השיפור היו הגבוהים ביותר באנגלית. בבדיקת האפקט של מעבר ממכפ"ל בפועל למכפ"ל מרבי על פי משתני רקע אחדים, נמצא שהמשתנה שפת היבחנות קשור ביותר לגודל השיפור, כאשר השיפור בקרב נבחני ערבית היה גדול מן השיפור בקרב נבחני עברית (6 נקודות לעומת 2 נקודות בממוצע בהתאמה).

בבדיקת תוקף הניבוי מעבר לכל הסטודנטים (נבחני פעם אחת, נבחנים חוזרים שאינם משפרים ציון במכפ"ל מרבי ונבחנים חוזרים שכן משפרים ציון במכפ"ל מרבי) נמצאה (מחקר ב') ירידה מזערית בתוקף הניבוי של ציון הסכם ומרכיביו במעבר מציון בפועל לציון מרבי. בקבוצת הנבחנים החוזרים שמשפרים במכפ"ל מרבי הצטמצם מעט תוקף מכפ"ל במעבר לציון מרבי. תוקף ציון הסכם לא השתנה במעבר זה. נראה, אם כן, שהירידה (המזערית) בתוקף מעבר לכל הסטודנטים במעבר מציון בפועל לציון מרבי נובעת משני מקורות: האחד הוא ירידה בתוקף החזאי בתוך קבוצת המשפרים במכפ"ל מרבי, והשני הוא צמצום שונות רלוונטית לקריטריון בין הקבוצות.

הבדיקה שתוארה נעשתה בהנחה ששינוי במדיניות הקבלה לא ישנה את החלטות הנבחנים להיבחן. הסקת מסקנות ממחקר זה צריכה להביא עובדה זו בחשבון.

מבוא

בתהליך הקבלה ללימודי התואר הראשון בישראל מקובל להביא בחשבון, עבור מועמדים שנבחנו במכפ"ל יותר מפעם אחת, את הציון הכללי הגבוה ביותר. ציון זה מתבסס על ממוצע משוקלל של הציונים שהתקבלו בתחומי המבחן במועד היבחנות מסוים. בארה"ב יש מגוון של גישות לשימוש בציוני נבחנים חוזרים. בהתייחס ל-SAT, הגישה הרווחת (בקרב המוסדות שמדווחים על מדיניותם) היא להתחשב בסכום של הציונים הגבוהים ביותר שהתקבלו בשלושת חלקי המבחן, גם אם הם התקבלו במועדי היבחנות שונים. כך, למשל, על פי מסמך שמפרסמת מועצת המכללות (CollegeBoard, 2010) בארה"ב, 50% מתוך 1,447 המוסדות שמדיניות הקבלה שלהם מוצגת במסמך משתמשים בסכום הציונים הגבוהים ביותר בחלקים ("highest section"). רק 12% משתמשים בציון הכללי הגבוה ביותר ("highest sitting"). שימוש בסכום הציונים הגבוהים ביותר בחלקים הוא הנוהל המקובל ביותר גם בקרב מוסדות סלקטיביים (Vigdor & Clotfelter, 2003).

מחקרים שבדקו את השלכות השימוש בציונים קובעים שונים עסקו הן בתוקף הניבוי של מערכת המיון והן בהוגנותה. תוקף הניבוי נבדק בקרב נבחנים חוזרים וכן מעבר לכל המועמדים – נבחנים חוזרים ונבחני פעם אחת. שאלת ההוגנות נבדקה הן מן ההיבט של תוקף דיפרנציאלי (האם התוקף בקרב נבחני פעם אחת דומה לתוקף בקרב נבחנים חוזרים) והן מן ההיבט של ניבוי דיפרנציאלי (האם לציון קבלה נתון יש משמעות ניבויית – במובן של הציון הצפוי בלימודים – דומה בקרב נבחני פעם אחת ונבחנים חוזרים). מחקרים שונים התמקדו בציונים קובעים שונים (ראשון, אחרון, הפרש בין ציון ראשון לאחרון, ממוצע פשוט, ממוצע משוקלל, הציון הכללי הגבוה ביותר וסכום הציונים הגבוהים ביותר בחלקים). סקירה נרחבת ומפורטת של עבודות שנעשו בתחום זה מופיעה אצל בולט, סנטרה וקורטני (Boldt, Centra, & Courtney, 1986). חלק מן העבודות שנסקרו (Cowen & Abrahams, 1982; Olsen & Schrader, 1958, 1959a, 1959b) התייחסו לציון הגבוה ("highest score") אך לא ברור אם הכוונה לציון הכללי הגבוה ביותר או לציונים הגבוהים ביותר בחלקים. חלק אחר מן העבודות שנסקרו (Pitcher, 1965, 1976, 1977) עסקו במבחן הכניסה לבית הספר למשפטים (LSAT) ולא התייחסו כלל לציון הגבוה (הציונים שנבדקו הם ראשון, אחרון וממוצע). עבודות מאוחרות יותר שעסקו בציוני LSAT מרובים (Dalessandro & McLeod, 1999; Stilwell, Thornton, & Pashley, 2002; Thornton, Stilwell, & Reese, 2006; Wightman, 1990) התייחסו גם לציון הכללי הגבוה ביותר, אך לא לסכום הציונים הגבוהים ביותר בחלקים (הציונים שנבדקו הם ציון ראשון, ציון אחרון, הציון הכללי הגבוה ביותר והציון

הממוצע). מחקרים שבוצעו בעבר במרכז הארצי לבחינות ולהערכה (אורן וכהן, 1988; קנת-כהן ואורן, 1989) לא בחנו אף הם את סכום הציונים הגבוהים ביותר בחלקים (הציונים שנבדקו הם ציון ראשון, ציון שני, הציון הכללי הגבוה ביותר, הציון הממוצע והציון הנמוך). בולט, סנטרה וקורטני (Boldt et al., 1986) עצמם כן בחנו את השימוש בסכום הציונים הגבוהים ביותר בחלקים. הציונים הקובעים הנוספים שנבדקו היו הציון הכללי הגבוה ביותר, הציון האחרון או הציון הממוצע. מחקרים התבססו על מדגם של 51,000 סטודנטים (מתוכם 25,770 נבחנו פעמיים ו-3,561 נבחנו שלוש פעמים) מ-87 קולג'ים. בבדיקת תוקף הניבוי בקרב נבחנים חוזרים נמצא שמקדמי התוקף הממוצעים בקרב נבחני פעמיים היו .38, .38, .37 ו-39. עבור סכום הציונים הגבוהים ביותר בחלקים, הציון הכללי הגבוה ביותר, הציון האחרון והציון הממוצע בהתאמה. מקדמי התוקף הממוצעים שהתקבלו בקרב נבחני שלוש פעמים היו .25, .26, .25 ו-27. בהתאמה. ניתן להיווכח שההבדלים בתקפיות הניבוי של החלופות השונות, ובפרט ההבדל בין תוקף הניבוי של סכום הציונים הגבוהים ביותר בחלקים לבין תוקף הניבוי של הציון הכללי הגבוה ביותר, הוא מזערי (ממצא עקבי מעבר למחקרים בתחום). בולט, סנטרה וקורטני בחנו גם את ההיבט של ניבוי דיפרנציאלי, ומצאו שניבוי ציוני הקריטריון של הנבחנים החוזרים באמצעות משוואת הרגרסיה של נבחני פעם אחת היה ניבוי חסר בכל ארבע הגישות להגדרת ציונים קובעים. ניבוי החסר המינימאלי נגרם בשימוש בסכום הציונים הגבוהים ביותר בחלקים וניבוי החסר המקסימלי נגרם בשימוש בציון הממוצע. החוקרים מסכמים שמשיקולי תוקף ניבוי נמצא יתרון מסוים לשימוש בציון הממוצע (ממצא עקבי מעבר למחקרים בתחום) ואילו משיקולי הוגנות יש עדיפות לשימוש בסכום הציונים הגבוהים ביותר בחלקים.

