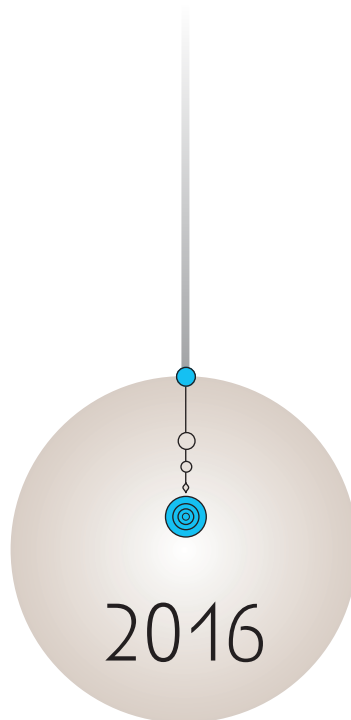


ט"ו בשבט תשע"ו 25.1.2016  כפר המכביה, רמת גן

הכינוס השניים-עשר של האגודה הישראלית לפסיכומטריקה



בתמיכת:



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

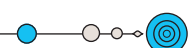
האגודה הישראלית לפסיכומטריקה (אפי) היא גוף המאגד בתוכו אנשי מקצוע מתחום המדידה וההערכה החינוכית והפסיכולוגית.

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
הרשמה, התכנסות, כיבוד קל	
	10:40 – 9:00
מושב א'	
דברי פתיחה וברכות יו"ר אפי: יואל רפ; יו"ר הכינוס: נעמי גפני	9:15 – 9:00
הענקת פרס אפי והצגת העבודה:	9:40 – 9:15
<ul style="list-style-type: none"> ■ הבדלים מגדריים במאפייני תפיסת העצמי של תלמידים מחוננים לירז טל	
סימפוזיון: אתגרים מתודולוגיים במחקרים רחבי היקף	10:40 – 9:40
יו"ר: חגית גליקמן	
<ul style="list-style-type: none"> ■ תכנון המדגם ואיסוף הנתונים בסקר מיומנות (PIAAC) צחי מקובקי, צביקה אמיר, אורי הדר	
<ul style="list-style-type: none"> ■ ניתוח נתוני המשוב הארצי בגיאוגרפיה על בסיס תורת התגובה לפריט צור קרליץ, שאול שלפק, הלנה קימרון, עינת נוטע-קורן, יואל רפ	
<ul style="list-style-type: none"> ■ בין מבחן המיצ"ב למבחן TIMSS – תיקוף תוצאות מחקר TIMSS מיכל בלר, נורית ליפשוטט, נתנאל גולדשמידט, ענבל רון קפלן, הדס גלברט	
	11:00 – 10:40
הפסקה	
	12:00 – 11:00
מושב ב' 1 (מושב מקביל) מחקרי תוקף של מבחני מיון להשכלה גבוהה בישראל	
יו"ר: ענת בן-סימון	
<ul style="list-style-type: none"> ■ האם ממוצעי הציונים בתואר ראשון בפסיכולוגיה מכוילים בין המוסדות להשכלה גבוהה? אייל גמליאל, גיל גולדצויג, חנן גולדשמידט	
<ul style="list-style-type: none"> ■ תיקוף ציוני הקבלה לתואר שני בפסיכולוגיה חנן גולדשמידט, גיל גולדצויג, אייל גמליאל	
<ul style="list-style-type: none"> ■ תוקף בחינת יע"ל (ידע בעברית לנבחני שפות זרות) איילת גולדצויג, יונתן סער	
	12:20 – 12:00
מושב ב' 2 (מושב מקביל) סוגיות במבחני בקיאות בשפה האנגלית והשפעה של ניסוח ההוראות על הביצוע במבחן	
יו"ר: עינת נוטע קורן	
<ul style="list-style-type: none"> ■ האם היבחנות חוזרת באותו נוסח במבחן אדפטיבי באנגלית נותנת יתרון לנבחן? טוני גוטנטיג, אבי אללוף, מיכל באומר, נדב פודולר, מרינה פרונטון	
<ul style="list-style-type: none"> ■ בדיקת הקשר בין סולם CEFR לבין סולם אמי"ר רות פורטס, נעמי גפני	
<ul style="list-style-type: none"> ■ השפעת הניסוח של הוראות במטלות למידה ומבחנים על ביצועי תלמידים איתמר שץ	
	12:20 – 12:00
הפסקה	
	13:20 – 12:20
מושב ג' 1 (מושב מקביל) מידה במציאות רב תרבותית	
יו"ר: יואל רפ	
<ul style="list-style-type: none"> ■ תכנית אבחון מותאם תרבות נועם פישמן, טלי סמני, ליאת בסיס	
<ul style="list-style-type: none"> ■ העדפות שפה בקרב נבחני ערבית בבחינה הפסיכומטרית כרמל אורן, איילת גולדצויג	
<ul style="list-style-type: none"> ■ קבלה להשכלה גבוהה וצדק חברתי – דוברי ערבית ועולים חדשים, והבחינה הפסיכומטרית נעמי גפני, אבי אללוף, ראובן סטולר	



מושב ג' 2 (מושב מקביל) **היבטים פסיכומטריים במיון והכשרה של עובדי הוראה**
יו"ר: מיכל בלר

■ **היבטים פסיכומטריים בהכשרה להוראה וקריירה בהוראה**
 דוד מעגן

■ **פיתוח והשמה של מערכת מיון מבוססת מרכז הערכה עבור מכון "אבני ראשה" למיון המועמדים לתוכנית הכשרת פרחי ניהול בתי ספר**
 דוד ציגלר, אביטל מושינסקי, ליסה לוי, איתי סודרי, הלנה קימרון, חני שלטון

■ **פיתוח כלי למדידת מעמד המורה בחברה הישראלית**
 זהבית כהן, יצחק גילת

ארוחת צהריים

13:20 – 14:45

מושב ד' 1 (מושב מקביל) **מערכי מטלות פתוחות והתנהגויות של עובדים בעלי**
יו"ר: אסטלה מלמד **כישורי יתר**

