

# הכינוס התטיעי טל האגודה הישראלית לפסיכומטריקה

2013

כ"ד בשבט תשע"ג 4.2.2013  
מרכז מורשת בגין, ירושלים

בתמיכת:



מרכז ארצי לבחינות ולהערכה (נ"ר)  
NATIONAL INSTITUTE FOR TESTING & EVALUATION  
المركز القطري للامتحانات والتقييم  
מיסודן של האוניברסיטאות בישראל



האגודה הישראלית לפסיכומטריקה (אפי) היא גוף המאגד בתוכו אנשי מקצוע מתחום המדידה וההערכה החינוכית והפסיכולוגית.

**מטרות האגודה:**

- לעודד את הפעילות בתחום הפיתוח והיישום של מבחנים סטנדרטיים ושיטות להערכת ביצוע
- לקדם את המחקר והעשייה בתחומי המדידה החינוכית והפסיכולוגית
- לשכלל את שיטות ההערכה וכלי המדידה ולבסס את יישומם המושכל
- להתוות, להעריך ולהפיץ כלים מתודולוגיים עדכניים ושיטות הערכה חדשות
- להתריע על שימוש לקוי במבחנים ועל פרשנות מוטעית של תוצאותיהם
- לעודד קשר מקצועי, חילופי מידע ושיתוף פעולה בין העוסקים בתחום המדידה וההערכה

## תכנית

| 9:00 – 8:30   | התכנסות  |
|---------------|--|
| 9:15 – 9:00   | <b>דברי פתיחה</b> יו"ר אפי: ענת בן-סימון, יו"ר הכינוס: צור קרליץ וספי פומפיאן, עורכת ידיעון אפי: עינת נוטע-קורן  |
| 10:45 – 9:15  | <b>מושב א – יו"ר: חגית גליקמן</b>  |
|               | <ul style="list-style-type: none"> <li> <b>פערים בין בנים לבנות בהישגים במתמטיקה במבחנים מערכתיים בישראל – בחינת מגמות והתאמת הנתונים למודלים שונים של פערים בין-מגדריים</b><br/>                     יואל רפ, עינת נוטע-קורן, ענבל רוז-קפלן, אימאן עואדיה, ענר רוגל, הדס גלברט, אמג'ד בשארה                 </li> <li> <b>המוטיבציה בביצוע מחקר פיזה – מגמות ושינויים בין מחקר 2006 לבין מחקר 2009</b><br/>                     עמיחי רגבי, ענבל רוז-קפלן, יואל רפ                 </li> <li> <b>מידול אוריינות ההערכה של מורים מתחילים בישראל: האפקט של מרכיבי ההכשרה להוראה, תפיסות ביחס להערכה וחוללות עצמית ביחס להערכת לומדים</b><br/>                     עדי לוי-ורד, פאדיה נאסר-אבו אלהיג'א                 </li> <li> <b>אינפלציה בציוני הבגרות בישראל</b><br/>                     דוד מעגן, ליבי שפירא                 </li> </ul> |
| 11:10 – 10:45 | הפסקה  |
| 12:40 – 11:10 | <b>מושב ב – יו"ר: ריטה סבר</b>   |
|               | <ul style="list-style-type: none"> <li> <b>לתקן או לא לתקן? על היתרונות שבשימוש בציונים מתוקננים של פריטי שאלונים ומבחנים</b><br/>                     לירון צוברי, אייל גמליאל, רז ליבוביץ                 </li> <li> <b>בגרות ופסיכומטרי והקשר ביניהם – על הנטל העודף של מבחנים</b><br/>                     שלמה יצחקי, טאינה פודלוב, אביאל קרנצ'לר                 </li> <li> <b>מבוגרים בעלי דיסקלקוליה אינם שונים מקבוצת ביקורת בעיבוד אוטומטי של מידע נומרי בסיסי</b><br/>                     יוסף צלגוב, בר זוהר-שי, טלי ליבוביץ, ענת בן-סימון, אורלי רובינשטיין                 </li> <li> <b>המבנה הגורמי של תפקודי השפה המשמשים באבחון דיסקליה במת"ל</b><br/>                     ענת בן-סימון, מאיר ברנרן, נורית ענבר-וייס                 </li> </ul>  |
| 13:00 – 12:40 | <b>הענקת הפרס לעבודות מחקר בתחום המדידה וההערכה לשנת 2013</b>  |
|               | <ul style="list-style-type: none"> <li> <b>הערכת עמיתים בקרב סטודנטים לרפואה ככלי לבחינת תוקף הניבוי של מבחני הקבלה הלא-קוגניטיביים ללימודי רפואה באוניברסיטה העברית</b><br/>                     אביאל חדד                 </li> <li> <b>פיתוח כלי מבוסס יכולת למדידת אינטליגנציה רגשית-חברתית</b><br/>                     יוסי חסון                 </li> </ul>   |

14:20 – 13:00 ארוחת צהריים

15:50 – 14:20 מושב ג – יו"ר: ענת קדם

- **מבט על השימוש במבחנים פסיכולוגיים בישראל והאתגרים החדשים**  
סול פיין
- **פיתוח מודל לסיווג עולם המקצועות עבור מיון תעסוקתי הבודק יכולת ביטוי בכתב**  
הילה שורק-פוזס, אבי אללוף
- **השפעת המגדר והמוצא האתני של המשתתפים והמעריכים על ההצלחה במרכז ההערכה**  
אמיר פלק, שאול פוקס
- **בחינת מיומנות ידנית ברפואת שיניים באמצעות סימולטור מדמה מציאות**  
גלעד בן-גל, נעמי גפני, אמתי זיו, רומינה סרנה, ארוין וייס

16:15 – 15:50 הפסקה

17:45 – 16:15 מושב ד – יו"ר: כרמל אורן

- **דעת הקהל והתוקף הנראה של הבחינה הפסיכומטרית**  
צור קרליץ, עדו ליטמנוביץ'
- **כישורים לימודיים ודפוסי קבלת החלטה בבחירת לימודים – ממצאים ממחקר על נשירה במערכת ההשכלה הגבוהה**  
תמי קנת-כהן, איתמר גתי
- **הוגנות הליך הקבלה להשכלה הגבוהה כלפי מועמדים מבקשי התאמות**  
דביר קלפר, אליוט טורוול, תמי קנת-כהן, כרמל אורן
- **תועלת קורסי ההכנה לבחינה הפסיכומטרית, בהשוואה להכנה עצמית, בקרב נבחני עברית וערבית**  
איילת גולדצויג, כרמל אורן, יונתן סער

18:15 – 17:45 אסיפת חברי אפי

## פערים בין בנים לבנות בהישגים במתמטיקה במבחנים מערכתיים בישראל - בחינת מגמות והתאמת הנתונים למודלים שונים של פערים בין-מגדריים

יואל רפ, עינת נוטע-קורן, ענבל רון-קפלן, אימאן עואדיה,  
ענר רוגל, הדס גלברט, אמג'ד בשארה

הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

לאורך שנים רבות נצרכה בתודעה הסברה כי בנים מוכשרים יותר מבנות במתמטיקה, ולצדה התפתחה ספרות אקדמית נרחבת והתקיים פולמוס מתמשך סביב השאלה האם באמת קיים פער ביכולת זו בין המינים, ואם כן מה מקורו - תורשתי או סביבתי. ממצאים אמפיריים שנאספו במבחני הישגים בית-ספריים רחבי היקף במתמטיקה מראים כי אכן קיים במקרים רבים פער לטובת בנים וכי שיעור הבנות המצטיינות במתמטיקה נמוך יחסית. אחת התאוריות הפופולריות להסבר פערים אלו היא היפותזת השונות הגדולה יותר של בנים - greater male variability hypothesis (Ellis, 1894). הדוגלים במודל זה טוענים כי בשל נסיבות ביולוגיות-גנטיות קיימת שונות גדולה יותר בקרב בנים מאשר בקרב בנות במישורים קוגניטיביים ורגשיים ובהיבטים של תפקודים פיזיים. הכללת מודל זה על כישורי החשיבה המתמטית משמעה כי קיימת שונות גדולה יותר בקרב הבנים בהישגים במתמטיקה. מסיבה זו, יש סיכוי גדול יותר למצוא בנים בשני הקצוות של סקלת היכולת במתמטיקה. מחד גיסא ימצאו יותר בנים מצטיינים במתמטיקה, ומאידך גיסא ימצאו גם יותר בנים חלשים מאוד. אך מאחר שברוב המבחנים במערכות חינוך התלמידים בעלי היכולת הנמוכה מאוד אינם נבחנים כלל, נוצר מצב של קיצוץ הזנב התחתון בהתפלגות של הבנים, ומכאן שנרשם באוכלוסייה הנבחנת יתרון ממוצע לבנים על-פני הבנות.

Mertz ו-Kane (2012) הצליחו (לטענתם) להכריע את הפולמוס המתמשך ולנפץ את "מיתוס הפער הבין-מגדרי במתמטיקה". לדידם, הפערים בין בנים לבנות במתמטיקה אינם נובעים מגורמים מולדים-ביולוגיים, אלא בעיקר מגורמים סביבתיים חברתיים-תרבותיים. מסקנתם זו מתבססת על סקירה של ההישגים במתמטיקה במחקרים בינלאומיים משווים שהועברו בעשרות ארצות ובמספר שכבות גיל (מחקר טימס (TIMSS) 2003 ו-2007 בכיתות ד' ו-ח' ומחקר פיזה (PISA) 2009), ועל נתונים מאולימפיאדת המתמטיקה לאורך עשור ועוד.

לטענתם, אילו היו תקפות התאוריות הטוענות כי הפער הבין-מגדרי במתמטיקה נובע מבסיס ביולוגי, צריך היה למצוא עדות לפערים במתמטיקה לטובת הבנים בכל התרבויות והמדינות, בקרב תלמידים בכל הגילים ולאורך כל השנים. כמו כן, אילו מודל השונות המוגדלת של בנים היה תקף, הרי שיחס השונויות בהישגים במתמטיקה צריך היה להיות גדול מ-1 בכל המקרים הללו. צמד החוקרים הראה כי הממצאים אינם כאלו, ולכן יש לחפש הסברים חלופיים, בעיקר סביבתיים, לגודל הפער הבין-מגדרי במתמטיקה - כגון, שוויון הזדמנויות בחברה ובכלכלה (או היעדרו) והבדלים בהכנסות בין גברים לנשים במדינות השונות. הם ניתחו את הפערים הבין-מגדריים על פי מדדים של שוויון הזדמנויות בחברה ובכלכלה בארצות השונות ואת הקשר בין הפערים במתמטיקה למדדים אלו.