מטרת המחקר הנוכחי הייתה לבחון את ההשלכות, ביחס לציון הקבלה, של מעבר מן המדיניות הנוכחית, של חישוב ציון כללי במכפ"ל בכל מועד היבחנות ושימוש בציון הכללי הגבוה ביותר ("מכפ"ל בפועל"), למדיניות שמשתמשת בציון כללי שמחושב על פי הציונים הגבוהים ביותר בשלושת תחומי המבחן – חשיבה מילולית, חשיבה כמותית ואנגלית – גם אם ציונים אלה התקבלו מהיבחנות במועדים שונים ("מכפ"ל מרבי"). כדי לספק הקשר להשוואה בין החלופות נבדקה גם האפשרות של שימוש בציון מכפ"ל אשר התקבל במועד ההיבחנות הראשון ("מכפ"ל ראשון").

ההשלכות של שינוי מדיניות הקבלה ממכפ"ל בפועל למכפ"ל מרבי התייחסו לכמה (נבחנים ונקודות בציון) ומי (על פי משתני רקע נבחרים) מרוויחים מן השינוי. נושאים אלה (מחקר א') נבדקו בקרב 67,806 נבחנים במכפ"ל בשנת 2008. ההשלכות שנבדקו נוגעות גם לשינוי הצפוי בתוקף הניבוי של מכפ"ל ומרכיביו וכן של ציון הסכם. נושא זה (מחקר ב') נבדק בקרב 57,632 רשומות של תלמידי שנה א' במחוזות תשס"ה, תשס"ו ותשס"ז באוניברסיטאות בישראל.

שיטה

מחקר א'

מדגם

המדגם כלל 67,806 נבחנים במכפ"ל בשנת 2008, שכל היבחנותיהם היו אחרי 1.1.1992.

משתני המחקר

היבחנות חוזרת ושיפור:

הנבחנים סווגו על פי הצירוף של סטטוס היבחנות חוזרת (האם נבחן יותר מפעם אחת) ושיפור הציון במכפ"ל מרבי (האם הציון במכפ"ל מרבי גבוה מהציון במכפ"ל בפועל). בהתאם לכך הוגדרו שלוש קטגוריות (קבוצות) במשתנה היבחנות חוזרת ושיפור: נבחני פעם אחת, נבחנים חוזרים שציונם במכפ"ל מרבי זהה לציונם במכפ"ל בפועל ("חוזרים שלא משפרים"), ונבחנים חוזרים שציונם במכפ"ל מרבי גבוה מציונם במכפ"ל בפועל ("חוזרים שמשפרים").

ציוני מכפ"ל:

לכל נבחן חושבו, על בסיס כל היבחנותיו שלפני 31.12.2008, הציונים הבאים:

1. ציונים בתחומי מכפ"ל:

עבור כל אחד מתחומי מכפ"ל – חשיבה מילולית חשיבה כמותית ואנגלית – הוגדרו שלושה ציונים:

מילולי/כמותי/אנגלית ראשון – הציון הראשון בתחום

מילולי/כמותי/אנגלית בפועל – הציון בתחום בהיבחנות שבה התקבל הציון הכללי הגבוה ביותר

מילולי/כמותי/אנגלית מרבי – הציון הגבוה ביותר בתחום מבין כל ההיבחנות

2. ציוני מכפ"ל כללי:

עבור הציון הכללי הוגדרו שלושה ציונים:

מכפ"ל ראשון – הציון הכללי בהיבחנות הראשונה

מכפ"ל בפועל – הציון הכללי הגבוה ביותר מבין כל ההיבחנות

מכפ"ל מרבי – ממוצע משוקלל של מילולי מרבי, כמותי מרבי ואנגלית מרבי (במשקלות של 0.4, 0.4 ו-0.2 בהתאמה). ממוצע זה הועבר לסולם הציון הכללי על פי הפרמטרים המקובלים (סטולר, 2010).

משתני רקע:

לכל נבחן הוגדרו המאפיינים הבאים¹:

1. גיל: צעיר (עד 20) או מבוגר (מעל 20)

2. מגדר

3. השכלת הורים: נמוכה (אף לא לאחד מההורים יש תואר ראשון) או גבוהה (לפחות לאחד

מההורים יש תואר ראשון)

4. שפת היבחנות במכפ"ל: ערבית או עברית

עיבוד הנתונים

חושבו ממוצעים וסטיות תקן של ציוני מכפ"ל על פי היבחנות חוזרת ושיפור ועל פי משתני רקע.

מחקר ב'

מדגם

המדגם התבסס על רשומותיהם של 46,512 תלמידי שנה א' במחזורים תשס"ה (2004/05),

תשס"ו (2005/06) ותשס"ז (2006/07) באוניברסיטאות בישראל, שכל היבחנותיהם במכפ"ל היו אחרי 1.1.1992. רשומה מוגדרת כנתוני סטודנט ביחידת עיבוד, כאשר יחידת עיבוד היא צירוף של חוג לימוד, מוסד ומחזור. בסך הכול נכללו במחקר 57,632 רשומות (תצפיות) מ-661 יחידות עיבוד.

סטודנט נכלל במחקר אם היו ברשותנו ציונים לגיטימיים² שלו בקריטריון ובכל החזאים (יפורטו בהמשך).

חוג נכלל במחקר אם למדו בו לפחות 20 סטודנטים. לא נכללו מסלולים למצטיינים או לימודים כלליים.

משתני המחקר

היבחנות חוזרת ושיפור (הוגדר כמו שתואר במחקר א')

קריטריון:

ממוצע ציוני שנה א' בלימודי הבוגר

¹ לנבחנים מרובי רשומות נבחר המידע המוקדם ביותר; החלוקה לקטגוריות במשתנים בהם הייתה לנו שליטה על כך (גיל, השכלת הורים) התבססה על שאיפה להגדיר קטגוריות שוות גודל.
² בקריטריון: ציון בטווח של 10-100, המתבסס על 8 נקודות זכות לפחות. בבגרות: ציון בטווח של 10-130.

חזאים:

1. ציונים בתחומי מכפ"ל (הוגדרו כמו שתואר במחקר א')³:

מילולי/כמותי/אנגלית ראשון

מילולי/כמותי/אנגלית בפועל

מילולי/כמותי/אנגלית מרבי

2. ציוני מכפ"ל כללי (הוגדרו כמו שתואר במחקר א')³:

מכפ"ל ראשון

מכפ"ל בפועל

מכפ"ל מרבי

3. ממוצע ציוני תעודת הבגרות (בגרות), אשר חושב ע"י האוניברסיטאות (כולל במוסעים)

4. ציוני סכם:

הוגדרו שלושה ציוני סכם – סכם ראשון, סכם בפועל וסכם מרבי – אשר חושבו כצירוף במשקלות שווים של מכפ"ל ראשון, בפועל או מרבי עם בגרות באופן הבא:
מכפ"ל הרלוונטי ובגרות תוקננו במוסד ובמחזור (על בסיס רשומה יחידה לכל סטודנט במוסד ומחזור) וצורפו במשקלות שווים. הציון שהתקבל הועבר לסולם עם ממוצע 50 וסטיית תקן 10 בקרב 57,632 התצפיות.