14:45 – 15:45

■ **פיתוח מדד יחיד לאיכות מעריכים של מבחני ביצוע**
 אבי אללוף, דביר קלפר, האשם ניג'ם, ג'ני שמולביץ

■ **שיפור המהימנות של מערכי חיבורים: בחינת כלי מהספרות על חוכמת ההמונים**
 מאיר ברנרון, אילן יניב, אבי אללוף

■ **התנהגויות לא נאותות בקרב עובדים בעלי כישורי יתר**
 סול פיין, מיכל אדווארד

מושב ד' 2 (מושב מקביל) **הערכת מוכנות ללימודים ונשירה מבית הספר**
יו"ר: רות פורטס

■ **הערכת מוכנות ללימודים אקדמיים**
 מיכל פיליפס, איתמר גתי, נועה סקה

■ **הנשירה הסמויה בישראל: ניתוק בקרב תלמידי בתי הספר**
 רות ברוך-קוברסקי, דליה בן רבי, וצ'יסלב קונסטנטינוב

■ **קבלת החלטות עבר/נכשל, שאפשר להגן עליהן בעזרת השימוש ביכולת הנבחנים ובקושי הפריטים**
 בועז שולרוף, פיל ג'ונס, רולף טרנר

הפסקה

15:45 – 16:00

מושב ה' **כריית נתונים, תיקון לקיצוץ תחום וקיום משותף של יהודים וערבים**
יו"ר: גיל גולדצויג

16:00 – 17:00

■ **מודלים של כריית נתונים לחקירת סוללת המבחנים הממוחשבים של מת"ל**
 דביר קלפר, ענת בן-סימון, אפי לוי, נועה סקה

■ **השוואה בין שתי שיטות לחישוב טעויות תקן ורווחי סמך למתאמים שתוקנו לקיצוץ תחום עקיף**
 תמר קנת-כהן, דביר קלפר, אליוט טורוול

■ **האם אנחנו מוכנים לחיות ביחד? המשגה ומדידה של קיום משותף בין יהודים וערבים בישראל**
 מאור שני, קלאוס בואנקה, מרגריט שרייר

הפסקה

17:00 – 17:10

אסיפת חברי אפי

17:10 – 17:40

הבדלים מגדריים במאפייני תפיסת העצמי של תלמידים מחוננים

לירז טל

האוניברסיטה העברית בירושלים

בישראל אחוז הבנות המאותרות ומשתתפות בתכניות לטיפוח מחוננים נמוך משמעותית מאחוז הבנים. שאלת הפער המגדרי באוכלוסיית המחוננים בישראל מתמיהה לנוכח ההישגים הלימודיים בפועל של בנים ובנות מחוננים. סקירת הספרות בנושא מציעה כי חלק מהגורמים נעוצים במשתנים פסיכולוגיים- אישיותיים של העצמי. המחקר הנוכחי בודק האם ובאיזה אופן מאפיינים משתנים פסיכולוגיים של תפיסת העצמי את אוכלוסיית המחוננים וכן את תרומתם של משתנים אלו להסברת הפער המגדרי בקרב אוכלוסיית המחוננים בישראל. נוסף על כך, המחקר מאפשר הבנה מקיפה על השלכות מדיניות ההעדפה המתקנת לבנות שאומצה במטרה להגיע לשוויון מגדרי ועל תחושתיה של אוכלוסיית הבנות המחוננות שנכנסו לתכנית בעקבותיה.

"שאלון מאפייני תפיסת העצמי בקרב ילדים מחוננים" נבנה, הוערך ותוקף בשני מחקרי חלוץ על מנת לבחון את שאלות המחקר. שאלון זה מורכב מארבעה כלי מחקר שונים הבוחנים היבטים פסיכולוגיים שונים של תפיסת העצמי. מדגם נרחב בפריסה ארצית כלל 524 תלמידים מחוננים הלומדים במסגרות החינוכיות לגביהם מיושמת מדיניות ההעדפה המתקנת. מסגרות הלימוד בהן נערך המחקר כללו שלושה מרכזי העשרה (יום למידה בשבוע) וחמישה בתי ספר ובהם כיתות מחוננים (שבוע למידה מלא) הפזורים ברחבי הארץ.

הבדלי מגדר: בהתאם להשערות נמצא כי לבנים תחושת מסוגלות גבוהה יותר בתחומים של מסוגלות אקדמית ומתמטיקה ואילו לבנות תחושת מסוגלות עצמית גבוהה יותר בתחום המילולי ובמדד יוזמה ומנהיגות. כמו כן, לבנות רמת חרדת בחינות גבוהה יותר.

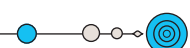
גיל: קיימת מגמה התפתחותית של העמקת הפערים המגדריים בתפיסת מאפייני העצמי.

מסגרת: בהתאם לספרות המחקרית, מסתמן יתרון פסיכולוגי רגשי מובהק למסגרת הטיפוח של מרכז העשרה על פני כיתת מחוננים.

העדפה: בהתאם להשערות, נמצאו הבדלים מובהקים במדדים של דימוי עצמי חברתי ותחושת מסוגלות עצמית במתמטיקה. בשאר המדדים לא נמצאו הבדלים מובהקים.

לסכום, למחקר הנוכחי תרומה ייחודית בבניה ופיתוח כלי מחקרי, השאלון "מאפייני תפיסת העצמי בקרב ילדים מחוננים". שאלון זה משמש לחקר עולמם הרגשי והסובייקטיבי של אוכלוסיית המחוננים.

כמו כן, המחקר הנוכחי תורם ממצאים ותובנות חדשות לתחום טיפוח תלמידים ותלמידות מחוננים בכלל ובישראל בפרט ומאפשר להסיק מסקנות לגבי הגורמים המשפיעים על מימוש הפוטנציאל האקדמי של בנים ובנות מחוננים, בגילאים שונים ובמסגרות לימוד שונות. הבנה זו תאפשר בסופו של דבר קידום שוויון הזדמנויות בחינוך מחוננים וצמצום הפערים המגדריים. כמו כן, המחקר הנוכחי מאפשר הערכה של החלק הרגשי הכרוך ביישום מדיניות של העדפה מתקנת.