העבודה הנוכחית סוקרת את תמונת המצב של פערי המתמטיקה הבין-מגדריים ויחס השונויות בהישגים במתמטיקה **בקרב תלמידים בישראל**, כפי שהם משתקפים במבחני הישגים מערכתיים לאומיים ובין-לאומיים מן השנים האחרונות. בפרט, העבודה מתבססת על נתונים ממבחני המיצ"ב במתמטיקה מהשנים תשס"ט-תשע"ב לכיתות ה' ו-ח', ועל נתוני טימס 2007 ו-2011 ופיזה 2006 ו-2009 בישראל.

ניתוח הנתונים התבצע על רקע סקירתם המקיפה של Mertz ו-Kane והפרשנויות שהם נתנו לפערים הבין-מגדריים במתמטיקה במחקרים הבינלאומיים טימס ופיזה בארצות השונות. נידונה השאלה האם נתוני ישראל, על הרב-תרבותיות שבה, מתיישבים עם הגישות המוצעות בסקירה זו (תאוריית השונות המוגדלת, תאוריית הפער בשל האי-שוויון בחברה, תאוריית התרבות המוסלמית ועוד). הניתוחים המוצגים הם עבור כלל הנבחנים בארץ, וגם לפי קבוצות אוכלוסייה שונות (מגזרי שפה וקבוצות שונות של רקע חברתי-כלכלי).

## המוטיבציה בביצוע מחקר פיזה - מגמות ושינויים בין מחקר 2006 לבין מחקר 2009

עמיחי רגבי, ענבל רון-קפלן, יואל רפ

הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

מחקר פיזה (PISA) הוא מחקר בין-לאומי השוואתי בתחום החינוך הנערך על ידי ארגון ה-OECD אחת לשלוש שנים ומשתתפות בו כ-70 מדינות, ובהן ישראל. ניתוח ההישגים במחזורי 2006 ו-2009 העלה כי ב-2006 ישראל השיגה ציונים נמוכים, יחסית למדינות ה-OECD, וזאת בכל שלושת תחומי הדעת. ב-2009 עלה הציון של ישראל בתחום אוריינות הקריאה בכרבע סטיית תקן. עם זאת, רמת ההישגים במתמטיקה ובמדעים נותרה דומה. הפער בין ההישגים הישראליים בפועל במבחנים מסוג זה ובין הציפיות להישגים גבוהים יותר, מוביל להשערה כי הישגים נמוכים יחסית אלו נובעים בחלקם, מן המוטיבציה הנמוכה של התלמידים בישראל להתאמץ במבחנים נמוכי סיכון מסוג זה. לפיכך, העלאת המוטיבציה להיבחנות עשויה לתרום להעלאת תוקף המדידה של מבחנים אלו וגם להעלאת ההישגים. לאור השערה זו עשתה הרשות הארצית למדידה והערכה בחינוך (ראמ"ה), האחראית בישראל על העברת מחקרים בינלאומיים מסוג זה, מאמצים לעידוד המוטיבציה להיבחנות לקראת מחזור 2009.

מטרת המחקר הנוכחי היא לבדוק באיזו מידה צלחו פעולות עידוד אלו, וזאת בדמות בדיקת השינויים ברמת המוטיבציה בין 2006 ל-2009 והקשר בין שינויים אלו לבין ההישגים בשני מחזורי המחקר. בעבר הציעה ראמ"ה דרכים למדידת המוטיבציה להיבחנות. אחת הדרכים מבוססת על הפרש שיעור האי-השבות בתחילת המבחן לעומת שיעורן לקראת סופו. במחקר הנוכחי נבדקו השינויים שחלו בהפרש זה בין המקבץ השלישי של שאלות המבחן לבין המקבץ הראשון של השאלות (להלן "מדד המוטיבציה"). הממצאים הראו כי בשני תחומי אוריינות - קריאה ומתמטיקה - חלה עלייה במוטיבציה להיבחנות. בתחום הקריאה הממצא התיישב עם העלייה בהישגים ובתחום המתמטיקה העלייה במוטיבציה לוותה בעליית הישגים קלה בלבד. ממצא זה נמצא בהלימה עם הפעילות שנעשתה בראמ"ה לעידוד המוטיבציה להיבחנות במחזור 2009. בתחום אוריינות המדעים, לעומת זאת, נמצאה ירידה במוטיבציה בביצוע פרקי המדעים במבחן.

הסבר אפשרי לממצאים אלו הוא כי מבין שלושת התחומים, תחום אוריינות הקריאה הוא הרגיש ביותר למוטיבציה להיבחנות, וזאת משום שלא מדובר בתחום שבו צריך לדעת או להספיק חומר. על מנת לשפר את ההישגים בתחום זה, על התלמיד בעיקר להשקיע מאמץ בקריאת הטקסטים והשאלות הקשורות אליהם ובהשבה עליהן. לפיכך, בתחום זה העלייה במוטיבציה להיבחנות מתואמת עם עלייה ניכרת בהישגים, בעוד שבשני התחומים האחרים העלייה במוטיבציה היא תנאי הכרחי אך לא מספיק לעלייה בהישגים. במתמטיקה, השיפור בהישגים מותנה גם בידע מתמטי, כך שמוטיבציה מוגברת בתחום זה תתבטא אמנם בניסיונות רבים יותר לפתור שאלות, אך לא בהכרח בפתרונות נכונים רבים יותר. למרות זאת, עצם הניסיון לפתור, ללא קשר לנכונות הפתרון, מעלה את תוקף המדידה של המבחן, ביחס למצב שבו לא פותרים את השאלה כלל. הממצא בתחום אוריינות המדעים, ההפוך למצופה, דורש הסבר אחר ומעמיק יותר, וידון בעבודה הנוכחית. יש לציין כי מגמה דומה נמצאה במדינות רבות אחרות שהשתתפו במחקר וייתכן שהדבר נובע מאופי מקבצי השאלות הספציפיות בתחום זה שניתנו במחזור 2009, ממיקומם בחוברות המבחן ב-2009 לעומת 2006 ומההקשר שבו הוצגו השאלות.

# מידול אוריינות ההערכה של מורים מתחילים בישראל: האפקט של מרכיבי ההכשרה להוראה, תפיסות ביחס להערכה וחוללות עצמית ביחס להערכת לומדים

עדי לוי-ורד, פאדיה נאסר-אבו אלהיג'א

אוניברסיטת תל-אביב

הערכת לומדים, על מאפייניה ועל השלכותיה, ניצבת בעשור האחרון במוקד דיון סוער בשדה החינוך בארץ ובעולם. מחקרים מצביעים באופן קונסיסטנטי על חוסר ידע, חוסר במיומנויות והכשרה דלה של מורים בנושא הערכת הישגי לומדים. המושג "אוריינות בהערכה" מורכב ממכלול של כשירויות, חלקן מעשיות וחלקן תאורטיות, העוסקות בשאלות 'למה', 'מתי' ו'איך' להעריך. מטרת המחקר הנוכחי להציג ולבחון מודל מבני הקושר בין האוריינות של מורים חדשים בהערכה לבין שלושה מרכיבים מהותיים לצורך רכישתה: א. הכשרת מורים בתחום ההערכה (מבחינת הידע הנרכש, ההזדמנויות להתנסות והמודלינג של המרצים בהערכה); ב. תפיסות המורים ביחס למטרות ההערכה; ו-ג. תחושת המסוגלות של המורים לבצע הערכה ראויה. המדגם כלל 327 נחקרים, מתוכם 187 מורים מתמחים ו-140 מורים בפועל. הנתונים נאספו באמצעות ארבעה כלי מדידה כמותיים: א. מבחן אוריינות בהערכה, ב. שאלונים לאפיון ההכשרה בהערכה, ג. שאלון לבדיקת חוללות עצמית ביחס לביצוע פעולות הערכה, ו-ד. שאלון תפיסות ביחס להערכה.

ממצאי המחקר הצביעו על מידת התאמה גבוהה בין המודל המשוער (המשופר) לנתונים. כלל המשתנים במודל (הכשרה בהערכה, תפיסות וחוללות עצמית) הסבירו במשולב אחוז גבוה ומובהק ( $R^2 = 62\%$ ) מהשונות במשתנה אוריינות בהערכה. נמצא כי לתפיסות הנחקרים ביחס להערכה ולחוללות העצמית יש אפקט חיובי ישיר על האוריינות בהערכה, בעוד שלמרכיבי ההכשרה בהערכה יש הן אפקט חיובי ישיר והן אפקט חיובי עקיף (דרך החוללות העצמית) על האוריינות בהערכה. בנוסף, נמצא קשר חיובי בעוצמה בינונית בין מרכיבי ההכשרה בהערכה לבין תפיסות הערכה ( $r = .44, p < .05$ ). כלומר, ככל שמידת ההכשרה בהערכה (מבחינת הידע, ההתנסות והמודלינג) הייתה רבה יותר, כך תפסו המורים את ההערכה כמשמשת למטרות שיפור הלמידה וההוראה, וכבעלת תקפות ומהימנות; וככל שמידת ההכשרה בהערכה הייתה נמוכה יותר, כך נטו המורים לתפוס את ההערכה כמשמשת למטרות אחריותיות ב"ס, כמזיקה, כלא מדויקת וכבעלת השפעה מועטה על ההוראה.

ממצאי המחקר מספקים תובנות חשובות ביחס להסבר אוריינותם של מורים מתחילים בהערכה. אחוז השונות הגבוה באוריינות ההערכה אשר ניתן להסביר על ידי שילוב בין מרכיבים הכשרתיים, תפיסתיים ואפקטיביים, עשוי לספק תשתית ממשית לחשיבה מחודשת ביחס להכשרת מורים בהערכה ולגיבוש מדיניות מתאימה בנושא.



## אינפלציה בציוני הבגרות בישראל

דוד מעגן, ליבי שפירא

הלשכה המרכזית לסטטיסטיקה

"אינפלציה בציונים" היא מצב שבו במהלך השנים מתרחשת עלייה בציונים אשר אינה משקפת שיפור הולם בידע ובמיומנויות התלמידים. מאחר שציוני בחינות הבגרות אינם מכוילים, על מנת לזהות את התופעה נדרשת בחינה של הציונים לאורך זמן מול מדד "מכיל", חיצוני למערכת, אשר משקף יכולות לאורך זמן. הציון הפסיכומטרי עומד בדרישות אלה, והוא משמש בסיס לניתוח של 12 קוהרטות של זכאי בגרות בשנים 1991-2002, ברמת התלמיד הבודד וברמת בתי הספר, תוך הבחנה בין המגזר העברי למגזר הערבי ותוך פיקוח על מאפייני הרקע.