משתנים מתנים (moderators):

1. תחום לימוד, שהוגדר על-פי פקולטות באופן הבא:

מילולי – מדעי הרוח והאומנויות, חוגים מילוליים ממדעי החברה, משפטים

כמותי – חוגים כמותיים ממדעי החברה, טבע, הנדסה, רפואה ועזר-רפואה

2. סלקטיביות, שהוגדרה לפי ממוצע מכפ"ל (בפועל) בחוג ומוסד באופן הבא:

חושב ממוצע מכפ"ל לפי חוג לימוד ומוסד (מעבר למחזורים). חוגים שהממוצע שלהם היה מעל לחציון הממוצעים אופיינו כבעלי סלקטיביות גבוהה וחוגים שהממוצע שלהם היה מתחת לחציון הממוצעים אופיינו כבעלי סלקטיביות נמוכה. האפיון המתאים (שחושב, כאמור, מעבר למחזורים) הוצמד לכל אחת מיחידות העיבוד לפי חוג לימוד ומוסד.

עיבוד הנתונים

בכל אחת מיחידות העיבוד חושבו הסטטיסטים הבאים:

1. ממוצעים וסטיות תקן של הקריטריון והחזאים.

2. מקדמי מתאם פירסון, מתוקנים לקיצוץ תחום (מקדמי תוקף ניבוי), בין כל אחד מהחזאים לבין הקריטריון. התיקון לקיצוץ תחום בוצע עבור בררה חד-משתנית (על-פי ציון קבלה),

³ על בסיס ההיבחנות שקדמו לתחילת הלימודים.

במקרה התלת-משתני (Gulliksen, 1950). תיקון זה מניח שידועה סטיית התקן של ציון הקבלה⁴ ברמת המועמדים. כאומד לסטיית התקן של ציון הקבלה בקרב המועמדים שימש שורש הממוצע המשוקלל של שונות ציון הסכם בקרב המועמדים לכל אחד מחוגי הלימוד במחזורים תשנ"ב ותשנ"ג (8.5; לשם השוואה, הערך שנמצא למדד זה ביחידות העיבוד של הלומדים הוא 6.1). מקדמי תוקף הניבוי חושבו מעבר לכל הסטודנטים וכן לפי היבחנות חוזרת ושיפור. בעיבוד האחרון נכללו רק יחידות עיבוד שבהן יש לפחות 5 סטודנטים בקבוצה הנדונה.

בפרק התוצאות מוצגים ממוצעים משוקללים של הסטטיסטיים שלעיל מעבר ליחידות העיבוד.

תוצאות

מחקר א'

מטרתו של מחקר זה הייתה, כאמור, לבחון את שיעור המועמדים שציוניהם ישתפרו בעקבות שינוי מדיניות הקבלה משימוש במכפ"ל בפועל לשימוש במכפ"ל מרבי ואת גודל השיפור הצפוי. כמו כן, נבדקה השאלה אילו קבוצות באוכלוסייה (שהוגדרו על פי משתני רקע נבחרים) ירוויחו משינוי זה. מחקר זה בוצע על 67,806 נבחנים במכפ"ל בשנת 2008. 63% מקרבם היו נבחני פעם אחת. נבחנים חוזרים שמשפרים במכפ"ל מרבי היוו 21% מן המדגם (ו-58% מן הנבחנים החוזרים) ונבחנים חוזרים שלא משפרים במכפ"ל מרבי היוו 16% מן המדגם (ו-42% מן הנבחנים החוזרים). מספר ההיבחנויות הממוצע בקרב ה"משפרים" היה 2.8 (עם סטיית תקן של 1.2) לעומת 2.3 (עם סטיית תקן של 0.7) בקרב ה"לא משפרים". ממצא זה מעיד על קשר חיובי בין מספר ההיבחנויות לבין שיפור הציון בשיטת מכפ"ל מרבי. בלוח 1 מוצגים ממוצעים וסטיות תקן של ציוני מכפ"ל בשלוש הקבוצות.

⁴ שהוא ציון סכם המחושב (על פי נתוני המועמדים למחזורים תשנ"ב ותשנ"ג) כצירוף של מכפ"ל ובגרות במשקלות שווים בקרב המועמדים למחזור בכל מוסד (תיאור מפורט של הגדרת ציון הסכם מצוי אצל קנת-כהן, ברונר ואורן, 1998). ציון סכם זה אינו זהה לסכם בפועל שחושב ברמת הלומדים, אך המתאם בין השניים גבוה (0.96).

לוח 1: ממוצעים (וסטיות תקן) של ציוני מכפ"ל לפי היבחנות חוזרת ושיפור

נבחנים חוזרים					נבחני פעם אחת (N=42,714)	
משפרים (N=14,503)			לא משפרים (N=10,589)			
ציון ראשון	ציון בפועל	ציון ראשון	ציון בפועל	ציון ראשון	ציון ראשון	
542.7 (109.4)	529.7 (110.0)	488.2 (101.4)	583.1 (100.6)	507.6 (92.1)	525.1 (112.9)	מכפ"ל
105.2 (20.7)	103.3 (21.1)	95.1 (19.9)	111.9 (19.4)	97.7 (18.2)	102.2 (22.0)	מילולי
111.6 (19.8)	109.6 (20.3)	101.9 (19.0)	118.1 (18.3)	105.3 (17.4)	106.7 (20.4)	כמותי
103.6 (24.5)	99.8 (25.3)	95.0 (23.3)	112.7 (23.0)	100.2 (22.0)	103.7 (25.7)	אנגלית

התבוננות בשורה הראשונה (ציוני מכפ"ל) מלמדת שהנטייה להיבחנות חוזרת קשורה, באופן לא מפתיע, לציון ראשון נמוך: בקבוצת נבחני פעם אחת הציון הראשון הוא 525 ובשתי קבוצות הנבחנים החוזרים הציון הראשון הוא 508 ו-488. חשוב להבהיר שקבוצת נבחני פעם אחת במחקר הנוכחי מורכבת מאנשים שביצעו את בחינתם היחידה ב-2008. חלקם ייבחנו שוב בעתיד, ויהפכו לנבחנים חוזרים, וחלקם לא ייבחנו יותר וישארו נבחני פעם אחת. ניתן להניח שממוצע הציון הראשון של נבחנים חוזרים עתידיים דומה לממוצע הציון הראשון של הנבחנים החוזרים במחקר הנוכחי (496). על פי הנחה זו (ובהנחה שהפרופורציות של נבחני פעם אחת ונבחנים חוזרים לא ישתנו) אומדן ממוצע הציון הראשון של נבחני פעם אחת "אמיתיים", אלו שיישארו נבחני פעם אחת, הוא 544.

השוואה בין שתי הקבוצות של הנבחנים החוזרים מלמדת שהנטייה להרוויח ממכפ"ל מרבי קשורה למכפ"ל בפועל נמוך יותר: ממוצע מכפ"ל בפועל בקבוצת הלא משפרים הוא 583 ובקבוצת המשפרים 530. הקשר בין ציון גבוה במכפ"ל בפועל לבין אי שיפור במכפ"ל מרבי עשוי לנבוע משני מקורות: אפשר שהלא משפרים היו מרוצים מציונם הגבוה במכפ"ל בפועל ולפיכך נטו פחות לחזור ולהיבחן, כפי שצוין קודם, וסיכוייהם להרוויח ממכפ"ל מרבי פחתו בהתאם. בנוסף, יתכן שאי השיפור נובע מסיבה סטטיסטית: ככל שהציון במכפ"ל בפועל גבוה יותר, הסיכוי לשפר אותו נמוך יותר (רגרסיה לממוצע).