סימפוזיון

אתגרים מתודולוגיים במחקרים רחבי היקף

יו"ר: חגית גליקמן
הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

מציגים

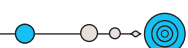
צחי מקובקי, צור קרליץ, מיכל בלר

הקדמה

מבחני הישגים סטנדרטיים ורחבי היקף – large scale assessment – (מבחני בגרות, מבחני מיצ"ב, משוב ארצי, מבחנים בינלאומיים) מאפיינים על פי רוב מערכות חינוך ומטרתם למדוד ולהעריך את ההישגים הלימודיים של תלמידי בתי הספר ושל כלל המערכת או חלקים נרחבים בה. מבחנים כאלו משמשים לעתים גם מערכות מיון למוסדות להשכלה גבוהה, לצבא ולמקומות עבודה. בשנים האחרונות השימוש במבחנים כאלו התפשט למחוזות חדשים כאשר גוף בינלאומי מרכזי, ה-OECD, החל להשתמש בהם כדי להעריך את רמת המיומנויות והידע של מבוגרים במדינות שונות מתוך מטרה להשוות את יכולותיהם של אזרחים בוגרים במדינות מפותחות בהקשר של הנדרש מהם בעולם העבודה (ע"ע מחקר פיאאק PIAAC). בכל המקרים האלו, מומחים למדידה נדרשים לעצב מערכות מבחנים בעלות סממנים מחקריים, לפתח כלי מדידה תקפים, מהימנים, הוגנים, ולהתמודד עם אתגרים מתודולוגיים ויישומיים בלתי מבוטלים, מתוך חתירה לספק בסופו של דבר מערך נתונים מקיף, רחב, מייצג, רלבנטי, אמין ומהימן.

בסימפוזיון המוצע, יוצגו אתגרים ודילמות מחקריות-פסיכומטריות-סטטיסטיות המאפיינים מערכות מבחנים רחבי היקף, אשר עמן מתמודדים מומחי המדידה והסטטיסטיקאים. המתודות והכלים הסטטיסטיים והפסיכומטריים עשויים להיות שונים ממחקר למחקר אך הם מדגימים קשיים במערכים, או בסביבת ההיבחנות, בתגובת האוכלוסייה הנבחנת, אשר מאיימים כולם, אם לא יטופלו בזהירות הראויה, על תוקף התוצאות של המחקרים ומערכות המבחנים שבהם נעשה שימוש לטובת אותם מחקרים. על מנת להבטיח כי לא תיווצר תמונה לא רלוונטית, מוטה, מטעה ו/או בלתי מדויקת של התוצאות ובסופו של דבר להביא לעיוות בפרשנות של הנתונים ובתוקף המחקרי של כלי המחקר ושל המסקנות העולות ממנו, איש המדידה נדרש להיות ער לכשלים אפשריים, למערכים מוגבלים, לבחירה הולמת של כלי העיבוד והמודלים התאורטיים המתאימים לכל מצב ומצב. בין הסכנות האלו אפשר למנות: ייחודיותה של האוכלוסייה הישראלית, על רבדיה, ומה נגזר מכך לבניית המדגמים, מבנים שונים ומגוונים של מבחנים, על נוסחיהם והשילוב בין פריטים מסוגים שונים במבחנים אלו, איומים של תוקף השלכתי וסביבתי בעיקר מצד גורמים בעלי עניין (פגיעה בטוהר בחינות מצד התלמידים או מוריהם, השבה לא רצינית על מבחנים ושאלונים, הכנה לימודית בלתי מייצגת של התלמידים הנבחנים).

בסימפוזיון זה יוצגו שלוש עבודות המתארות קשיים ואתגרים במבחנים רחבי היקף בשלבים שונים של המדידה. העבודה הראשונה תציג קשיים ואתגרים האופייניים לשלב שמתחיל טרם העברת המבחנים עצמם (בניית המדגם) ובשלב העברתם (שלב איסוף הנתונים) ואת דרכי ההתמודדות עמם. העבודה השנייה תתמקד בקשיים ואתגרים האופייניים לשלב עיבוד הנתונים – בחירת המודלים והכלים הפסיכומטריים-סטטיסטיים המתאימים ביותר לניתוח, שיקולים בפסילת תגובות, הכנת תשתית נתונים שתשרת באופן הטוב ביותר את האחראים על דיווח התוצאות. העבודה השלישית תציג אתגרים וניתוחים שלאחר העברת המבחנים ועיבוד הנתונים – כיצד לבדוק תהיות וחששות כי משהו השתבש בהליך ההיבחנות ומאיים על תוקף התוצאות. העבודה הראשונה מבוססת על נתוני מחקר פיאאק – PIAAC שהועבר בישראל במהלך 2014, מחקר בינלאומי של ה-OECD שמשווה מיומנויות אוריינות של מבוגרים בשלושה תחומי דעת: קריאה, מתמטיקה ופתרון בעיות בסביבת העבודה. המחקר הועבר בישראל בהובלת הלשכה המרכזית לסטטיסטיקה ובשיתוף ראמ"ה. העבודה השנייה מבוססת על נתוני המשוב הארצי בגאוגרפיה, מחקר רחב היקף אשר ראמ"ה מובילה בסיוע המרכז הארצי לבחינות ולהערכה – הגוף העוסק בעיבוד הנתונים של מחקר זה. העבודה השלישית מבוססת על נתוני טימס 2011, מחקר בינלאומי ביוזמה ובהובלת ה-IEA, אשר בוחן את הידע והמיומנויות של תלמידים בכיתה ח' במדעים ובמתמטיקה ואשר ראמ"ה אחראית על ההשתתפות של ישראל בו. מכלול העבודות מלמד כי עבודת המדידה וההערכה דורשת ערנות מתמדת, ביקורתיות וחשדנות לאורך כל הדרך ולאורך כל הליך המדידה, החל מפיתוח כלי המבחן, תכנון המדגם, עיבוד הנתונים וניתוחם. חשדנות וביקורתיות אשר מתורגמות לעיבודים ובדיקות נוספות של הנתונים, יכולות בכל שלב ושלב לצמצם, אם לא למנוע הטיות או לאפשר גילוי בדיעבד של הטיות בתוצאות באופן שיחזק ככל האפשר את תוקפם של מבחנים רחבי היקף.