מטרת המחקר היא לבחון האם ועד כמה קיימת מגמת אינפלציה בציוני הבגרות בישראל. מטרה נוספת היא לבחון האם מגמת האינפלציה בציוני הבגרות משתנה בין בעלי יכולות גבוהות (עפ"י ההישג הפסיכומטרי) לבין בעלי יכולות נמוכות.

אוכלוסיית המחקר כללה 485,036 זכאי בגרות אינטרניים במועד סיום התיכון בשנים 1991-2002. לא נכללו זכאי בגרות בחינוך החרדי, "נבחני משנה" וזכאי בגרות אקסטרניים. לקובץ זה הושלמו נתוני הרקע של הזכאים מקובצי התלמידים לשנים 1991-2002 ונתוני הבחינה הפסיכומטרית הראשונה לכל זכאי הבגרות שנבחנו בבחינה הפסיכומטרית במהלך השנים 1991-2007. לצורך בחינת מגמת העלייה בציוני הבגרות במהלך השנים, שנות מחזור סיום תיכון (1991-2002), הומרו למשתנה מגמת הזמן (T) הנע בין 1 ל-12. ציוני הבגרות הגולמיים הומרו לציוני תקן.

לבחינת ה"אינפלציה" בציוני הבגרות נעזרנו בניתוח בשתי רמות – ברמת הפרט וברמת המוסד. אנו מניחים כי עדות לקיומה של "אינפלציה" בציוני בגרות תמצא אם ברגסיה רב-משתנית – המפקחת על מאפייני הרקע של התלמידים ועל כישוריהם הקוגניטיביים כפי שהם באים לידי ביטוי בציון הפסיכומטרי – מגמת הזמן (T) תהיה מובהקת.

מניתוח ממצאי המודלים ברמת הפרט עולה כי משתנה מגמת הזמן (T) הוא מובהק ( $p < 0.01$ ) בשני המגזרים (עברי וערבי). עוד עולה כי מגמת האינפלציה גבוהה יותר משמעותית במגזר הערבי (0.081) לעומת העברי (0.049). מניתוח ההבדלים במגמת האינפלציה בין בעלי היכולות הגבוהות (Q3) והנמוכות (Q1) עולה כי מגמת האינפלציה גבוהה יותר בקרב בעלי יכולות נמוכות מאשר בקרב בעלי יכולות גבוהות, בשני המגזרים.

ניתוח נוסף ברמת המוסד נועד, בין היתר, להתגבר על בעיית ההטיה העשויה לנבוע ממידת הסלקטיביות של היבחות בבחינה הפסיכומטרית בקרב זכאי בגרות. מניתוח ממצאי המודלים ברמת המוסד עולה כי משתנה מגמת הזמן (T) הוא מובהק ( $p < 0.01$ ). בשני המגזרים (עברי וערבי). עוד עולה כי מגמת האינפלציה גבוהה יותר משמעותית בבתי ספר במגזר הערבי (0.093) מאשר בבתי ספר במגזר העברי (0.053).

## לתקן או לא לתקן? על היתרונות שבשימוש בציונים מתוקננים של פריטי שאלונים ומבחנים

רז ליבוביץ  
המכללה האקדמית ת"א-יפו
אייל גמליאל  
המרכז האקדמי רופין
לירון צוברי  
האוניברסיטה העברית

מדידות טיפוסיות של משתנים במדעי החברה מבוצעות על ידי אגרגציה של מספר פריטים לכדי ציון כולל, שאמור לשקף את המשתנה הנמדד. לרוב, לכל פריט שנבחר לייצג את המשתנה הנמדד יש חשיבות שווה במדידת התכונה. למרות זאת, החישוב בפועל של הציון הכולל באמצעות מיצוע (או סכימה) של ציוני הגלם של הפריטים מעניק להם חשיבות דיפרנציאלית, בשל הממוצעים ו/או הפיזורים השונים של הפריטים. אפשר להבטיח שכל פריט ישפיע על הציון הכולל במידה שווה באמצעות הבאת הפריטים לכדי סולם משותף, כגון המרת ציוני הגלם של הפריטים בציוני תקן.

מחקר זה בחן אמפירית את ההבדלים בין השיטה הנהוגה לחישוב הציון הכולל לבין שיטה חלופית, שבה לפריטים שנבחרו לייצג את התכונה הנמדדת יש חשיבות שווה, באמצעות השוואה של המהימנות, התוקף וההוגנות של הציונים הכוללים, לגבי 64 משתנים שנמדדו באמצעות שאלוני דיווח עצמי אנונימיים ("נמוכי סיכון") ולגבי ארבעה מבחני הישגים בלימודי תואר ראשון ("עתירי סיכון").

עבור כל בסיס נתונים הופקו ציוני גלם של הפריטים המרכיבים את כלי המדידה וציון כולל המבוסס עליהם, וכן ציוני תקן של הפריטים המרכיבים את כלי המדידה וציון כולל המבוסס עליהם.

הממצאים הראו כי מקדמי העקביות הפנימית, שחושבו בהתבסס על ציוני גלם של הפריטים ועל ציוני התקן שלהם, היו דומים הן עבור המשתנים שנמדדו בשאלוני הדיווח העצמי והן עבור ארבעת מבחני הישגים. מקדמי המתאם בין שני סוגי הציונים הכוללים, שחושבו בהתבסס על ציוני גלם של הפריטים ועל ציוני התקן שלהם, לבין ציוני קריטריון ("מקדמי תוקף"), היו אף הם דומים.

עבור כל כלי מדידה חושב ממוצע הפרשים בערך מוחלט בין ציון הדירוג של כל משתתף בהסתמך על הציון הכולל שחושב מציוני הגלם של הפריטים, לבין ציון הדירוג שלו בהסתמך על הציון הכולל שחושב מציוני התקן של הפריטים. ממוצעי הפרשים המוחלטים בין הדירוגים של שני הציונים הכוללים (המאונים) של המשתתפים היו כ-2% עבור שאלוני הדיווח העצמי, וכ-3% עבור מבחני הישגים. בדיקה נוספת העלתה שהפרשים בין הדירוגים היו גבוהים יותר (בערך מוחלט) בקרב משתתפים ובבחנים שציונם היה סמוך לממוצע.

ממצאי מחקר זה מורים כי ליתרון התאורטי של הציון הכולל המבוסס על ציוני תקן של פריטים על פני השיטה הנהוגה לחישוב הציון הכולל (בהתבסס על ציוני הגלם), לא מתלווה יתרון (או חיסרון) פסיכומטרי בהתייחס למקדמי מהימנות ולמקדמי תוקף, בהקשר של שאלוני דיווח עצמי ובהקשר של מבחני הישגים. לעומת זאת, מיקומם היחסי של מרבית המשתתפים השתנה מעט בעקבות אופן חישוב הציון הכולל, הן עבור שאלוני הדיווח עצמי ועוד יותר עבור מבחני הישגים.

אם מחקרי המשך יאששו את ממצאי מחקר זה, הרי שמומלץ יהיה לשקול שוב את הנוהג הקיים של חישוב ציונים כוללים בהתבסס על ציוני גלם של פריטים, ולהמירו בחישוב ציונים כוללים בהתבסס על ציוני תקן של פריטים עבור כלי מדידה עתירי סיכון, וכן עבור כלי מדידה נמוכי סיכון שמתבצעת בהם השוואה בין מדגמים שונים.

## בגרות ופסיכומטרי והקשר ביניהם - על הנטל העודף של מבחנים

טאינה פודלוב, אביאל קרנצלר

הלשכה המרכזית לסטטיסטיקה

שלמה יצחקי

הלשכה המרכזית לסטטיסטיקה והאוניברסיטה העברית

בחינות הבגרות במקצועות ליבה שונים נערכות על ידי משרד החינוך במתכונת ארצית, ומיועדות לדרג תלמידים על פי רמת התמחותם ויכולותיהם בתחומים השונים. גם הבחינה הפסיכומטרית משמשת את מוסדות הלימוד האקדמיים למיון מועמדים לחוגים השונים.

המטרה של מאמר זה היא לבחון באיזו מידה בודק המבחן הפסיכומטרי תחום יכולות אחר מזה שבוחנת בחינת הבגרות באותו התחום. כלומר, המאמר אינו עוסק בשאלת כושר הניבוי של הצלחה בלימודים אקדמיים, אלא מתרכז בשאלה האם שני המבחנים, הבגרות והפסיכומטרי, הם מבחנים לאותו תחום יכולות.

החידוש המתודולוגי בעבודה זו הוא בשיטת המחקר: השאלה שעליה מנסה המאמר לענות היא, כיצד לקבוע ששני מבחנים בודקים אותו תחום יכולות. במילים אחרות, מתי נוכל לומר שהידע והיכולות המאפשרים הצלחה במבחן האחד זהים לידע וליכולות הנדרשים במבחן השני?

המאמר מסביר את שיטת ה-LMA, הכלי הגרפי שפותח לצורך בדיקה אם הקשר בין שני משתנים הוא מונוטוני. קשר מונוטוני בין משתנים משמעו ששימונו של הקשר אינו משתנה לכל אורך התחומים שבהם מוגדרים המשתנים. ההדגמה מבוצעת על ידי השוואת תוצאות מבחנים פסיכומטריים לעומת מבחני בגרות. הכלי הוא אי פרמטרי ועל כן אינו תלוי התפלגות. הכלי מבוסס על הרעיון שהתפלגות הקושי של השאלות במבחן כמוה כהפעלת טרנספורמציה מונוטונית לא יורדת על תוצאות המבחן. אנו בודקים אם הקשר בין תוצאות המבחנים כמוהו כהפעלת טרנספורמציה מונוטונית לא יורדת על הציון של אחד המבחנים. אם הבדיקה תמצא שאכן הקשר הוא מונוטוני, אזי שני המבחנים בודקים אותו סוג יכולת, או שהיכולות מתואמות חזק וזאת כיוון שהתוצאות היו דומות לחזרה על המבחן הראשון עם שינוי בהתפלגות הקושי של השאלות. לעומת זאת, אם הקשר אינו מונוטוני, אזי שני המבחנים בודקים תחומי יכולת שונים.