התבוננות בקבוצת המשפרים במכפ"ל מרבי מלמדת שגודל השיפור במעבר בין מכפ"ל בפועל למכפ"ל מרבי הוא בממוצע 13 נקודות. לשם השוואה, השיפור במעבר ממכפ"ל ראשון למכפ"ל בפועל בקבוצה זו הוא 42 נקודות. כלומר, האפקט על הציון של מעבר ממכפ"ל בפועל למכפ"ל מרבי קטן מהאפקט של מעבר ממכפ"ל ראשון למכפ"ל בפועל, אך

גם הוא אינו מבוטל. עוד נציין שהמתאם בין מכפ"ל בפועל למכפ"ל מרבי בקבוצת המשפרים במכפ"ל מרבי הוא 0.996⁵.

הציונים בתחומי מכפ"ל בשיטת מכפ"ל מרבי בקרב הנבחנים שמשפרים במכפ"ל מרבי (105, 112 ו-104 עבור מילולי, כמותי ואנגלית בהתאמה; ראו עמודה שמאלית בלוח 1) מייצגים את הציונים בתחום הנידון מעבר לכל הנבחנים שהציון הכללי שלהם בשיטת מכפ"ל מרבי גבוה מצינום הכללי בשיטת מכפ"ל בפועל – חלקם שיפרו את ציונם בתחום הנידון ואחרים לא. כדי ללמוד על גודל השיפור בציון בכל אחד מתחומי מכפ"ל במעבר משיטת מכפ"ל בפועל לשיטת מכפ"ל מרבי נתמקד רק במשפרי הציון בתחום הנידון (שהם קבוצה חלקית של קבוצת המשפרים במכפ"ל מרבי). בלוח 2 מוצג אחוז המשפרים ציון וגודל השיפור בתחומי מכפ"ל במעבר משיטת מכפ"ל בפועל לשיטת מכפ"ל מרבי.

לוח 2: אחוז* המשפרים ציון בתחומי מכפ"ל במעבר משיטת מכפ"ל בפועל לשיטת מכפ"ל מרבי וממוצעים (וסטיות תקן) של ציוניהם בשתי השיטות

ציון מרבי	ציון בפועל	אחוז* המשפרים	
103.4 (21.1)	98.0 (21.3)	35	מילולי
110.3 (20.1)	105.0 (20.4)	38	כמותי
101.6 (24.7)	94.8 (25.2)	54	אנגלית

*מקרב הנבחנים שמשפרים במכפ"ל מרבי (N=14,503)

עיון בלוח מלמד שאחוז המשפרים, ובמידה מסוימת, גם גודל השיפור (5, 5 ו-7 נקודות במילולי, כמותי ואנגלית בהתאמה) גבוהים יותר באנגלית. יתכן שזהו ארטיפקט של העובדה שמשקל אנגלית בציון הכללי נמוך יותר: הקשר בין העובדה שהציון בתחום בהיבחנות מסוימת הוא הגבוה ביותר לבין העובדה שהציון הכללי בהיבחנות זו הוא הגבוה ביותר חלש יותר באנגלית מאשר במילולי או בכמותי. טענה זו משתקפת גם בממצא השיפור בציון במעבר מציון ראשון לציון בפועל הוא הנמוך ביותר באנגלית בקרב המשפרים במכפ"ל מרבי (לוח 1), וגם, אם כי במידה פחותה, בקרב הלא משפרים במכפ"ל מרבי.

עד כה התמקד הדיון בשלוש הקבוצות – נבחני פעם אחת, נבחנים חוזרים שלא משפרים במכפ"ל מרבי ונבחנים חוזרים שמשפרים במכפ"ל מרבי. כאמור, האחרונים, המהווים 21% מן המדגם, משפרים בממוצע ב-13 נקודות. האפקט הכולל של מעבר למכפ"ל מרבי על כל נבחני 2008 – מעבר לשלוש הקבוצות – הוא, לפיכך, 3 נקודות. אפקט זה נבדק על פי משתני רקע אחדים – שפת היבחנות במכפ"ל, השכלת הורים, מגדר וגיל. הממצאים מוצגים בלוח 3. נדגיש שאין מדובר בניתוח סיבתי אלא בממצאים תיאוריים בלבד.

⁵ מקדמי המתאם בין שלושת הציונים – מכפ"ל ראשון, מכפ"ל בפועל ומכפ"ל מרבי – בקרב שתי קבוצות הנבחנים החוזרים ומעבר לשלוש קבוצות הנבחנים מוצגים בנספח 1.

לוח 3: ממוצעים (וסטיות תקן) של ציוני מכפ"ל על פי משתני רקע

מכפ"ל מרב	מכפ"ל בפועל	מכפ"ל ראשון	N בקטגוריה	קטגוריה	N כללי*	משתנה רקע
521.6 (110.6)	518.5 (110.8)	497.1 (106.0)	37,277	נשים	64,676	מגדר
565.8 (108.2)	563.1 (108.3)	541.5 (106.0)	27,399	גברים		
523.1 (114.9)	519.0 (114.8)	487.8 (103.5)	34,019	צעירים	67,463	גיל
553.9 (107.0)	552.4 (107.0)	542.3 (106.4)	33,444	מבוגרים		
503.8 (108.5)	500.2 (108.5)	477.4 (101.2)	30,658	נמוך	57,901	השכלת הורים
578.7 (104.7)	576.1 (104.7)	551.9 (103.3)	27,243	גבוה		
567.3 (101.9)	565.7 (101.7)	547.6 (99.0)	46,576	עברית	64,632	שפת היבחנות
465.9 (107.0)	460.2 (105.3)	431.8 (88.0)	18,056	ערבית		

*של נבחים שידוע ששייכים לאחת משתי הקטגוריות של משתנה הרקע

ממצאי לוח 3 מעידים שבמגדר לא נמצא הבדל בין הקטגוריות בגודל השיפור במעבר ממכפ"ל בפועל למכפ"ל מרבי. בהתייחס לשלושת משתני הרקע האחרים מוצאים שהשיפור גדול יותר בקרב נבחים צעירים (4 נקודות) לעומת נבחים מבוגרים (2 נקודות); בקרב נבחים שהשכלת הוריהם נמוכה (4 נקודות) לעומת נבחים שהשכלת הוריהם גבוהה (3 נקודות); ובעיקר, בקרב נבחי ערבית (6 נקודות) לעומת נבחי עברית (2 נקודות). ניתן להתרשם ששתי הקטגוריות במשתני הרקע השונים שונות זו מזו בהישג הראשוני (מכפ"ל ראשון). נציין גם שיש הבדלים בין הקטגוריות במספר ההיבחניות הממוצע (לא מופיע בלוח). ככלל, התמונה שתוארה קודם ברמת המדגם כולו, של קשר בין הישג ראשוני, מספר היבחניות ושיפור במכפ"ל מרבי מתקיימת גם כאן, במובן שהקטגוריות שנוטות לשיפור יותר הן אלה שהישגן הראשוני נמוך יותר וממוצע מספר ההיבחניות שלהן גבוה יותר. כאמור, ההתבוננות שהוצעה כאן היא תיאורית בלבד ולא כוללת טיפול מורכב במשתנים. יתרה מזו, חלק ממשתני הרקע השונים מתואמים זה עם זה (כך, למשל, המתאם בין שפת היבחנות לבין גיל הוא 0.51. והמתאם בין שפת היבחנות לבין השכלת הורים הוא 0.24). חלק מהקשרים שנמצאו בין משתני הרקע לבין גודל השיפור נובעים ממתאם זה (לשם המחשה, ראו בנספח 2 ניתוח של האפקט של משתני הרקע מגדר, גיל והשכלת הורים בנפרד עבור נבחי עברית וערבית).