תכנון המדגם ואיסוף הנתונים בסקר מיומנויות (PIAAC)

צחי מקובקי, צביקה אמיר, אורי הדר
הלשכה המרכזית לסטטיסטיקה

סקר מיומנויות (PIAAC) (The Program for the International Assessment of Adult Competencies) הוא הסקר הבינלאומי המקיף ביותר, שנערך אי-פעם בנושא כישורי בוגרים. הסקר מעריך את כישורי בני 16-65. הוא מתמקד בכישורי מפתח, הנדרשים לאנשים בעידן המידע, כדי להשתתף באופן מוצלח בחיי הכלכלה והחברה של המאה ה-21. ארגון המדינות המפותחות (OECD) הוא היוזם והמארגן של הסקר ובו משתתפים למעלה מ-30 מדינות מפותחות. את הסקר מלווה מנהלת בינלאומית אשר קובעת סטנדרטים אחידים לתכנון וביצוע הסקר, עורכת בקורות על כל שלבי הסקר, כדי להבטיח את הצלחת הסקר ולאפשר השוואה תקפה של התוצאות בין המדינות. ב-2014 פורסמו תוצאות הגל הראשון של 24 המדינות המשתתפות וב-2016 יתווספו אליהן תוצאות המדינות המשתתפות בגל השני, ובכללן ישראל.

השתתפות ישראל בסקר היא בהובלת הלשכה המרכזית לסטטיסטיקה (למ"ס) בשיתוף פעולה עם הרשות הארצית למדידה והערכה בחינוך (ראמ"ה) אשר הייתה אחראית להליך תרגום והתאמת הכלים לעברית, ערבית ורוסית ולבדיקת המבחנים. סקר זה הינו מערכת מבחנים סטנדרטיים רחבי היקף (שלצדם שאלוני רקע) הגדולה ביותר שהתקיימה בארץ שלא במסגרת מערכת החינוך. דבר זה אשר לכשעצמו מציב אתגרים יוצאי דופן. הסקר מהווה חידוש מתודולוגי וטכנולוגי חשוב בהפעלת טכנולוגיה ושאלונים שפותחו בחו"ל לשילוב ראיון אישי ומבדקים למרואיינים בוגרים.

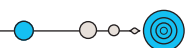
מחקרים מראים שככל שרמת המיומנות גבוהה יותר כך יכול האדם לשפר את רווחתו, תחושותיו ובריאותו. ארגון ה-OECD בחר במיומנויות היסוד ככלי חשוב בקביעת המדיניות וקידום החברה והכלכלה. מיומנויות היסוד שנמדדו בסקר בעזרת מבדקים היו:

- **אוריינות קריאה:** היכולת להבין מידע ממגוון סוגי טקסט ולהשתמש בו. השאלות והטקסטים נוגעים לצורות הצגה מוכרות כגון: מודעות, הוראות שימוש וידיעות בעיתון.
- **אוריינות מתמטית:** היכולת להשתמש, ליישם, לפרש ולהעביר הלאה מידע ומושגים מתמטיים. הבדיקה מתייחסת לביצוע חישובים בסיסיים המשמשים בחיי היום יום.
- **פתרון בעיות בסביבה טכנולוגית:** היכולת למצוא, להעריך מידע ממקורות שונים ולהשתמש בו לשם פתרון בעיות, כגון חיפוש ספר בספרייה או ניווט בתוך אתר אינטרנט וכל זאת באמצעות המחשב.

אוכלוסיית הסקר כוללת את בני 16-65 באוכלוסייה הקבועה במדינת ישראל. כדי להוציא מדגם שיכסה בצורה מקיפה אוכלוסייה זו ויאפשר מדגם יעיל, נבחר קובץ מרשם התושבים כמסגרת הדגימה. הסקר בכל המדינות המשתתפות תוכנן לקבל כ-5,000 נדגמים משיבים. ועדת ההיגוי של הסקר, הכוללת נציגים ממשרדי ממשלה וחוקרים, החליטה לתגבר את המדגם באוכלוסייה הערבית ובאוכלוסייה החרדית.

העבודה הנוכחית עוסקת באתגרים שעמדו בפני הלמ"ס בשלב קביעת מסגרת הדגימה ותכנון המדגם ובשלב איסוף הנתונים.

- האתגרים שעמדו בפני עורכי הסקר שיתוארו ויידונו בעבודה בהרחבה הם:
- כיסוי חסר וכיסוי עודף הנובעים ממסגרת דגימה מבוססת מרשם תושבים
 - שילוב מידע ממקורות מנהליים נוספים להעשרת התכונות הידועות במסגרת הדגימה כמו למשל זיהוי האוכלוסייה החרדית במרשם התושבים ("חרדים מנהליים")
 - תכנון המדגם - הגדרת 84 שכבות דגימה, הקצאת המדגם בין השכבות, תגבור תת-אוכלוסיות, שיטת הדגימה
 - אתגרים מיוחדים בתכנון שלב איסוף הנתונים - סקר ללא חובת השבה (להבדיל מסקרי ליבה שעורך הלמ"ס) נושא מיוחד, אורך השאלון ושיטת איסוף לא שגרתיות (פתירת תרגילים בנוסף לשאלון רקע)
 - קשיים מיוחדים שהתגלו בשלב איסוף הנתונים - מבצע צוק איתן, קשיים בגיוס סוקרים, נושא חוק גיוס חרדים שעמד על הפרק לא עזר לקבלת שיתוף הפעולה הדרוש
 - שיעורי ההשבה שהתקבלו בסקר לפי משתני רקע דמוגרפיים-חברתיים
 - הטיפול שנערך לאחר האיסוף כדי להתמודד עם ההטייה הנובעת משיעורי ההשבה