הנתונים שעליהם התבססנו כללו את הנבחנים בבחינות הבגרות במתמטיקה ובאנגלית בשנים 1999-2000, אשר עמדו בדרישות הסף של האוניברסיטאות והשתתפו במבחן פסיכומטרי לאחר מכן - סה"כ 17,867 נבחנים. על סמך הנתונים שנבדקו, התשובה המתקבלת ממחקר זה היא שהמבחנים בודקים אותו תחום יכולות, אולם קיימת בהם טעות אקראית גדולה. הטעות האקראית במבחני המתמטיקה גדולה יותר מאשר במבחן באנגלית.

## מבוגרים בעלי דיסקלקוליה אינם שונים מקבוצת ביקורת בעיבוד אוטומטי של מידע נומרי בסיסי

יוסף צלגוב, בר זוהר-שי, טלי ליבוביץ
ענת בן סימון
אורלי רובינשטיין

אוניברסיטת בן-גוריון בנגב
מרכז ארצי לבחינות ולהערכה
אוניברסיטת חיפה

דיסקלקוליה התפתחותית היא לקות ביכולת החישובית. יש הטוענים כי לבעלי דיסקלקוליה התפתחותית יש קשיים בעיבוד אוטומטי של מידע מספרי, מה שמבטא מגבלת ייצוג מידע מספרי בזיכרון לטווח ארוך. סמן אחד לעיבוד אוטומטי של מידע מספרי הוא אפקט ה-SNARC, המתבטא בתגובה מהירה יותר למספרים קטנים בצד שמאל, ולמספרים גדולים בצד ימין. אפקט זה מפורש בדרך כלל כתומך בייצוג מנטאלי של מספרים על קו ישר על פי גודלם, משמאל לימין.

סמן נוסף הוא אפקט תואמות הגודל (SiCE) Size Congruity Effect. האפקט מתקבל כאשר נבדקים משוים את הגודל הפיזיקלי של צמדי מספרים השונים בערכם הנומרי ובגודלם הפיזיקלי. הוא מתבטא בזמני תגובה קצרים יותר בתנאי התואם (לדוגמה 3,5) מאשר זמני התגובה בתנאי הלא-תואם (לדוגמה 5,3). ה-SiCE גדל עם המרחק המספרי בין המספרים המשווים, בהתאם לטענה של ייצוג על פני קו ישר, וקטן ככל ששני המספרים גדולים יותר, בהתאמה לטיעון כי הייצוג המנטלי הוא ייצוג דוחס.

במחקר הנוכחי השווה עיבוד אוטומטי של מידע נומרי אצל הלוקים בדיסקלקוליה וקבוצת ביקורת. במחקר השתתפו 29 מבוגרים (מהם 13 פגועי דיסקלקוליה) שאובחנו על פי ביצועם במת"ל כ"תקינים" או כבעלי דיסקלקוליה (מתחת לאחוזון 20 במבחנים הרלוונטיים במת"ל). לא נכללו במדגם בעלי ADHD. חמישה מפגועי הדיסקלקוליה סבלו מדיסגרפיה קלה (רמה 2 במת"ל), ושניים מדיסקלקציה קלה (רמה 2 במת"ל).

שתי קבוצות הראו את אפקט ה-SNARC ואת אפקט ה-SiCE. בשתייהן ה-SiCE קטן עם הגודל הנומרי של המספרים המשווים, וגדל עם הפער הנומרי בערכם. תוצאות אלה מעידות כי פגועי דיסקלקוליה אינם נבדלים מנבדקים תקינים בייצוג מנטלי של מספרים בזיכרון לטווח ארוך, וכך מערערות את הטענה כי דיסקלקוליה מקורה בפגיעה בייצוג זה.

## המבנה הגורמי של תפקודי השפה המשמשים באבחון דיסלקסיה במת"ל

ענת בן-סימון, מאיר ברנרן, נורית ענבר-וייס

מרכז ארצי לבחינות ולהערכה

באבחון טיפוסי של לקות למידה נחשפת המאבחנת לעשרות נתונים, הכוללים, בין השאר, את ההיסטוריה ההתפתחותית והלימודית של הפונה, תוצאות אבחונים קודמים ותוצאות הביצוע במבחנים דיאגנוסטיים. באבחון מבוסס מת"ל נאספים מעל 300 משתנים הכוללים 53 מדדי ביצוע במבחנים דיאגנוסטיים. כיוון שהאבחנה הסופית בנוגע לקיומה של לקות למידה נעשית בסופו של דבר באמצעות שיפוט (שקלול) קליני של כלל המשתנים, אחת הדרכים לייעל את תהליך השיפוט היא לקבץ את מדדי הביצוע במבחני מת"ל לקטגוריות תפקוד. השימוש בקטגוריות תפקוד במקום במדדי ביצוע לצורך גיבוש אבחנה צפוי לא רק להקל על המאבחנת את גיבוש האבחנה אלא גם לתרום למהימנות ולסטנדרטיזציה של הליך קבלת ההחלטות באבחון לקויות למידה, שיש להם חשיבות רבה באבחון המשמש לקביעת זכאות להתאמות.

מטרתו של מחקר זה היא לבחון את המבנה הגורמי של מבחני מת"ל בכלל ושל מבחני השפה בפרט לצורך חשיפת הקשרים בין תפקודי השפה השונים וקבוצות מדדי ביצוע לקטגוריות תפקוד בעלות משמעות תאורטית וביסוס אמפירי. המשתתפים במחקר הם 2,652 מועמדים להשכלה הגבוהה וסטודנטים שפנו למרכזי אבחון מבוססי מת"ל לצורך אבחון לקות למידה ובדיקת הזכאות להתאמות בבחינות. כל המשתתפים אובחנו באמצעות מת"ל בין השנים 2008-2011.

לצורך חקירת המבנה הגורמי של מבחני השפה במת"ל וקבוצת לקטגוריות תפקוד בוצע ניתוח גורמים היררכי על תוצאותיהם של כלל הפונים בכל מבחני מת"ל. בשלב הראשון יושם ניתוח גורמים מגשש על כל מדדי הביצוע של מבחני מת"ל. בשלב השני בוצע ניתוח גורמים היררכי על ציוני השארית שהופקו לאחר בקרה של הגורם הראשי שהתקבל בניתוח הראשון. בשלב השלישי יושם ניתוח שונות מגשש על מבחני השפה בלבד, ולאחר מכן בוצע ניתוח גורמים על ציוני השארית של מבחני השפה לאחר בקרת הגורם הראשי שהתקבל בשלב הראשון של הניתוח. לבסוף, חושבו המתאמים בין גורמי השפה לבין עצמם, ונבדק התוקף המבחני של גורמי התפקוד השפתיים באבחון דיסלקסיה.

בניתוח הגורמים שבוצע על כלל מבחני מת"ל התקבלו 13 גורמים שהסבירו 63% מהשונות. כמו כן, נמצא גורם ראשי בולט שמרבית (44 מתוך 53) מדדי הביצוע היו טעונים עליו. בניתוח הגורמים של מבחני השפה התקבלו שישה גורמי תפקוד שהסבירו 66% מהשונות. הגורמים פורשו באופן הבא: (1) שיום (שליפה לקסיקלית); (2) קידוד פונולוגי; (3) הבנת טקסט עברית/אנגלית (דיוק); (4) הבנת טקסט עברית/אנגלית (קצב); (5) זיכרון אורתוגרפי (קריאה וכתובה של טקסט); ו-(6) שטף מילולי. בניתוח הגורמים היררכי של מבחני השפה התקבלו שמונה גורמי תפקוד שכללו את ששת הגורמים הראשונים ושני גורמים נוספים: מודעות פונולוגית ומודעות תחבירית (ז'ת). כל הגורמים נמצאו מבחינים באופן מובהק בין משתתפים עם דיסלקסיה לבין משתתפים ללא לקות למידה.

ממצאי המחקר מאוששים את תאוריות הליקוי הכפול העומד בבסיס דיסלקסיה (לקות בקידוד פונולוגי ו/או בשליפה לקסיקלית) ומדגישים את הקשר ההדוק בין כל היבטי הכתיבה (איכות כתב היד, הספק הכתיבה ושגיאות הכתיב) ואת הקשר ההדוק בין תהליכי הכתיבה והקריאה המסביר את קשיי הכתיבה של בעלי דיסלקסיה. בנוסף, מציעות תוצאות המחקר גם מודל מבוסס תאורטית ואמפירית לקבוצות משתני השפה לקטגוריות תפקוד שפתי העשוי להקל על הליך גיבוש האבחנה של דיסלקסיה ולתרום לזיהוי המקורות ללקות.

## פרס לעבודת מחקר בתחום המדידה וההערכה לשנת 2013

### הערכת עמיתים בקרב סטודנטים לרפואה ככלי לבחינת תוקף הניבוי של מבחני הקבלה הלא-קוגניטיביים (מרק"ם) ללימודי רפואה באוניברסיטה העברית

אביאל חדד, האוניברסיטה העברית

בהנחיית אבי ישראל ונעמי גפני

רקע: במרבית האוניברסיטאות בישראל שונו מבחני הקבלה ללימודי רפואה בשנים האחרונות כך שיכללו גם מבחנים לא קוגניטיביים: מבחני מרק"ם (מערכת ראיונות קצרים מובנים) ומו"ר (מערכת מיון ללימודי רפואה), המבוססים בעיקרם על שיטת MMI (Multiple Mini Interview), שפותחה באוניברסיטת מקמסטר שבקנדה, ולפיה המועמדים ללימודי רפואה עוברים סדרה של 8 ראיונות קצרים מובנים. במבחני מו"ר ומרק"ם, בנוסף למערכת הראיונות הקצרים, המועמדים מתבקשים למלא שאלון ביוגרפי ושאלון שיפוט והחלטה. תוקף הניבוי של מבחני MMI נבדק בצורה חלקית במספר מצומצם של מחקרים ברחבי העולם.

מטרה: מטרת המחקר היא בחינת תוקף מבחני מרק"ם ומו"ר בכלל ומבחן MMI בפרט.

שיטה: בהתבסס על שאלון ממחקר קודם, נבנה שאלון הערכת עמיתים אינטרנטי. 99 סטודנטים לרפואה באוניברסיטה העברית בירושלים עברו מבחני הערכת עמיתים בסיום השנה הרביעית ללימודיהם. נבדקו מקדמי המתאם בין תוצאות מבחני הערכת העמיתים לבין תוצאות מבחני מרק"ם ומו"ר שעברו הסטודנטים טרם קבלתם ללימודי רפואה.

תוצאות: נמצאה קורלציה מתוקנת של  $r=0.39$  ושל  $r=0.37$  בין מבחני הערכת העמיתים לבין הציון המסכם במבחני מרק"ם ומו"ר בהתאמה. נמצאה קורלציה חלשה ולא משמעותית מבחינה סטטיסטית בין מבחני הערכת העמיתים לבין ביצוע שאלון שיפוט והחלטה (0.07) ושאלון ביוגרפי (0.13).