מחקר ב'

מטרתו של מחקר זה הייתה, כאמור, לבחון את השינוי הצפוי בתוקף הניבוי של ציון הסכם ושל מכפ"ל ומרכיביו כתוצאה ממעבר משימוש במכפ"ל בפועל לשימוש במכפ"ל מרבי. מחקר זה בוצע בקרב 57,632 רשומות של תלמידי שנה א' באוניברסיטאות בישראל. שאלת

התוקף נבחנה בראש ובראשונה מעבר לשלוש הקבוצות שהוגדרו על פי היבחנות חוזרת ושיפור – נבחני פעם אחת ושתי הקבוצות של נבחנים חוזרים – מאחר שתהליך הבררה מתבצע מעבר לשלושתן. עבור כל חזאי הוגדרו שלושה ציונים – ציון ראשון, ציון בפועל וציון מרבי.⁶ בנוסף, נבדק תוקף הניבוי של החזאים השונים בקרב כל אחת משלוש הקבוצות. בלוחות 4 ו-5 מוצגים מקדמי התוקף⁷ של החזאים, כאשר בלוח 4 מוצגים הנתונים עבור ציון סכם, בגרות ומכפ"ל, ובלוח 5 מוצגים הנתונים עבור הציונים בתחומי מכפ"ל.

לוח 4: מקדמי התוקף של ציון סכם, בגרות ומכפ"ל לפי היבחנות חוזרת ושיפור ומעבר לכולם

קבוצה	מס' יח' עיבוד	מס' תצפיות	סכם ראשון	סכם בפועל	סכם מרבי	בגרות	מכפ"ל ראשון	מכפ"ל בפועל	מכפ"ל מרבי
כולם	661	57632	0.49	0.48	0.47	0.37	0.46	0.44	0.43
נבחני פעם אחת	659	30867	0.53			0.46	0.44		
נבחנים חוזרים	575	14270	0.37	0.38		0.30	0.34	0.37	
משפרים	573	12024	0.31	0.30	0.30	0.17	0.36	0.36	0.35

לוח 5: מקדמי התוקף של הציונים בתחומי מכפ"ל לפי היבחנות חוזרת ושיפור ומעבר לכולם*

קבוצה	מילולי ראשון	מילולי בפועל	מילולי מרבי	כמותי ראשון	כמותי בפועל	כמותי מרבי	אנגלית ראשון	אנגלית בפועל	אנגלית מרבי
כולם	0.40	0.37	0.36	0.43	0.39	0.38	0.36	0.34	0.33
נבחני פעם אחת	0.37			0.41			0.29		
נבחנים חוזרים	0.25	0.28		0.33	0.33		0.24	0.27	
משפרים	0.28	0.25	0.26	0.31	0.30	0.30	0.28	0.30	0.30

* השכיחויות זהות לאלה שבלוח 4.

ניתן להיווכח שמעבר לכל הסטודנטים התוקף הולך ויורד (באורח מתון) במעבר מהציון הראשון, לציון בפועל ולציון המרבי. ממצא זה מתקיים ביחס למכפ"ל ולכל אחד מתחומי וביחס לציון הסכם (בניתוח לפי המשתנים המתנים לא נמצאו הבדלים בין הצירופים השונים של תחום לימוד וסלקטיביות בעניין זה. התוצאות מניתוח זה מוצגות בנספח 5).

בקבוצת הנבחנים שהציון המרבי שלהם שונה מציונם בפועל ("חוזרים שמשפרים") מוצאים אינדיקציות חלשות בכיוון המגמה שתוארה: תוקף הציון המרבי זהה לתוקף הציון בפועל (ושניהם נמוכים מתוקף הציון הראשון) ביחס לציון הסכם, ותוקף הציון המרבי נמוך מתוקף הציון בפועל (שזהה לתוקף הציון הראשון) ביחס לתחומי מכפ"ל אין זהות במקדמי התוקף של שלושת הציונים, אך אין מגמה עקבית של סדר ביניהם.

יש, אם כן, אינדיקציות לכך שהירידה (המזערית) בתוקף מעבר לכל הסטודנטים במעבר מציון בפועל לציון מרבי נובעת משני מקורות: האחד הוא ירידה בתוקף החזאי בתוך קבוצת המשפרים במכפ"ל מרבי, והשני הוא צמצום שונות רלוונטית לקריטריון בין הקבוצות.

⁶ כאשר אצל חלק מן הסטודנטים זהים שלושת הציונים או חלקם.

⁷ אלו מקדמים מתוקנים לקיצוץ תחום. מקדמי תוקף נצפים מוצגים בנספח 4. בנספח 3 מוצגים ממוצעים וסטיות תקן של הקריטריון והחזאים.

סיכום ודין

במחקר הנוכחי נבדקו ההשלכות של שימוש במכפ"ל מרבי לצורך בררת מועמדים ללימודים באוניברסיטאות, נוהל רווח בקבלה ללימודים במכללות בארה"ב. ממצאי המחקר מעידים שמעבר למכפ"ל מרבי עשוי להעלות את ציונם של כחמישית מן הנבחנים בעשירית סטיית תקן⁸. נבחנים אלה מאופיינים בציון התחלתי נמוך יחסית. מבין תחומי מכפ"ל אחוז המשפרים וגודל השיפור היו הגבוהים ביותר באנגלית. בבדיקת האפקט על פי משתני רקע אחדים, נמצאה שפת היבחנות כקשורה ביותר לגודל השיפור, כאשר השיפור בקרב נבחני ערבית היה גדול מן השיפור בקרב נבחני עברית (אם כי גם במקרה זה השיפור לא היה רב). הקשר בין משתני רקע מסויימים, ובפרט בין שפת היבחנות במכפ"ל, לבין גודל השיפור מוסבר, כמשוער, על ידי הנטייה להיבחנות חוזרת, והיא, בתורה, מוסברת על ידי ציון התחלתי נמוך. כך, למשל, נבחני ערבית (ובהתאם, גם נבחנים צעירים ונבחנים שהשכלת הוריהם נמוכה) נבחנים יותר פעמים והציון הראשון שלהם נמוך יותר. ויגדור וקלוטפלט (Vigdor & Clotfelter, 2003) מדווחים על ממצא דומה, קרי, שנבחנים מאוכלוסיות ממיצב חברתי-כלכלי נמוך נוטים להיבחן יותר, לכאורה בניגוד לתפיסה שהיבחנות חוזרת היא נחלתם של העשירים. בניתוח של מתאמים חלקיים, שבו החזיקו קבוע ציונים קודמים, נמצא שהיבחנות חוזרת אמנם קשורה לעושר רב יותר ולהשכלת הורים גבוהה יותר. המתאמים החלקיים בין משתני רקע לבין היבחנות חוזרת מוסוויים, אם כן, על ידי הקשר החזק בין ציון התחלתי נמוך לבין הנטייה להיבחנות חוזרת (ובין מיצב חברתי-כלכלי נמוך לבין ציון התחלתי נמוך). כדי להשיג הבנה מלאה יותר של התופעה מן הראוי לבדוק מודלים תיאורטיים שונים שיביאו בחשבון את מכלול הגורמים. עם זאת, ניתוח המתאמים החלקיים אינו מבטל את המסקנה שמתקבלת מן המתאמים הפשוטים, שעל פיה קבוצות חלשות עשויות להרוויח יותר ממעבר לשיטת מכפ"ל מרבי.