ניתוח נתוני המשוב הארצי בגיאוגרפיה על בסיס תורת התגובה לפריט

צור קרליץ, שאול שלפק
מרכז ארצי לבחינות ולהערכה (מאל"ו)

הלנה קימרון, עינת נוטע-קורן, יואל רפ
הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

בשנת הלימודים תשע"ד נערך לראשונה המשוב הארצי המדגמי בגיאוגרפיה ("המשוב"), במטרה לספק מידע למשרד החינוך ולציבור על הישגים, הערכים, וההקשר הפדגוגי בלימודי הגיאוגרפיה של תלמידי כיתות ט'. המשוב כלל מבחני הישגים ושאלונים למדגם מייצג ארצי של 3,764 תלמידים, וכמו כן סקרים למנהלים ולמורים שמלמדים גיאוגרפיה.

מערך המחקר היה מטריציוני ואפשר דגימת תוכן רחבה: כל תלמיד קיבל אחד מחמשת נוסחי המבחן כאשר בסה"כ, המבחן כאשר כל נוסח כלל בממוצע כ-32 פריטים ובסה"כ, המבחן כלל 128 פריטים שונים. בין ארבעה מהנוסחים לא הייתה חפיפה, ואילו הנוסח החמישי חפף באופן חלקי את כל אחד מהנוסחים האחרים, ופריטי העוגן הללו שימשו לקישור ולכיול בין כל הנוסחים. מבנה המבחן ומערך ההיבחנות הציבו אתגרים לא מבוטלים בשלב עיבוד הנתונים שהעיקריים בהם היו אלו:

מספר קטן של משיבים לפריט (כל נבחן נחשף לכרבע מהפריטים וכל פריט הועבר לכרבע מהנבחנים). בנוסף, היו תלמידים שדילגו או לא השיבו על פריטים.

- קיימים הבדלי קושי בין הנוסחים שנבעו מהרכב הנושאים בהם.
- המבחן כלל פריטים דיכוטומיים ופוליטומיים (בעלי עד 5 רמות ציון).

בהרצאה נתאר את האתגרים והדילמות המחקריות איתן התמודדנו בשלב טיוב הנתונים וניתוחם. השאיפה הייתה למצות את עושר המידע מן הנתונים ולבצע עיבודים העומדים בסטנדרטים גבוהים של איכות פסיכומטרית. הפרמטרים שהופקו מניתוח זה שימשו כבסיס להליך קביעת התקנים להישגים במבחן. ניתוח הנתונים והדיווח עליהם התבססו על תורת התגובה לפריט (תת"פ), ובפרט על המודל לציון חלקי (partial credit model; Masters, 1982). שבו משתמשים לניתוח מבחנים בינלאומיים כמו פיזה וטימס. מודל זה הוא חד-פרמטרי, כלומר שהוא כולל רק פרמטרים של קושי עבור כל פריט. עבור פריט דיכוטומי, המודל מייצג את ההסתברות שנבחן בעל יכולת θ יקבל ציון 1 בפריט. פרמטר הקושי של הפריט הוא ערך ה- θ שבו ההסתברות להשיב נכון היא 0.5. עבור פריטים פוליטומיים המודל מייצג את ההסתברות שנבחן בעל יכולת θ יקבל ציון מסוים בפריט, לעומת ציונים אחרים. פרמטרים אלו נקראים סיפי הקושי, כך שהסף הראשון הוא הנקודה בה יש הסתברות של 0.5 לקבל ציון גדול מ-0, הסף השני מתייחס להסתברות של 0.5 לקבלת ציון גדול מ-1, וכן הלאה. השימוש בסיפי קושי בתוך פריט מאפשר גם לזהות פריטים בעייתיים מבחינת יכולת האבחנה שלהם.

לאחר בדיקת הנחות היסוד של המודל (חד מימדיות ואי תלות מקומית) הליך כיול הפרמטרים בוצע תחילה בכל נוסח בנפרד, ולאחר מכן, בהליך משותף שכלל את התגובות לכל הנוסחים. קובץ הנתונים המשותף הינו אופייני למערך דגימה מטריציוני: ישנם חללים מבניים (עבור כל פריט ישנם נתונים רק עבור הנבחנים שקיבלו את הנוסח הכולל את הפריט), וכל הפריטים החופפים המופיעים בנוסחים שונים מסודרים באופן תואם (למשל, לנבחן מנוסח 1 ולנבחן מנוסח 5 יהיו תשובות תחת אותו פריט עוגן). הפרמטרים של פריטי העוגן שהתקבלו בשני הליכים אלו הוששו זה לזה והוששו גם לתוצאות על בסיס תורת המבחנים הקלאסית. בעבודה יוסברו שלבי הניתוח השונים והאתגרים הייחודיים של כל שלב ושלב.

בין מבחן המיצ"ב למבחן TIMSS – תיקוף תוצאות מחקר TIMSS

מיכל בלר

מכללת לוינסקי לחינוך

נורית ליפשוט, נתנאל גולדשמידט, ענבל רון קפלן, הדס גלברט
הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

בשנת תשע"א נערכו בישראל שני מבחנים רחבי היקף בכיתות ח' בתחומי הדעת מתמטיקה ומדעים – מחקר TIMSS הבינלאומי ומבחן המיצ"ב הארציים הנערכים מדי שנה על ידי ראמ"ה.