מסקנות: תוקף מבחני MMI הודגם במחקר זה ומצדיק את השימוש בהם בתהליך הקבלה ללימודי רפואה. נדרשים מחקרים נוספים לבדיקת תוקף מבחני MMI.

## פרס לעבודת מחקר בתחום המדידה והערכה לשנת 2013

### פיתוח כלי מבוסס יכולת למדידת אינטליגנציה רגשית-חברתית

יוסי חסון, המכללה האקדמית ת"א-יפו

בהנחיית דלית לב-ארי

במהלך שני העשורים האחרונים, המושג "אינטליגנציה רגשית" (א"ר) עורר עניין רב בספרות הפופולרית והמחקרית. מושגים, מודלים וכלי מדידה שונים של א"ר פותחו ונלמדו בתחומי מחקר רבים, ויושמו במקומות שונים, למשל בבתי ספר ובמקומות עבודה (Ciarrochi, Forgas & Mayer, 2006). שלושת המודלים העיקריים העוסקים בא"ר פותחו על ידי מאייר וסאלוביי (1993), בר-און (1997) וגולמן (1995). בין שלושת המודלים אפשר למצוא דימיון מועט בלבד, ובעיקר חוסר הסכמה בנוגע להגדרת המושג א"ר והאופן הנכון למדוד אותה. בחינה של שלושת המודלים מעלה שהמושג א"ר סובל מבעיות קונספטואליות ומתודולוגיות.

מטרת המחקר הנוכחי היא לבחון האם מתודת מדידה חדשה עשויה לחזק את האספקטים הפסיכומטריים של מודל האינטליגנציה הרגשית-חברתית של בר-און (Bar-On, 1997) ולסייע בפתרון בעיות של כלי מדידה אחרים של א"ר.

במסגרת המחקר פותח מבחן מבוסס יכולת למדידת אינטליגנציה רגשית-חברתית. על מנת לבדוק את תוקף המבנה של המבחן ואת מהימנותו, נבחנו קשריו עם מבחנים מתחומים אחרים, ותרומתו לחיזוק מאפייניו הפסיכומטריים של המודל של בר-און. המחקר כלל מדגם של 95 מועמדים ליחידה קרבית עילית בצה"ל שהשלימו סדרה של מבחנים: שני מבחני אינטליגנציה כללית (מנבא דירוג פסיכוטכני ראשוני ו-BMCT), מבחן אישיות (BTS), מבחן דיווח עצמי של אינטליגנציה רגשית-חברתית (EQ-i) והמבחן מבוסס היכולת החדש.

מתוצאות המחקר עולה כי המבחן החדש מהימן, ונמצא בקורלציה סבירה עם מדדי אינטליגנציה כללית. כמו כן, נמצאה קורלציה עם מדדי אישיות, אך במידה נמוכה מזו שנמצאה במבחן EQ-i. ממצאים אלה מעידים על תוקף מבחן טוב יותר של המבחן החדש בהשוואה למבחן EQ-i. בנוסף, נמצא קשר נמוך אך מובהק בין המבחן החדש למבחן EQ-i, והדבר תומך באופן חלקי בתוקף מתכנס.

## מבט על השימוש במבחנים פסיכולוגיים בישראל והאתגרים החדשים

סול פיין

חברת מידות

מבחנים פסיכולוגיים נמצאים בשימוש נרחב בישראל בזירות קליניות, חינוכיות וארגוניות. בהתאם לכך, עוברים פסיכולוגים בישראל הכשרות מקצועיות מקיפות בנושא השימוש הנכון במבחנים הן במהלך לימודיהם המתקדמים והן במהלך התמחויותיהם, כחלק מתהליך הסמכתם כמומחים בתחום. עם זאת, בישראל, בשונה מהמצב השורר במדינות אירופיות לא מעטות ובארה"ב, לא קיימות הסמכות לפסיכולוגים בתחום השימוש במבחנים, לא חלה חובה על פסיכולוגים להמשיך וללמוד בתחומיהם לאחר שהוסמכו כמומחים ורק בודדים מהמבחנים הפסיכולוגים המתפרסמים עוברים בדיקות ובקרת איכות ושימוש בלתי תלויים.

הפיקוח המקל יחסית בתחום המבחנים הפסיכולוגיים בישראל עלול ליצור תמונה מעורפלת בנוגע לאופן השימוש בהם בפועל. מחקר זה סקר מדגם של פסיכולוגים בישראל (N = 338), העוסקים במגוון תחומים המצריכים שימוש במבחנים פסיכולוגיים. בהקשר דומה, בשנת 2000 ה- European Federation of Psychologists' Associations (EFPA) סקר עמדות של פסיכולוגים משש מדינות אירופיות שונות בנושאים הקשורים למבחנים והשימוש בהם (Muniz et al., 2001). יותר מעשור לאחר פרסום סקר זה, EFPA מצאה לנכון לערוך סקר חוזר ולהרחיבו לכ-17 מדינות באירופה. המחקר הנוכחי, שנערך בישראל, מהווה חלק מהסקר האירופי המורחב הזה. הרצאה זו תסכם את ממצאיו, תוך התייחסות לממצאי הסקר המקורי מ-2001.

באופן כללי, ממצאי המחקר מעידים על השימוש הנרחב במבחנים פסיכולוגיים בישראל ועל ההערכה הרבה השוררת בקרב פסיכולוגים כלפיהם, במידה דומה לעמיתיהם באירופה. עם זאת, ממצאי המחקר מצביעים על שורה של נקודות תורפה המציבות אתגרים ספציפיים לתחום המבחנים הפסיכולוגיים בישראל ועל הצורך בפיקוח מקצועי הדוק יותר כדי להבטיח את רמתם המקצועית של המשתמשים בהם ואת איכותם של המבחנים עצמם.

על סמך הממצאים מובאות מספר המלצות אופרטיביות המיועדות לקידום התחום בישראל. המלצות אלו עוסקות, בין היתר, בסטנדרטים מקצועיים בני אכיפה לשימוש במבחנים, במסלול הסמכה פורמלית לפסיכולוגים (ולמשתמשים אחרים) במבחנים, בחובת ההשתלמות לעוסקים בתחום, בפרסום במידע טכני על המבחנים שנמצאים בשימוש בישראל, בטיפול בתלונות על שימוש לא נאות במבחנים ועוד. יש לקוות כי מחקר זה והמלצותיו יתרמו להגברת המודעות לנושא המבחנים הפסיכולוגיים בישראל, הן בקרב המשתמשים בהם והן בציבור הרחב.



## פיתוח מודל לסיווג עולם המקצועות עבור מיון תעסוקתי הבודק יכולת ביטוי בכתב

אבי אללוף  
מרכז ארצי לבחינות ולהערכה

הילה שורק-פוזס  
המכללה האקדמית תל-אביב יפו

יכולת ביטוי בכתב חשובה הן במערכת החינוך והן בעולם התעסוקה. בשנים האחרונות התחזקה ההכרה בחשיבות יכולת ההבעה בכתב בהשכלה הגבוהה, והדבר בא לידי ביטוי בהכללת פרק כתיבה במבחני מיון למוסדות להשכלה גבוהה בארה"ב (למשל SAT, ACT, GRE) ולאחרונה בבחינה הפסיכומטרית בישראל. הדגשת יכולת הכתיבה באה גם כתגובה לירידה ביכולת הביטוי בכתב באוכלוסייה הצעירה. תהליכי המיון התעסוקתי הקיימים כיום בישראל נעזרים בכלים המועברים בפורמט כתוב לצורך התרשמות כללית מיכולת הכתיבה של המועמד, אולם מדובר בכלים שפותחו למטרות אחרות. למרות החשיבות שיש ליכולת הביטוי בכתב גם בעולם התעסוקה, לא פותח כלי ייעודי לכך שיאפשר להעריכה באופן שיטתי וסטנדרטי. כשלב מקדים לפיתוח כלי מיון כזה יש להמשיג מהם סגנונות הכתיבה התעסוקתיים במקצועות שבהם יכולת הכתיבה רלוונטית.

מטרת המחקר הנוכחי הייתה לנסח מודל המסווג את עולם המקצועות מנקודת מבט העוסקת בכתיבה בעולם התעסוקתי. התאוריות המקובלות בתחום לסיווג מקצועות, שפותחו על ידי אן רואו (1956) וג'ון הולנד (1966), מסווגות את עולם המקצועות לשדות או לתחומים המשקפים עולמות תוכן מקצועיים. קרבה תוכנית זו בין מקצועות עשויה להעיד על סגנון כתיבה בעל מאפיינים דומים בקרב מקצועות מסוימים. המודל שפותח במחקר הנוכחי נועד לשקף הבחנה בין מאפיינים של יכולת ביטוי בכתב כפי שהם מתבטאים בתפקידים ובמקצועות שונים, תוך הסתייעות בתאוריות אלו.

בשלב הראשון נבנה מודל סיווג התחלתי. באמצעות מאגר ניתוחי העיסוקים האינטרנטי O\*NET (The Occupational Information Network) אותרו כל המקצועות שבהם נדרשת יכולת ביטוי בכתב מסוג כלשהו. לאחר מכן, קובצו המקצועות לפי הקרבה התוכנית שביניהם, בהתאם לסיווג המקצועות של רואו ושל הולנד.

בשלב השני נבדקה תקפותו של הסיווג שהתקבל למול הערכות מומחים. כל מומחה התבקש לציין מהם מאפייני הכתיבה הנדרשים בכל אחד מהמקצועות שהניב תהליך האיתור. הנתונים נותחו בשיטת "Bottom-Up" ובאופן דומה לאסטרטגיה שבבסיס של ניתוח גורמים: ראשית, רוכזה רשימת מאפיינים של יכולת ביטוי בכתב. שנית, הוערך משקלו של כל מאפיין בכל מקצוע לחוד ובכל המקצועות יחדיו. לבסוף, אותרה התקבצות של מקצועות לפי שיפוט מידת החפיפה בין המאפיינים הייחודיים של המקצועות. המודל הסופי שיקף את ההתקבצות השיטתית של המקצועות, לפי מאפייני יכולת הביטוי בכתב, אשר נמצאה אופטימלית מבחינת דמיון בין עולמות התוכן של המקצועות ומבחינת דמיון בין מאפייני הכתיבה הנדרשים בהם. נמצא דמיון רב בין המודל ההתחלתי לבין המודל הסופי. עם זאת, מספר קטגוריות איחדו מקצועות מעולמות תוכן שונים בשל מאפייני כתיבה דומים. עולה כי ממד הקרבה התוכנית בין מקצועות משקף סגנונות כתיבה שונים בעולם העבודה, אולם יש סגנונות כתיבה אשר אופייניים למקצועות מסוימים גם ללא קשר לתוכנם.