בבדיקת תוקף הניבוי מעבר לכל הסטודנטים נמצאה ירידה מזערית בתוקף הניבוי של ציון הסכם ומרכיביו במעבר מציון בפועל לציון מרבי. בקבוצת הסטודנטים שמשפרים במכפ"ל מרבי היה תוקף הציונים השונים, ובפרט התוקף של הציון בפועל והציון המרבי, זהה ביחס לציון הסכם, אך לא ביחס למכפ"ל, שם הניבוי הציון המרבי תוקף נמוך יותר. חשוב להדגיש שההבדלים שנמצאו בתקפיות הינם מזעריים ביותר, ממצא שהוא עקבי עם ממצאים שנאספו בארה"ב על הבדלים מזעריים בתוקף של חלופות שונות להגדרת הציון הקובע (Boldt et al., 1986). נראה, אם כן, שהירידה (המזערית) בתוקף מעבר לכל הסטודנטים במעבר מציון בפועל לציון מרבי נובעת משני מקורות: האחד הוא ירידה בתוקף החזאי בתוך קבוצת המשפרים במכפ"ל מרבי, והשני הוא צמצום שונות רלוונטית לקריטריון בין הקבוצות.

⁸ ב-13 נקודות, כאשר ממוצע מכפ"ל בקרב נבחני 2008 הוא 530 וסטיית התקן היא 114.

חשוב להדגיש שבדיקת ההשלכות של שינוי מדיניות הקבלה בוצעה בהנחה ששינוי זה לא ישפיע על דפוסי ההיבחנות. בדיקת השלכות אלו תוך הבאה בחשבון של השפעת מדיניות הקבלה על דפוסי ההיבחנות מודגמת בעבודתם של ויגדור וקלוטפלטר (Vigdor & Clotfelter, 2003), שהשתמשו בנתונים על מועמדים לשלוש אוניברסיטאות סלקטיביות, כדי לנתח את החלטתו של מועמד ללימודים להיבחן שוב ב-SAT. הם מידלו החלטה זו כבעיית חיפוש אופטימלי, והשתמשו במודל כדי להעריך את האימפקט של שימוש בציונים קובעים חלופיים – סכום הציונים הגבוהים ביותר בחלקים (ציון מרבי), ציון ראשון, ציון ממוצע, ציון אחרון וממוצע שני ציונים (ראשונים) עבור כל מועמד⁹ – על ההתנהגות של היבחנות חוזרת. כצפוי, המדיניות של שימוש בסכום הציונים הגבוהים ביותר בחלקים (ציון מרבי) נמצאה כמספקת את התמריצים הגדולים ביותר לחזור ולהיבחן. אחרי שהביאו בחשבון את ההשפעה של כל מדיניות על מספר ההיבחנות, הישוו החוקרים בין ציונים קובעים חלופיים על פי ארבעה קריטריונים, ביניהם *דיוק* (*precision*), שהוגדר כסטיית התקן של הפרשים בין אומדן היכולת על ידי הציון הקובע לבין הערך האמיתי של הפרמטר¹⁰. הממצאים ביחס ל*דיוק* מורים שהמדיניות של שימוש בציון מרבי מפיקה אומדן מדויק יותר של היכולת האמיתית מזה של הממוצע, וזאת בניגוד לממצאי מחקרים רבים שנסקרו בעבודתם של בולט, סנטרה וקורטני (Boldt et al., 1986), שהעידו שלממוצע יש התוקף הגבוה ביותר (כאמור, בהסתייגות שהפערים בין תקפיות הניבוי של הציונים הקובעים השונים אינם גדולים). הסיבה לסתירה לכאורה, כפי שמסבירים זאת ויגדור וקלוטפלטר היא ש"כמות המידע היא אנדוגנית למדיניות" (Vigdor & Clotfelter, 2003, p. 29). במילים אחרות, ההשוואה הנכונה בין תוקפו של הממוצע לתוקפו של הציון המרבי צריכה להביא בחשבון שתחת המדיניות של שימוש בממוצע, מועמדים ייבחנו פחות פעמים מאשר תחת המדיניות של שימוש בציון מרבי, ולפיכך אומדן היכולת שלהם יתבסס על פחות מידע. הדברים שלעיל ממחישים שהנחת דפוסי ההיבחנות הקבועים היא הנחה חזקה ושאם מביאים בחשבון שהיא אינה מתקיימת, התוצאות עשויות להשתנות באופן משמעותי. באופן מפורט, ניתן לשער שללא הנחת דפוסי ההיבחנות הקבועים יהיה תוקף הניבוי של ציון מרבי גבוה מזה שנאמד במחקר הנוכחי.

המחקר הנוכחי התמקד בהשלכות המעבר לציון מרבי על תוקף הניבוי. כפי שהוזכר במבוא, רבות מן העבודות שעסקו בהשוואה בין ציונים קובעים שונים התייחסו גם (לעיתים רק) לשאלות של הוגנות (לרוב תחת הפרדיגמה של הניבוי הדיפרנציאלי). נושא זה ראוי לחקירה עתידית. לבסוף, בנוסף לסוגיות של תוקף ניבוי והוגנות, מן הראוי להביא בחשבון גם

⁹ מעניין לראות שהחלופה של הציון הכללי הגבוה ביותר לא נבדקה כלל.
¹⁰ שלושת הקריטריונים הנוספים שנבדקו היו *נכונות* (*accuracy*), שהיא הפרש הממוצע בין אומדן היכולת על ידי הציון הקובע לבין הערך האמיתי של הפרמטר, *הטיה* (*bias*), שהיא הפרש בין הנכונות הממוצעת עבור נבחנים עם עלות גבוהה להיבחנות (למשל, מועמדים ממיצב חברתי-כלכלי נמוך) לבין הנכונות הממוצעת עבור נבחנים עם עלות נמוכה להיבחנות *ועלות* (*cost*), שהיא מספר ההיבחנות הממוצע למועמד.

את התוקף ההשלכתי (consequential validity) של המדיניות המוצעת (כגון, תפקוד והתנהגות בעת ההיבחנות בתחומים שונים של המבחן), נושא שחורג מעבר לטווח המחקר הנוכחי.

מקורות

אורן, כ' וכהן, א' (1988). הוגנות השימוש בבחינה הפסיכומטרית השנייה של נבחנים חוזרים בקרב תלמידי מכינת הטכניון (דוח מס' 67). ירושלים: מרכז ארצי לבחינות ולהערכה.

סטולר, ר' (2010). דוח סטטיסטי למועדים: אוקטובר, דצמבר 2009 (דוח פנימי מס' 365). ירושלים: מרכז ארצי לבחינות ולהערכה.

קנת-כהן, ת' ואורן, כ' (1989). על היעילות וההוגנות של הגדרת ציונים קובעים שונים בבררת נבחנים חוזרים במכפ"ל (דוח מס' 110). ירושלים: מרכז ארצי לבחינות ולהערכה.

קנת-כהן, ת', ברונר, ש' ואורן, כ' (1998). תוקף צירופים שונים של מרכיבי המבחן הפסיכומטרי ותעודת הבגרות בניבוי הישגי שנה א' באוניברסיטאות בישראל (דוח מס' 249). ירושלים: מרכז ארצי לבחינות ולהערכה.

Boldt, R. F., Centra, J. A., & Courtney, R. G. (1986). *The validity of various methods of treating multiple SAT scores*. College Board Report 86-4 (ETS Research Report 86-25). New York: College Entrance Examination Board.