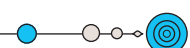
מחקר TIMSS הוא מחקר בינלאומי בתחום החינוך שנועד להשוות בין הישגיהם של תלמידי בתי ספר במתמטיקה ובמדעים במדינות שונות לאורך שנים. מאז שנת 1999 נערך המחקר בישראל בקרב מדגם ארצי מייצג של תלמידי כיתות ח', דוברי עברית ודוברי ערבית. במחקר TIMSS 2011 השתתפו כ-4,700 תלמידי כיתות ח' ב-150 בתי ספר.

הישגי התלמידים בישראל בשנת 2011 במחקר TIMSS הציבו את ישראל במקום גבוה, הן בהשוואה להישגי מדינות אחרות והן בהשוואה להישגי ישראל במחזורי קודמים. העלייה בהישגים בישראל בין מחזורי 2007 ל-2011 (53 נקודות במתמטיקה ו-48 נקודות במדעים) הייתה גדולה ואף חריגה בהשוואה למידת השיפור המקובלת בקרב מדינות שמשותפות במחזורי מחקר רצופים. העלייה הגדולה בהישגי התלמידים בישראל ב-2011 גרמה לכלי התקשורת ולחוקרי חינוך בישראל להיות ספקנים ביחס לתוקף הממצאים, ולראותם כמוטים כלפי מעלה. ההטיה יוחסה בין השאר להטיה לכאורה במדגם, לשינויים בתרגום ולתופעות נוספות. בין השאר, עלה החשש כי בניגוד להנחיות ראמ"ה הייתה הכנה אינטנסיבית ייעודית של הכיתות שעלו במדגם בתחומים הנבדקים במבחן – מתמטיקה ומדעים – וכי הוקצו משאבי הוראה ייחודיים לאותן כיתות במטרה לשפר את הישגיהם ובכך להעלות את התוצאה של ישראל (באופן שאינו מייצג את העשייה החינוכית בכלל המערכת). מעבר לשאלה הערכית שבגישה כזו, עצם קיומם של תהליכי הכנה חד פעמיים לקראת מבחנים מדגמים שאמורים לנטר את מערכת החינוך ולשרת את מקבלי החלטות, פוגעים במדידה – באמינותם של המבחנים, במהימנותם ובתוקפם.

בשנת הלימודים תשע"א, נערכו בישראל, כבכל שנה, מבחני מיצ"ב. מטרת מבחנים אלו היא לבדוק את השליטה של תלמידי בתי הספר ברמת הדרישות המצופה על פי תכנית הלימודים במקצועות מדע וטכנולוגיה, אנגלית, מתמטיקה ושפת-אם (עברית/ערבית). בשנה זו, השתתפו במיצ"ב בכל תחום דעת כרבע מבתי הספר בארץ, אשר מהווים מדגם מייצג של האוכלוסייה.

שני המחקרים, ה-TIMSS והמיצ"ב, מתבססים על מבחנים סטנדרטיים אחידים הנערכים בבתי ספר באותה שכבת גיל ובקרב אותה אוכלוסייה ברמה ארצית. שני המבחנים דומים בתכנים. יתרה מכך, בחינות המיצ"ב באותה שנה נערכו סמוך למועד שבו נערך מחקר TIMSS.

מטרתו של מחקר זה היא לבחון האם היה ממש בחששות כי נתוני TIMSS 2011 בישראל היו מוטים כלפי מעלה בשל הליך הכנה מאסיבי שהתמקד בכיתות ובבתי הספר שנדגמו למחקר. המחקר התבסס על השוואת תוצאות מחקר TIMSS 2011 לתוצאות המיצ"ב בכיתה ח' באותה שנת לימודים (תשע"א). המחקר ידגים בין השאר כיצד ניתן מתודולגית לתקף תוצאות של מחקר אחד באמצעות אחר.



ההשערה כי הייתה הכנה ממוקדת של כיתות שנדגמו למחקר TIMSS נבדקה באמצעות השוואת הביצועים במבחני המיצ"ב בין תלמידים שהשתתפו בשתי מערכות המבחנים לעמיתיהם שהשתתפו במיצ"ב אך לא השתתפו ב-TIMSS. במידה שהייתה הכנה ממוקדת בכיתות שנבחנו ב-TIMSS - צפוי כי תבוא לידי ביטוי בציוני בחינות המיצ"ב במתמטיקה ובמדעים באותן הכיתות ובתי ספר.

נערכה בדיקה לגבי נבחני מתמטיקה ומדעים בעברית ובערבית. תוצאות המחקר מצביעות על כך כי בשלושה מתוך ארבעת המקרים לא נמצאה עדות להכנה ממוקדת של כיתות המדגם. בקבוצה הרביעית הממצאים תומכים בהשערה כי הכנה ייעודית כזו התקיימה.

בשנת 2015 נערך מחזור נוסף של מחקר TIMSS ונערכו במקביל מבחני מיצ"ב במדגם מייצג של כשליש מבתי הספר בישראל. בדיקה דומה נעשתה וממצאים ראשוניים מבדיקה זו יוצגו.

האם ממוצעי הציונים בתואר ראשון בפסיכולוגיה מכילים בין המוסדות להשכלה גבוהה?