מודל הסיווג שהוצע במחקר זה מוסיף על סיווגים קיימים של עולם המקצועות. בהיבט היישומי, ניתן להשתמש במודל ובהבחנות שהוא מציע בין תוכן לבין מאפייני כתיבה רלוונטיים לצורך פיתוח מבחנים ייעודיים, שניתן יהיה להעריכם באופן שיטתי ואשר ישימו דגש, לפי בחירת הגורם המפתח, על תכנים ו/או על מיומנויות כתיבה רלוונטיים למקצועות שונים.

## השפעת המגדר והמוצא האתני של המשתתפים והמעריכים על ההצלחה במרכז ההערכה

אמיר פלק, שאול פוקס

אוניברסיטת בר-אילן

עקרון הסטנדרטיזציה במיון דורש שכל המשתתפים ייבחנו בתנאים זהים, וזאת במטרה לשמור על האובייקטיביות, ההוגנות והתוקף של הכלי. למרות שנהוג לייחס למרכזי הערכה רמות גבוהות של תוקף והוגנות (Thornton & Rupp, 2006), נטען שתוצאותיהם תלויות בגורמים שונים שאנשי המקצוע אינם מודעים להם, אשר פוגמים בסטנדרטיזציה שלהם (Fox & Haboucha, 2011). נושא זה דורש התייחסות מחקרית לאור ההשלכות של היעדר סטנדרטיזציה על הוגנות מרכזי ההערכה. המחקר הנוכחי עונה על הצורך הזה, ובוחן את האופנים שבהם פרופורציות המגדר והמוצא האתני (יהודים לעומת בני מיעוטים) של כלל המשתתפים והמעריכים במרכז משפיעות על סיכויי ההצלחה של כל משתתף.

מחקרים מצאו שנשים מצליחות במרכזי הערכה יותר מגברים ושאמריקאים לבנים מצליחים יותר מאפרו-אמריקאים או מהיספנים (Dean, Roth & Bobko, 2008). להבדיל מהמחקרים המתמקדים במגדר ובמוצא האתני של המשתתף עצמו, המחקר הנוכחי בחן כיצד סיכויי ההצלחה של משתתפים בעלי רקע שונה (יהודים ובני מיעוטים, נשים וגברים) הושפעו מההרכב המגדרי והמגזרי של כלל המשתתפים והמעריכים במרכז. לפי Thornton & Gibbons (2009), הרכב המגדר והרקע האתני של הקבוצה עלולים להשפיע על התפקוד הממשי או הנתפס של כל אחד מחבריה. עם זאת, המחקרים המעטים שבחנו הנחה זו לא הניבו תוצאות עקביות.

הנתונים נלקחו ממרכז הערכה הנערך מטעם משרד החינוך, למיון מועמדים לתפקידים ניהוליים בדרג זוטרה. המרכז הורכב ממספר תרגילים במערך קבוצתי. ביצועי המשתתפים בכל תרגיל הוערכו על ידי צוות מעריכים בעלי תפקידים בכירים במערכת החינוך ועל ידי פסיכולוג המרכז, אשר קבע את ההערכה הסופית. המועמדים הוערכו בתשעה ממדים (כגון מטען ערכי ומנהיגות), והציון המסכם הורכב ממיצוע הציונים בתשעת הממדים. נאספו נתונים עבור 384 מועמדים שנטלו חלק ב-44 מרכזים בין השנים 2005-2010. כמחצית מהמשתתפים היו גברים, וכ-20% מהם היו בני מיעוטים.

נמצא שציוני הנשים היו גבוהים מציוני הגברים, וציוני היהודים היו גבוהים מציוני בני המיעוטים. עוד נמצא קשר חיובי בין שיעור הנשים המשתתפות לבין ממוצע הציונים של הגברים במרכז. ככל ששיעור הנשים במרכז היה גבוה יותר (קרי, שיעור הגברים היה נמוך יותר), כך ציוני הגברים היו גבוהים יותר. האפקט של שיעור הנשים המשתתפות נמצא באינטראקציה עם שיעור הנשים המעריכות. ככל ששיעור הנשים המעריכות היה גבוה יותר (כלומר, שיעור הגברים המעריכים היה נמוך יותר), כך הקשר החיובי בין שיעור הנשים המשתתפות לבין ציוני הגברים היה חזק יותר. להבדיל מציוני הגברים, ציוני הנשים לא נמצאו קשורים לשיעורי המגדר של המשתתפים או של המעריכים. בנוסף לשיעור המגדר, גם שיעורי המוצא האתני נמצאו קשורים לציוני המשתתפים: ככל ששיעור המשתתפים בני המיעוטים היה גבוה יותר (קרי, שיעור המשתתפים היהודים היה נמוך יותר), כך ציוניהם של המשתתפים היהודים היו גבוהים יותר.

התוצאות שקיבלנו מצביעות על הבעיות הטמונות בהיעדר סטנדרטיזציה בהרכב הקבוצתי של המשתתפים והמעריכים במרכז ההערכה. הקושי לתכנן את המרכז כך שלא יושפע מגורמים אלו מהווה מגבלה על התוקף וההוגנות של מרכזי ההערכה. המחקר הנוכחי מצביע על הצורך להקדיש תשומת לב רבה יותר לנושא זה מצד חוקרים ואנשי מקצוע.

## בחנת מיומנות ידנית ברפואת שיניים באמצעות סימולטור מדמה מציאות

גלעד בן-גל, רומינה סרנה, ארוין וייס      נעמי גפני      אמתי זיו  
הפקולטה לרפואת שיניים, הדסה      מרכז ארצי לבחינות ולהערכה      מסר - המרכז לסימולציה רפואית

עבודתו של רופא השיניים דורשת הן ידע תאורטי נרחב ועדכני בתחום, והן מיומנות ידנית עדינה המאפשרת ביצוע פרוצדורות בפה המתרפא. על מנת לרכוש מיומנות זו במהלך הלימודים וההכשרה נהוג לדמות את סביבת הפה המתרפא בעזרת סימולציה, שבה הסטודנטים מבצעים פרוצדורות על גבי בובות עם לסתות ושיני אקריל. הערכת הביצוע מתבצעת על ידי המדריכים ומאופיינת בחוסר עקביות בין הערכות של אותו מעריך ובין מעריכים שונים. תופעה זו נובעת משימוש במדדים קשים למדידה. בנוסף, בדיקה כזו מדרגת ובודקת את התוצאה הסופית בלבד, ולא את הדיוק והיעילות בתהליך העבודה.

סימולטורים ממוחשבים של מציאות מדומה משמשים יותר ויותר בהוראה של תחומים שונים ברפואה, ובראשם הכירורגיה. לאחרונה פותחו סימולטורי חישה (Haptic) גם בתחום רפואת השיניים. סימולטור חישה מסוגל להחזיר משוב תחושת ליד המשתמש, ובעזרתו הופכת חוויית האימון למציאותית יותר. הסימולטור כולל מעין עט מחובר לזרוע, המספקת למשתמש משוב המבוסס על תמונה תלת ממדית המוצגת על גבי מסך. המשתמש מתבקש לבצע משימות קידוח מדומות, כולל קידוח דרך מראה וירטואלית בראייה בלתי ישירה, בדומה לנדרש בעבודה הקלינית. הקידוח מתבטא בשינויים ויזואליים בתמונה התלת ממדית הנצפית במסך.

הערכת הביצוע מתבצעת באופן אוטומטי ע"י הסימולטור, המערכת רושמת את זמן העבודה ואת רמת הדיוק. על מנת לבדוק את התכונות הפסיכומטריות של הערכה זו נערכה סדרה של ניסויים שבהם נבדקו היבטים שונים שלה: מהימנות פנימית, מידת הרלבנטיות (תוקף תוכן), תוקף ניבוי ותוקף מבנה.

בשלב ראשון נבדקה מידת הרלוונטיות של השימוש בסימולטור וקרבתה של מערכת ההדמיה לעולם האמתי, וזאת בעזרת שאלונים שנענו על ידי מומחים וסטודנטים בעקבות התנסות במערכת (N=33). תוצאות השאלונים הראו כי המשתתפים העריכו שלסימולטור יש פוטנציאל משמעותי בהוראה ובהערכה של מיומנות ידנית ברפואת שיניים, וכי המיומנות שהוא מדמה רלוונטית לרפואת שיניים.

לאור התוצאות הוחלט לבנות מבחן בסימולטור. לצורך כך נערך מבחן חלוץ שכלל 12 משימות והועבר ל-106 משתתפים, ובעקבותיו הורכב מבחן סופי שכלל אף הוא 12 משימות. המבחן הועבר ל-142 משתתפים, אשר חולקו לשלוש קבוצות בהתאם לניסיון הדנטלי שלהם: 88 סטודנטים לרפואת שיניים (שנים א,ד,ו), 37 רופאי שיניים ו-17 משתתפים שאינם מתחום רפואת השיניים.

חושבה מהימנות המבחן, ונבדק הקשר בין איכות הביצוע לבין ניסיון דנטלי וציון על עבודה מעשית בתום שנה א. נמצא כי למבחן מהימנות פנימית גבוהה: Cronbach Alpha = 0.82. המתאם (Spearman Rank Correlation) שנמצא בין הניסיון הדנטלי לבין הביצוע במבחן הוא 0.47.

בבדיקת הקשר בין ציוני הסימולטור עבור 46 תלמידי שנה א עם ציוני מבחן קידוח מעשי סטנדרטי (מדד לתוקף הניבוי), התקבל מתאם Pearson של 0.33 ( $P < 0.05$ ).

תוצאות הניסויים תומכות בתוקף התוכן, בתוקף המבנה ובתוקף הניבוי של מבחן באמצעות הסימולטור. נראה כי אפשר להשתמש בסימולטורים ככלי אובייקטיבי, סטנדרטי, אמין והוגן לבחינה של מיומנות ידנית ברפואת שיניים, ובכך לתרום לשיפור האימון והמקצועיות של סטודנטים לרפואת שיניים.