CollegeBoard (2010). *SAT Score-use practices by participating institution*. Retrieved from <http://professionals.collegeboard.com/profdownload/sat-score-use-practices-list.pdf>

Cowen, M. B., & Abrahams, N. M. (1982). *Selecting qualified candidates to the United States Naval Academy using college aptitude test scores*. NPRDC Special Report 82-20. San Diego, CA: Navy Personnel Research and Development Center.

Dalessandro, S. P., & McLeod, L. D. (1999). *The validity of Law School Admission Test scores for repeaters: A replication* (LSAC Technical Report No. 98-05). Newtown, PA: Law School Admission Council.

Gulliksen, H. (1950). *Theory of mental tests*. New York: John Wiley & Sons. {Reprinted in 1987. Hillsdale, NJ: Erlbaum.}

- Olsen, M., & Schrader, W. B. (1958). *Relation of preliminary and final Scholastic Aptitude Test scores to college grades: Preliminary report*. College Entrance Examination Board Research and Development Reports; also Statistical Report 58-15. New York: College Entrance Examination Board.
- Olsen, M., & Schrader, W. B. (1959a). *The use of preliminary and final Scholastic Aptitude Test scores in predicting college grades*. College Entrance Examination Board Research and Development Reports; also Statistical Report 59-19. Princeton, NJ: Educational Testing Service.
- Olsen, M., & Schrader, W. B. (1959b). *The comparative achievement of final-only and preliminary-final Scholastic Aptitude Test candidates*. College Entrance Examination Board Research and Development Reports; also Statistical Report 59-20. Princeton, NJ: Educational Testing Service.
- Pitcher, B. (1965). Internal College Board file. Referenced in: Boldt, R. F., Centra, J. A., & Courtney, R. G. (1986). *The validity of various methods of treating multiple SAT scores*. College Board Report 86-4 (ETS Research Report 86-25). New York: College Entrance Examination Board.
- Pitcher, B. (1976). Internal College Board file. Referenced in: Boldt, R. F., Centra, J. A., & Courtney, R. G. (1986). *The validity of various methods of treating multiple SAT scores*. College Board Report 86-4 (ETS Research Report 86-25). New York: College Entrance Examination Board.
- Pitcher, B. (1977). Internal College Board file. Referenced in: Boldt, R. F., Centra, J. A., & Courtney, R. G. (1986). *The validity of various methods of treating multiple SAT scores*. College Board Report 86-4 (ETS Research Report 86-25). New York: College Entrance Examination Board.
- Stilwell, L. A., Thornton, A. E., & Pashley, P.J. (2002). *The validity of Law School Admission Test scores for repeaters: 1997 through 2000 entering law school classes* (LSAT Technical Report No. 02-02). Newtown, PA: Law School Admission Council.
- Thornton, A. E., Stilwell, L. A., & Reese, L. R. (2006). *The validity of Law School Admission Test Scores for repeaters: 2001 through 2004 entering law school classes* (LSAT Technical Report No. 06-02). Newtown, PA: Law School Admission Council.

Vigdor, J. L., & Clotfelter, C. T. (2003). Retaking the SAT. *Journal of Human Resources*, 38, 1-33.

Wightman, L. F. (1990). *The validity of Law School Admission Test scores for repeaters: A replication* (Research Report No. 90-02). Newtown, PA: Law School Admission Council.

נספח 1: מקדמי מתאם בין שלושת ציוני מכפ"ל בקרב הנבחנים

לוח 6: מקדמי המתאם בין מכפ"ל ראשון, מכפ"ל בפועל ומכפ"ל מרבי בקרב שתי

קבוצות הנבחנים החוזרים ומעבר לשלוש קבוצות הנבחנים

קבוצה	ראשון ובפועל	ראשון ומרבי	בפועל ומרבי
כולם	0.937	0.931	0.998
נבחנים לא משפרים	0.894	0.894	1
חוזרים משפרים	0.918	0.916	0.996

נספח 2: ממוצעים (וסטיות תקן) של ציוני מכפ"ל בעברית ובערבית על פי

משתני רקע

לוח 7: ממוצעים (וסטיות תקן) של ציוני מכפ"ל בקרב נבחני עברית על פי משתני רקע

מכפ"ל מרבי	מכפ"ל בפועל	מכפ"ל ראשון	N	קטגוריה	משתנה רקע
552.0 (102.1)	550.2 (101.8)	531.3 (98.2)	24,512	נשים	מגדר
589.8 (96.7)	588.2 (96.4)	569.7 (95.1)	20,192	גברים	
572.7 (100.9)	570.5 (100.6)	537.7 (93.3)	16,458	צעירים	גיל
564.8 (102.1)	563.4 (101.9)	553.4 (101.4)	29,968	מבוגרים	
539.2 (100.0)	537.3 (99.6)	516.8 (94.3)	17,941	נמוך	השכלת הורים
597.6 (95.6)	595.7 (95.4)	574.2 (94.9)	20,833	גבוה	

לוח 8: ממוצעים (וסטיות תקן) של ציוני מכפ"ל בקרב נבחני ערבית על פי משתני רקע

מכפ"ל מרבי	מכפ"ל בפועל	מכפ"ל ראשון	N	קטגוריה	משתנה רקע
457.7 (103.2)	451.8 (101.8)	423.9 (85.5)	11,211	נשים	מגדר
489.7 (111.5)	483.5 (109.5)	450.4 (90.4)	6,001	גברים	
472.4 (107.1)	466.5 (105.5)	436.2 (88.1)	16,483	צעירים	גיל
393.9 (76.9)	390.2 (75.4)	379.9 (69.3)	1,399	מבוגרים	
451.7 (100.7)	445.7 (98.9)	418.9 (82.2)	11,911	נמוך	השכלת הורים
512.4 (113.5)	506.3 (111.9)	467.5 (94.0)	4,632	גבוה	

נספח 3: ממוצעים וסטיות תקן בקרב הלומדים

לוח 9: ממוצעים וסטיות תקן של ציון שנה א', ציון סכמ, בגרות ומכפ"ל לפי היבחנות חוזרת ושיפור ומעבר לכולם

מכפ"ל מרבי	מכפ"ל בפועל	מכפ"ל ראשון	בגרות	סכמ מרבי	סכמ בפועל	סכמ ראשון	ציון שנה א'	מס' תצפיות (ויחי' עיבוד)	קבוצה
632.8 (56.0)	630.3 (56.8)	597.2 (73.1)	100.2 (6.6)	50.0 (7.0)	50.0 (7.1)	50.0 (7.7)	81.1 (9.2)	57,632 (661)	כולם
633.7 (59.1)			100.0 (6.6)	50.0 (7.4)	50.2 (7.4)	52.3 (7.2)	83.0 (8.4)	30,867 (659)	נבחני פעם אחת*
645.3 (48.3)		559.4 (59.0)	100.2 (6.2)	50.6 (6.2)	50.8 (6.1)	47.1 (6.6)	79.8 (9.0)	14,270 (575)	לא משפרים
616.6 (51.9)	604.7 (52.4)	550.2 (65.9)	100.8 (6.6)	49.6 (6.4)	48.9 (6.4)	47.7 (6.8)	77.6 (9.9)	12,024 (573)	משפרים

* השוני בין שלושת ציוני הסכמ של נבחני פעם אחת ובין סכמ בפועל וסכמ מרבי של חוזרים שלא משפרים נובע מאופן

הגדרתם (ראו בפרק השיטה).