אייל גמליאל

המרכז האקדמי רופין

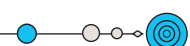
גיל גולדצויג

המכללה האקדמית תל-אביב יפו

חנן גולדשמידט

המרכז הבינתחומי הרצליה

תכניות הלימוד לתואר שני בפסיכולוגיה קלינית ממיינות מועמדים לפי מספר ציוני קבלה, וביניהם ממוצע לימודי תואר ראשון בפסיכולוגיה וציון מבחן המתא"ם. בעוד שציון מבחן המתא"ם הוא ציון מכיל עבור כל המועמדים, מידת הכיול של ממוצעי התואר הראשון במוסדות שונים להשכלה גבוהה אינה ברורה. חוסר כיול בין הציונים עלול לגרום לחוסר הוגנות של תהליכי הקבלה לתכניות הלימודים לתואר שני כלפי מועמדים שלמדו תואר ראשון במוסדות שונים: אלה שלמדו במוסדות "מקלים" בציונים יופלו לטובה, ואילו אלה שלמדו במוסדות "מחמירים" יופלו לרעה. מטרת מחקר זה הינה לבדוק את מידת הכיול של ממוצעי התואר הראשון של מועמדים ללימודי תואר שני בפסיכולוגיה המגיעים ממוסדות שונים להשכלה גבוהה. המחקר התבצע על שני בסיסי נתונים של מועמדים ללימודי תואר שני בפסיכולוגיה בשני מוסדות להשכלה גבוהה בארץ. כל אחד מבסיסי הנתונים כלל כאלף מועמדים שלמדו תואר ראשון ב-13 מוסדות להשכלה גבוהה, ובהם חמש אוניברסיטאות המחקר (תל-אביב, ירושלים, חיפה, בן-גוריון ובר-אילן), ושמונה מוסדות אחרים להשכלה גבוהה (מכללות ציבוריות ופרטיות והאוניברסיטה הפתוחה; להלן "מכללות"). מספר המועמדים שהגיעו מחמש האוניברסיטאות ומשלוש מכללות בשני בסיסי הנתונים היה גבוה מ-30, אולם מספר המועמדים שהגיעו מחמש מכללות נוספות היה קטן ולכן הם קובצו יחדיו (להלן: "המכללה המאוחדת"). בדיקת הכיול אמורה להיעשות אל מול קריטריון של הצלחה בלימודי תואר שני, אך בהיעדר נתונים כאלה בבסיסי הנתונים שעמדו לרשות המחקר נבדק הכיול של ממוצעי התואר הראשון אל מול ציון מבחן המתא"ם. הממצאים בשני בסיסי הנתונים היו דומים ומדווחים להלן יחד: ממוצע ציוני התואר הראשון היה 92 (ס"ת=2); ממוצע ציוני המתא"ם היה 113 (ס"ת=11); המתאם ביניהם היה 0.30. בכל בסיס נתונים חושבו תשע גרסיות מרובות לניבוי ממוצע התואר הראשון מידיעת ציון המתא"ם, בהן בשלב השני הוספו משתני המוסד (כמשתנה דמה). מקדם הרגרסיה של כל מוסד שיקף את מידת הכיול של ציוני המועמדים שלו: ערכים חיוביים מעידים על כך שהמועמדים שהגיעו מאותו מוסד קיבלו ממוצע תואר ראשון גבוה יותר מהמועמדים האחרים כאשר ציון המתא"ם מוחזק קבוע (כלומר מועמדים אלה "הרוויחו"), ולהיפך. בהחזקת ציון המתא"ם קבוע, ממוצע התואר הראשון של המועמדים שהגיעו משלוש אוניברסיטאות ומשתי מכללות היה נמוך יותר מזה של כל האחרים ב-0.2 עד 0.6 נקודות (כעשירית עד כרבע סטיית תקן של ממוצע ציוני הב.א.); ממוצע התואר הראשון של המועמדים שהגיעו משתי אוניברסיטאות, ממכללה אחת ומ"המכללה המאוחדת" היה גבוה מזה של האחרים ב-0.2 עד 1.3 נקודות (כעשירית עד כשש עשיריות סטיית תקן). ממצאי המחקר מסייגים את הסטריאוטיפ שאולי קיים בציבור, על פיו קיימת אינפלציה של ציונים במכללות לעומת אוניברסיטאות המחקר. ראשית, ממצאי ההטיה בקרב המוסדות היו מעורבים בקרב האוניברסיטאות ובקרב המכללות. שנית, כיול של ציוני הקבלה לתואר שני אמור להתבצע לגבי קריטריון של הצלחה בלימודי התואר השני, ולא לגבי ציון מבחן המתא"ם. נתוני מחקרים בעבר מנבאים שההטיה שתימצא במחקרי כיול שישמשו בקריטריון הצלחה של לימודי תואר שני יהיו קטנים יותר מאלה שנמצאו במחקר זה.



תיקוף ציוני הקבלה לתואר שני בפסיכולוגיה

חנן גולדשמידט
 המרכז הבינתחומי הרצליה
 גיל גולדצויג
 המכללה האקדמית תל-אביב יפו
 אייל גמליאל
 המרכז האקדמי רופין

תהליכי הקבלה לתואר ראשון ולתואר שני בפסיכולוגיה חולקים הגיון דומה של שילוב בין חזאי הבדוק הצלחה בלימודים קודמים (ממוצע בגרות וממוצע תואר ראשון, בהתאמה) ושל ציון בחינה סטנדרטית ואחידה (פסיכומטרי ומתא"ם, בהתאמה). מטרת מחקר זה היא לבדוק את התוקף של ממוצע התואר הראשון וציון מבחן המתא"ם בניבוי קריטריון ההצלחה בלימודי התואר השני בפסיכולוגיה עבור בוגרי תואר ראשון באוניברסיטאות המחקר לעומת מכללות. בסיס הנתונים כלל 322 מסיימי תואר שני במגמות שונות בפסיכולוגיה במוסד להשכלה גבוהה בישראל: 237 היו בוגרי תואר ראשון בפסיכולוגיה בחמש אוניברסיטאות המחקר (חיפה, תל-אביב, ירושלים, בר-אילן, ובן-גוריון), ו-85 היו בוגרי תואר ראשון במדעי ההתנהגות או בפסיכולוגיה בשמונה מוסדות אחרים (מכללות ציבוריות, מכללות פרטיות, או האוניברסיטה הפתוחה; להלן: "מכללות"). ממוצע התואר הראשון היה כ-92 (ס"ת=2), ממוצע ציוני מבחן המתא"ם היה כ-110 (ס"ת=11), וממוצע התואר השני היה כ-90.5 (ס"ת=2). ממוצעי הציונים בקרב בוגרי האוניברסיטאות היו גבוהים יותר מאלה של בוגרי המכללות בשלושת הציונים: הפער היה הגדול ביותר בציון מבחן המתא"ם (מעל חצי סטיית תקן), נמוך יותר בממוצע התואר הראשון (כשליש סטיית תקן), והנמוך ביותר בקריטריון ההצלחה בלימודי התואר השני (כרבע סטיית תקן). תוקף הניבוי של ממוצע התואר הראשון את ממוצע התואר השני היה דומה בקרב בוגרי שני סוגי המוסדות (0.26 עבור בוגרי האוניברסיטאות ו-0.32 עבור בוגרי המכללות), אך תוקף הניבוי של ציון מבחן המתא"ם היה שונה: 0.29 עבור בוגרי האוניברסיטאות ומתאם אפסי עבור בוגרי המכללות (המתאם בין שני החזאים היה גבוה מעט יותר עבור בוגרי האוניברסיטאות לעומת המכללות - 0.20 לעומת 0.15). את הפערים בממצאים לא ניתן להסביר במונחים של קיצוץ תחום, שכן סטיות התקן היו דומות בשלושת הציונים עבור בוגרי האוניברסיטאות והמכללות. ממצאי העבודה מורים שהפער בין בוגרי האוניברסיטאות לבוגרי המכללות בציוני מבחן המתא"ם הוא כפול לערך מהפער בין אותם סטודנטים בממוצעי ההצלחה בלימודי התואר השני. ממצא זה עקבי עם ממצאים קודמים ויש לו השלכה על הוגנות השימוש בציוני הקבלה ללימודי תואר שני ביחס לבוגרי שני מוסדות הלימוד: הפערים הגדולים יותר בציוני הקבלה לעומת ההצלחה בלימודים מצביעים על האפשרות שבוגרי המכללות מופלים לרעה בתהליכי הקבלה.