## דעת הקהל והתוקף הנראה של הבחינה הפסיכומטרית

צור קרליץ, עדו ליטמנוביץ'

מרכז ארצי לבחינות ולהערכה

מבחן הכניסה הפסיכומטרי לאוניברסיטאות (מכפ"ל) נוצר ומנוהל על ידי המרכז הארצי לבחינות ולהערכה (מאל"ו) על מנת להסדיר את המיון לאוניברסיטאות. למרות שהוא משרת היטב את לקוחותיו, מכפ"ל סופג ביקורת רבה מהציבור והתקשורת, והצעות חוק לביטולו או לשינויו עולות תדיר. ביקורת זו והשלכותיה הן הרקע למחקר זה, הראשון מסוגו, הסוקר באופן נרחב ומעמיק את האמונות והדעות בציבור ביחס למכפ"ל. לדעתנו, היכרות עם תפיסות הציבור את מכפ"ל יכולה לסייע לשיפור הבחינה, לחזק את קבלתה של הבחינה בציבור ולכוון את הדיון הציבורי לאפיקים חיוביים יותר.

נבו (1985) טוען שדעת הקהל משקפת את התוקף הנראה של הבחינה. בהתאם לכך, המחקר הנוכחי נועד לאסוף עדויות מהציבור על מנת לאשש או להפריך את טיעון התוקף הנראה של מכפ"ל. תוקף נראה הוא המידה שבה המבחן נתפס, על פניו, כמתאים למטרות שלשמן נוצר. במקרה של מכפ"ל מדובר במידה שבה ציוני נתפסים על ידי הציבור כרלוונטיים למדידת הצלחה בלימודים אקדמיים. למרות שמדידת אמונות ודעות היא כלי פופולרי בתחומי מחקר רבים, היא כמעט ואינה נמצאת בשימוש בתחום ההערכה בחינוך. זאת מכיוון שהתוקף הנראה אינו מבוסס על מודל סטטיסטי ואינו מתייחס למבחן באופן ישיר, אלא בוחן תפיסות סובייקטיביות של הדיוטות (Downing, 2006). אנו נציע הגדרה מחודשת לתוקף הנראה המושתתת על תפיסת התוקף כאוסף טענות המובילות מהתוצאות הנצפות של הנבחנים למסקנות עליהם, לפי מערך טיעון התוקף שכולל שני מרכיבים (Kane, 2006):

**טיעון פרשני** – רשת ההנחות והמסקנות העומדות בבסיס הפירושים והשימושים של ציוני המבחן, וכוללות הנחות ומסקנות לגבי ציונו, הכללה, השלכה והחלטה.

**טיעון תוקף** – עדויות המשמשות להערכת הטיעון הפרשני, השוואתו לטיעונים אלטרנטיביים וקבלת החלטה לגבי אישוש או הפרכתו.

על בסיס זה, הגדרתנו לתוקף נראה היא: "למבחן יש תוקף נראה אם אנשים חושבים שהטיעון הפרשני של המבחן הוא ברור, המסקנות שלו סבירות וההנחות שלו מתקבלות על הדעת". לצורך המחקר, ניסחנו שמונה טענות הנוגעות לתוקף הנראה של מבחני כניסה למערכת ההשכלה הגבוהה. המוקד של המחקר הוא תיקוף המבחן מבחינת המידה שבה הציבור מקבל את הטענות הללו ואת העובדות שבבסיסן.

המחקר כלל שאלון אינטרנטי המורכב מכ-70 שאלות שנבנו על בסיס טענות שהועלו נגד מכפ"ל בעשורים האחרונים בתקשורת, בהצעות חוק ובפניות למאל"ו. לשאלון השיבו 11,665 משתתפים (כ-16% הענות), מרביתם מועמדים למערכת ההשכלה הגבוהה וסטודנטים.

תוצאות המחקר סיפקו למאל"ו תמונת מצב עדכנית וברורה של עמדות הציבור לגבי הבחינה, לקראת כניסתה של המתכונת החדשה. מרבית הממצאים תואמים את ציפיות החוקרים, המבוססות על הדיון הציבורי והתקשורתי בבחינה. אולם, הממצאים שיוצגו בכנס מראים ששינוי נקודתי של מרכיבים הפוגעים בתוקף הנראה של הבחינה עשוי לשפר את תפיסת מכפ"ל על ידי הציבור.

## כישורים לימודיים ודפוסי קבלת החלטה בבחירת לימודים - ממצאים ממחקר על נשירה במערכת ההשכלה הגבוהה

איתמר גתי  
האוניברסיטה העברית

תמי קנת-כהן  
מרכז ארצי לבחינות ולהערכה

מטרתו של מחקר זה הייתה להעשיר את הידע על גורמי הנשירה במערכת ההשכלה הגבוהה בישראל. המחקר ניגש לנושא הנשירה במערכת ההשכלה הגבוהה מזווית חדשה - הן בהיבט של הגדרת נשירה והן בהיבט של המשתנים המנבאים אותה. בהגדרת הנשירה כלל המחקר לא רק נשירה מן המוסד, שהיא הקריטריון השכיח במחקרי הנשירה בארץ ובעולם, אלא גם נשירה מן החוג (החלפת חוגים). מחקרים על גורמי הנשירה התמקדו עד עתה בכישורים לימודיים ובמשתנים דמוגרפיים, וכן במגוון רחב של משתנים פסיכו-סוציאליים, הקשורים בעיקר להתנסות הלימודית והחברתית בזמן הלימודים. מחקר זה הוסיף לחקר הנשירה קבוצת משתנים חדשה - דפוסי קבלת החלטה בבחירת לימודים.

המחקר התבסס על נתוני שלושה מדגמים של מתחילי לימודים בשתי אוניברסיטאות - 705 בשנה"ל תש"ע ו-875 בשנה"ל תשע"א במוסד א, ו-304 בשנה"ל תשע"א במוסד ב - שמילאו שאלון מחקר באינטרנט סמוך למועד הרשמתם ללימודים. השאלון בדק מאפיינים אחדים של דפוסי קבלת החלטה בבחירת לימודים, כגון פרופיל/סגנון קבלת החלטות בבחירת מקצוע או אי-החלטיות, וכן היבטים שונים של תהליך קבלת החלטת ההרשמה, כגון מידת ההשקעה בתהליך או קיומה של ברירה עצמית. המחקר עקב אחר התמדה לעומת נשירה במעבר משנה א' לשנה ב'. "התמדה" הוגדרה התמדה בכל החוגים או עזיבת חוג אחד משניים (מעבר למסלול חד-חוגי), ו"נשירה" הוגדרה החלפת חוג אחד משניים, החלפת כל החוגים או נשירה מן המוסד. שיעור הנושרים על פי הגדרה זו היה 20% ו-18% בשני המדגמים של מוסד א, ו-14% במדגם של מוסד ב.

במחקר נעשתה הבחנה בין נשירה מתוכננת (למשל, נשירה המבוססת על התכנון ללמוד שנה א' בביוולוגיה כדי להתקבל ולעבור ללימודי רפואה) ובין נשירה לא מתוכננת, ובתוך נשירה לא מתוכננת - בין נשירה מרצון ובין נשירה מכורח (נשירה המלווה בכישלון אקדמי).

במוקד המחקר עמד הניסיון לנבא נשירה מרצון על ידי מאפייני הנרשם - כישורים לימודיים (ציון מכפ"ל וציון הברגרות), דפוסי קבלת החלטה בבחירת לימודים ומשתנים דמוגרפיים (גיל ומגדר). ניתוח רב-משתני לניבוי התמדה לעומת נשירה מרצון הראה שהן הכישורים הלימודיים והן דפוסי קבלת החלטה קשורים באופן עקיף למשתנה תלוי זה. מודל רב-משתני לניבוי התמדה לעומת נשירה מרצון על ידי משתני המחקר הסביר 6% מהשונות במשתנה זה. ניתוח רב-משתני שאינו כולל את ציון שנה א', שעשוי להיות רלוונטי יותר לצרכים מעשיים, היה אחוז השונות המוסברת 3%. במטרה ללמוד יותר על תופעת החלפת החוגים הופנתה שאלה פתוחה למחליפי חוגים (מרצון) שאותרו במחקר על הסיבות שבעטיין החליפו חוגים במעבר משנה א' לשנה ב'. ניתוח איכותני של תגובותיהם של 42 מחליפי החוגים המחיש שהחלפת חוגים, כהיבט של נשירה במערכת ההשכלה הגבוהה, היא התנהגות מורכבת ולא קלה לניבוי.

במחקר נבדק גם הניבוי של התמדה לעומת נשירה מכורח על ידי הכישורים הלימודיים. 8%-11% מהשונות של משתנה תלוי זה הוסברו על ידי הכישורים הלימודיים.

## הוגנות הליך הקבלה להשכלה הגבוהה כלפי מועמדים מבקשי התאמות

דביר קלפר, אליוט טורוול, תמי קנת-כהן, כרמל אורן

מרכז ארצי לבחינות ולהערכה

עבודה זו חוקרת את הוגנות הליך הקבלה להשכלה הגבוהה בישראל ביחס למועמדים המבקשים התאמות במבחן הכניסה הפסיכומטרי לאוניברסיטאות (מכפ"ל). מחקר זה מתעניין לא רק באלה שקיבלו התאמות, אלא גם באלה שבקשתם לא התקבלה, אם מסיבה טכנית (לא צירפו את כל המסמכים הדרושים) אם על בסיס שיפוט מקצועי.

התאמות הן שינוי באמות המידה הסטנדרטיות, ומטרתן להפוך את "מגרש המשחקים" להוגן יותר כלפי אוכלוסיות בעלות לקויות, באמצעות הפחתת השונות הלא רלוונטית הנגרמת מהלקויות. בחינה מותאמת שניתנת בצורה תקפה אמורה למדוד בקרב נבחנים בעלי לקויות אותה תכונה שנמדדת על ידי הבחינה הסטנדרטית בקרב האוכלוסייה הרגילה (Tindal & Fuchs, 1999).

אוכלוסיית מבקשי ההתאמות (בשנים 2000–2008) חולקה לחמש תת קבוצות:

- שלוש קבוצות של מקבלי התאמות: בעלי לקות למידה (N=958), בעלי בעיות קשב וריכוז (N=187) ובעלי בעיות פיזיות (N=1,096).
- שתי קבוצות של מבקשי התאמות שבקשתם לא התקבלה: חסרי נתונים (אי-הגשת כל המסמכים הדרושים בזמן ו/או בקשה שהוגשה מאוחר מדי, N=299) ומי שבקשתם לא התקבלה מסיבה מקצועית (אלה שבקשתם נבחנה על-ידי צוות מומחים ונדחתה, N=1,458).