לוח 10: ממוצעים וסטיות תקן של הציונים בתחומי מכפ"ל לפי היבחנות חוזרת ושיפור ומעבר לכולם*

אנגלית מרבי	אנגלית בפועל	אנגלית ראשון	כמותי מרבי	כמותי בפועל	כמותי ראשון	מילולי מרבי	מילולי בפועל	מילולי ראשון	קבוצה
124.3 (15.2)	123.6 (15.6)	118.8 (17.8)	123.6 (11.5)	123.2 (11.6)	117.7 (13.7)	122.5 (12.5)	122.1 (12.7)	115.5 (16.0)	כולם
125.7 (15.0)			122.5 (12.2)			123.3 (13.0)			נבחני פעם אחת
125.7 (14.1)		112.1 (16.6)	127.07 (10.0)		112.6 (11.9)	123.9 (11.1)		107.2 (13.6)	לא משפרים
119.1 (15.4)	115.8 (16.3)	109.0 (18.0)	122.6 (10.7)	120.7 (11.1)	111.5 (13.1)	119.0 (11.7)	117.3 (12.3)	105.9 (14.8)	משפרים

* השכיחויות זהות לאלה שבלוח 9.

נספח 4: מקדמי תוקף נצפים

לוח 11: מקדמי תוקף נצפים של ציון סכמ, בגרות ומכפ"ל לפי היבחנות חוזרת ושיפור ומעבר

לכולם

מכפ"ל מרבי	מכפ"ל בפועל	מכפ"ל ראשון	בגרות	סכמ מרבי	סכמ בפועל	סכמ ראשון	מס' תצפיות	מס' יח' עיבוד	קבוצה
0.30	0.31	0.34	0.21	0.33	0.34	0.37	57,632	661	כולם
0.29			0.32	0.40			30,867	659	נבחני פעם אחת
0.24		0.23	0.14	0.26		0.24	14,270	575	נבחנים לא משפרים
0.28	0.29	0.30	0.04	0.20	0.21	0.23	12,024	573	חוזרים משפרים

לוח 12: מקדמי תוקף נצפים של הציונים בתחומי מכפ"ל לפי היבחנות חוזרת ושיפור ומעבר

לכולם*

אנגלית מרבי	אנגלית בפועל	אנגלית ראשון	כמותי מרבי	כמותי בפועל	כמותי ראשון	מילולי מרבי	מילולי בפועל	מילולי ראשון	קבוצה
0.21	0.23	0.25	0.26	0.27	0.33	0.22	0.22	0.28	כולם
0.16			0.28			0.22			נבחני פעם אחת
0.17		0.14	0.23		0.24	0.15		0.14	נבחנים לא משפרים
0.24	0.25	0.23	0.23	0.23	0.26	0.17	0.17	0.22	חוזרים משפרים

* השכיחויות זהות לאלה שבלוח 11.

נספח 5: מקדמי תוקף¹¹ לפי תחום לימוד וסלקטיביות מעבר לכולם¹²

לוח 13: מקדמי תוקף של ציון סכמ, בגרות ומכפ"ל לפי תחום לימוד וסלקטיביות

מכפ"ל מרב	מכפ"ל בפועל	מכפ"ל ראשון	בגרות	סכמ מרב	סכמ בפועל	סכמ ראשון	מס' תצפיות	מס' יח' עיבוד	סלקטיביות	תחום לימוד
0.42	0.43	0.44	0.28	0.42	0.43	0.44	19,607	230	נמוכה	מילולי
0.48	0.49	0.50	0.43	0.51	0.52	0.54	8,483	67	גבוהה	מילולי
0.37	0.39	0.40	0.30	0.40	0.41	0.42	6,476	87	נמוכה	כמותי
0.45	0.46	0.48	0.44	0.52	0.52	0.53	23,066	277	גבוהה	כמותי
0.43	0.44	0.46	0.37	0.47	0.48	0.49	57,632	661	כולם	כולם

לוח 14: מקדמי תוקף של הציונים בתחומי מכפ"ל לפי תחום לימוד וסלקטיביות*

אנגלית מרב	אנגלית בפועל	אנגלית ראשון	כמותי מרב	כמותי בפועל	כמותי ראשון	מילולי מרב	מילולי בפועל	מילולי ראשון	סלקטיביות	תחום לימוד
0.33	0.34	0.34	0.33	0.34	0.37	0.38	0.38	0.40	נמוכה	מילולי
0.36	0.37	0.40	0.39	0.40	0.44	0.42	0.43	0.46	גבוהה	מילולי
0.28	0.29	0.30	0.33	0.33	0.37	0.28	0.29	0.32	נמוכה	כמותי
0.32	0.34	0.37	0.44	0.44	0.49	0.35	0.35	0.40	גבוהה	כמותי
0.33	0.34	0.36	0.38	0.39	0.43	0.36	0.37	0.40	כולם	כולם

* השכיחויות זהות לאלה שבלוח 13.

¹¹ מתוקנים לקיצוץ תחום

¹² כלומר, מעבר לשלוש הקבוצות המוגדרות על פי היבחנות חוזרת ושיפור

The predictive validity with respect to the criterion of first-year average – study 2 – was examined on the basis of 57,632 records of 46,512 students who began their academic studies at Israeli universities from 2004 through 2006.

The predictive validity was examined among repeaters who gained from using the highest PET as well as for the total group (repeaters and one-time test takers combined). Validities for the three ranking policies – using the highest, current or initial PET score – were generally very similar in both analyses, a result which is consistent with findings reported in the literature. However, a certain trend of lower validities for the highest PET score could be discerned. This trend was more evident in validities within the total group than in validities within the group of repeaters who gained from using the highest PET. This combination of results indicates that the slight decrease in validity in the total group when moving from the current PET to the highest PET derives from two sources: a decrease in validity **within** the repeaters who gain from highest PET and a reduction in the criterion-related-variance of PET **between** one-time test takers and repeaters.

The study was based on the assumption that the method of treating multiple scores did not affect retesting behavior, which is probably an unrealistic assumption. This fact should be taken into account when conclusions are drawn from the study.

The Consequences of Using Highest Domain Scores Across Test Dates

Tamar Kennet-Cohen, Shmuel Bronner, Carmel Oren

Abstract

The objective of the study was to examine the consequences – in the context of predictive validity – of using, for purposes of university admission, a total Psychometric Entrance Test (PET) score computed from the highest score in each domain, even if those scores were obtained on different dates. Such a score will be referred to as the "highest PET". The use of this score was compared with the current policy of using the highest total score obtained by an examinee at a single test administration. This score will be referred to as the "current PET". In order to have a context for this comparison, a policy of using the initial PET score was also examined.

The consequences examined were related to expected score gain and expected change in the predictive validity of relevant measures of the admissions process (PET and its components, and the Admission Score).

The issue of the score gain – study 1 – was examined among 67,806 PET examinees in 2008. Thirty-seven percent of these examinees took the test more than once. Among them, 58 percent gained from switching to the policy of using the highest PET. The average score gain was 13 score points. For comparison, the average score gain when switching from the initial PET to the current PET was 42 score points. Thus, the effect on the score of switching from the current PET to the highest PET is smaller than the effect of switching from the initial PET to the current PET, but it is still not trivial.

The magnitude of the score gain resulting from switching from the current PET to the highest PET was examined with respect to several background variables: testing language (Arabic or Hebrew), parents' education, age at testing and gender. Of these variables, testing language was found to have the strongest relation with the magnitude of score gain: Arabic-speaking examinees gained 6 score points on average, compared with 2 score points for Hebrew-speaking examinees.