בספרות הוצעו הגדרות שונות להטיה בברירה. כפי שמציינים מרבית החוקרים בתחום, השוני בין ההגדרות נובע משוני בשיפוטם ערכיים. מרבית החוקרים בתחום מסכימים שיש לקשר את בדיקת ההוגנות של תהליך המיון עם הקריטריון שלמענו הוא נוצר. לפיכך, נבדקה במחקר הנוכחי הוגנות מערכת המיון לאוניברסיטאות בהתייחס לקריטריון של הצלחה בלימודי שנה א' באוניברסיטה.

שאלת ההוגנות נבדקה משני היבטים: ההיבט של הטיה בברירה וההיבט של תוקף דיפרנציאלי. הטיה בברירה נבדקה בעזרת מודל הרגרסיה של קלירי (Cleary, 1968), שלפיו הטיה מוגדרת כניבוי דיפרנציאלי. השימוש במודל רווח במחקרים לאיתור הטיה בברירה (AERA/APA/NCME, 1999; Young, 2001). לצורך בדיקת התוקף הדיפרנציאלי נעשה שימוש במקדמי מתאם.

אחד הקשיים המרכזיים במחקרים העוסקים בבעלי לקויות הוא הקושי לאתר קבוצות נבחנים שיהוו קבוצות ניתוח גדולות מספיק. בנוסף, קיים קושי להגדיר את קבוצת ההשוואה המתאימה ביותר, שכן בעלי לקויות הלמידה אינם מהווים מדגם מייצג מקרב כלל הנבחנים. כדי להקטין את הסיכון לתוצאה מוטה שנובעת מתופעה זו, בחרנו להשתמש בציוני נטייה (propensity score) שבהם משתני הרקע מגדר וגיל נלקחים בחשבון בתהליך הדגימה של קבוצת ההשוואה (קבוצת נבחנים "רגילים" שלא ביקשו התאמות, N=120,503), כך שתהיה דומה ככל האפשר לקבוצות מבקשי התנאים.

לא נמצאה כל הטיה בברירה ביחס לקבוצות בעלי לקות הלמידה, בעלי בעיות הקשב והריכוז ואלה שבקשתם נדחתה מסיבות מקצועיות, בעוד שנמצאה הטיה קלה לטובת קבוצת בעלי הבעיות הפיזיות ולרעת קבוצת חסרי הנתונים.

השוואת מקדמי התוקף בין הקבוצות מגלה שעבור מכפ"ל מקדמים אלה היו דומים מאוד בקרב ה"רגילים" ובקרב קבוצות בעלי המוגבלויות, בעוד שעבור הבגרות ועבור ציון הסכם התוקף עבור קבוצת ה"רגילים" היה גבוה במקצת (תוצאה שבולטת במיוחד עבור קבוצות בעלי לקות הלמידה ובעלי בעיות הקשב והריכוז).

## תועלת קורסי ההכנה לבחינה הפסיכומטרית, בהשוואה להכנה עצמית, בקרב נבחני עברית וערבית

איילת גולדצויג, כרמל און, יונתן סער

מרכז ארצי לבחינות ולהערכה

בשנים האחרונות נראה כי מרבית הנבחנים במכפ"ל מתכוננים לבחינה באמצעות קורס הכנה מסחרי. נשאלת שאלת מידת התועלת המופקת מקורס הכנה כזה לעומת הכנה עצמית בשיטות אחרות. שאלה נוספת שעולה היא האם קיימים הבדלים בדפוסי ההכנה לבחינה הפסיכומטרית בין נבחני עברית לנבחני ערבית, והאם יש הבדל במידת התועלת שכל קבוצה מפיקה מקורסי ההכנה לבחינה.

כדי להסיק מסקנות סיבתיות, יש לנטרל משתנים שעשויים להיות קשורים להבדלים בין הקבוצות המשוות, למשל יכולת בסיסית שונה של הנבחנים. בהיעדר הקצאה רנדומלית של הנבחנים לשיטות הכנה, ניסינו לנכות את השפעתם של משתני רקע כאלה באמצעים סטטיסטיים. משתני הרקע היו מין הנבחן, המצב הכלכלי, השכלת ההורים, וממוצע הבגרות המשוקלל כמדד ליכולת בסיסית. איתור הנבחנים הסתמך על שאלון המשוב הניתן בתום הבחינה, והעיבודים בוצעו בכל שפה בנפרד.

בשלב הראשון, נבנה מודל רגרסיה לינארית כדי לאמוד את תרומתו הייחודית של קורס הכנה – ביחס להכנה עצמית, מעבר למשתני הרקע – לציון במכפ"ל. זאת בקרב 52,327 נבחנים לראשונה. נמצא כי בקרב נבחני עברית, משתתפי הקורס קיבלו בממוצע 4.7 נק' פחות ממי שהתכוננו בעצמם, ואילו בקרב נבחני ערבית, משתתפי הקורס קיבלו בממוצע 9.6 נק' יותר ממי שהתכוננו בעצמם. בשני המקרים, מתאמים חלקיים וסמי-חלקיים בין סוג הכנה לציון בבחינה נמצאו זניחים. משתני הרקע הסבירו רק 1.5% מהשונות בסוג ההכנה, ולכן ניכרים לא הוביל לשינוי משמעותי בדפוס התוצאות.

לנוכח ההצלחה המוגבלת של שיטה זו בנטרול הברירה העצמית של הנבדקים לקבוצות הכנה שונות, ומתוך הנחה שליכולת הבסיסית של הנבחנים תפקיד מרכזי ברירה זו, בחלקו השני של המחקר בחרנו בפרדיגמה אחרת, המאפשרת אומדן טוב יותר של יכולתם. לצורך זה אותרו 15,828 נבחנים חוזרים, אשר נבחנו פעמיים או יותר בתקופת המחקר. במחקר נכללו ההיבחנות הראשונה והשנייה בלבד, והנבחנים חולקו לארבע קבוצות בהתאם לשיטת ההכנה שבה נקטו לפני כל בחינה. תחילה, נבנה מודל רגרסיה דו-שלבי לניכוי השפעתם של משתני הרקע על השיפור הממוצע בין שתי הבחינות. לפי המודל, חושו לכל קבוצה ממוצעים מתוקנים של הציונים בשתי הבחינות והשיפור ביניהן. לאחר מכן בוצעה סדרת השוואות זוגיות בין הקבוצות, באופן שאפשר, תחת הנחות מסוימות, לנכות את היכולת הבסיסית כדי לאמוד את תרומתו הייחודית של קורס הכנה. נמצא כי בעברית, הקורס תרם מעבר להכנה עצמית, בממוצע, כ-5 נק' לפני הבחינה הראשונה וכ-26 נק' לפני הבחינה השנייה. לעומת זאת, בערבית לא נצפה הבדל גדול בין תרומת הקורס הייחודית לפני כל בחינה, כ-12 נק' בממוצע בשני המקרים.

נראה כי שיטה זו שיפרה את טיב האומדן. על פי הממוצעים הנצפים וממצאי חלק א בעברית, משתתפי הקורס קיבלו ציון נמוך יותר בבחינה ממי שלמדו בעצמם, ככל הנראה עקב הברירה העצמית לקבוצות. השיטה המוצגת אפשרה להתגבר על כך ולדייק יותר באומדן תועלת הקורס, לפיכך נתקבל בה אומדן לתרומתו החיובית. מעניין כי נתגלה הבדל בין נבחני עברית לערבית בהקשר זה.

על אף מגבלות מתודולוגיות המחייבות זהירות במתן פרשנות סיבתית לממצאים, ממצאי המחקר הנוכחי מחזקים ומרחיבים את הידע בתחום.

## ידיעון האגודה הישראלית לפסיכומטריקה

ידיעון האגודה הישראלית לפסיכומטריקה משמש במה תקופתית להעברת מידע ממוסדות האגודה לחבריה ובין החברים לבין עצמם. מופיעים בו נושאים מתחום המדידה וההערכה בישראל ובעולם. נוסף על העורכת, הידיעון מונחה על ידי ועדה מייעצת מתוך חברי האגודה.

הידיעון מתוכנן להופיע פעמיים בשנה.

### ידיעון האגודה מוקדש לדיווחים בנושאים הבאים:

1. מהנעשה באגודה: חדשות האגודה, דיווח ועדות האגודה, כינוס שנתי
2. קורסים אקדמיים, השתלמויות וכנסים בנושאי מדידה והערכה
3. מחקרים בישראל
4. תוכנות, ספרים, מאמרים ואתרי אינטרנט
5. חדשות ועדכונים מהנעשה בתחום המדידה וההערכה בישראל ובעולם
6. תקצירי מאמרים שפורסמו בעיתונות המקצועית לאחרונה
7. דעות ותגובות

**המעוניינים לשלוח דיווחים ומידע לידיעון יפנו אל העורכת:**

**ד"ר עינת נוטע-קורן, ראמ"ה**

**דוא"ל [enotea.rama@education.gov.il](mailto:enotea.rama@education.gov.il)**

### **חברי הוועדה המייעצת:**

פרופ' אייל גמליאל, המרכז האקדמי רופין

פרופ' שאול פוקס, אוניברסיטת בר-אילן

ד"ר סול פיין, מידות

ד"ר יואל רפ, ראמ"ה

גב' אסטלה מלמד, מט"ח



כשיחלו

---

כשיחלו

---

## פוטו-מטריה - הערכה ומדידה בדימויים ומטאפורות מצולמים מודי אפשטיין, ליאת בסיס

מודי אפשטיין הוא פסיכולוג העובד במחלקת מדעי ההתנהגות במשטרת ישראל. הוא טייל רבות בארץ ובעולם והנציח במצלמתו מראות רבים.

ליאת בסיס מנהלת את 'בסיס להערכה' ומרצה באוניברסיטת בר-אילן. היא אוהבת לחבר מושגים מעולם המדידה וההערכה לעולמות נוספים.

ליאת ומודי השתעשעו ברעיון לתת פרשנות פסיכומטרית לצילומים שצולמו בהקשרים שונים. כך נולדה "פוטו-מטריה". מכרים של ליאת ושל מודי, רובם פסיכומטריקאים, סייעו להם בבחירת הפרשנויות המוצלחות מבין עשרות רבות.

הצילומים יוצגו בכנס בשלב ההתכנסות ובהפסקות.



מיון והשמה



מכנה משותף



עקומת U הפוכה

תאובות ופצלות  
בנושא פאודר פילראית אפסיכומטריקה  
אפסר אפסנות א:

שרה שפירא, מרכז ארצי לבחינות ולהערכה [sara@nite.org.il](mailto:sara@nite.org.il)

אתר האגודה הישראלית לפסיכומטריקה [www.ispa.org.il](http://www.ispa.org.il)

---