בעוד שתוקף הניבוי של ממוצע התואר הראשון את ההצלחה בלימודי התואר השני היה דומה בין בוגרי האוניברסיטאות לבוגרי המכללות (כ-0.30), נמצא שוני ביניהם בתוקף הניבוי של ציוני מבחן המתא"ם: כ-0.30 עבור בוגרי האוניברסיטאות לעומת תוקף אפסי בקרב בוגרי המכללות. הממצאים עבור בוגרי האוניברסיטאות תואמים את שנמצא לגבי תוקף הניבוי של ממוצע הבגרות וציון הבחינה הפסיכומטרית את ההצלחה בלימודי תואר ראשון במדעי החברה באוניברסיטאות (תוקף ניבוי של כ-0.30 ללא תיקון לקיצוץ תחום). לא ניתן לבצע השוואה דומה עבור בוגרי המכללות בשל היעדר מחקרי תיקוף של ההצלחה בלימודי תואר ראשון במכללות. ככל שיישמע הדבר מפתיע, למרות שמספר התלמידים הלומדים לתואר ראשון במכללות הוא כפול ממספר הלומדים לתואר ראשון באוניברסיטאות, אין בנמצא מחקרי תוקף מקיפים עבור סטודנטים במכללות, בדומה למחקרים הקיימים עבור סטודנטים באוניברסיטאות.

תוקף בחינת יע"ל (ידע בעברית לנבחני שפות זרות)

איילת גולדצויג, יונתן סער
מרכז ארצי לבחינות ולהערכה (מאל"ו)

מבחן יע"ל מיועד להערכת רמת הקריאה והכתיבה בשפה העברית של מועמדים להשכלה הגבוהה, שנבחנו במבחן הכניסה הפסיכומטרי לאוניברסיטאות (מכפ"ל) בשפה שאיננה עברית. הבחינה משמשת למיון המועמדים לרמות עברית, ובמרבית האוניברסיטאות נדרש ציון סף מסוים כתנאי קבלה ללימודים. הבחינה מורכבת מחלק סגור (שאלות רבות-ברירה) ומחלק פתוח (כתיבת חיבור), וניתנים בה שלושה ציונים: סגור, פתוח, וכללי – המורכב משקלול ציוני שני החלקים, כשמשקלו של הסגור כפול ממשקלו של הפתוח.

עד כה נחקרה בחינת יע"ל בעיקר מן ההיבטים הקשורים למהימנות, ומרבית המחקרים מסתמכים על נתונים מן העשור הקודם. מטרת המחקר הנוכחי לבחון היבטים שונים של תוקף בחינת יע"ל על מדגם גדול ועדכני. מרבית הנבחים ביע"ל הם דוברי ערבית ולכן המחקר מתמקד באוכלוסייה זו.

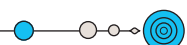
שוער כי רמת הידע בעברית מתואמת במידה רבה עם יכולות הנמדדות במכפ"ל, ולכן נבדקו הקשרים בין הישגים ביע"ל ובמכפ"ל. בנוסף שוער כי ימצא קשר בין יכולת ביטוי בכתב (ציוני החיבור) בשפה ראשונה ושנייה. בהתאם למשוער, נמצא כי המתאם בין ציון יע"ל כללי וציון מכפ"ל כללי – גבוה. מקרב תחומי מכפ"ל, התחום המתואם ביותר עם יע"ל הוא אנגלית, לאחר מכן מילולי, ולבסוף הכמותי. המתאם בין ציוני החיבורים ביע"ל ובמכפ"ל בינוני.

הקשר בין ציוני יע"ל ומאפיינים דמוגרפיים שונים נבדק לאור הקשרים בין רכישת שפה ומעמד סוציו-אקונומי, וכן משום שידוע כי רמת העברית באוכלוסייה הערבית משתנה כתלות במאפיינים דמוגרפיים. מבין המשתנים הדמוגרפיים שנבדקו, המשתנה המתואם ביותר עם יע"ל היה השכלת אם, בקשר חלש-בינוני. הקשר בין המצב הכלכלי ויע"ל היה חלש, ולא נמצא קשר בין ציוני יע"ל וגיל. כמשוער נמצא קשר בין משתנה המין וציון יע"ל לטובת הבנים, אך הוא היה חלש. הפער בין המינים בחיבור היה קטן יותר מבחלק הסגור.

במוקד המחקר עמדה בדיקת תוקף הניבוי של ציוני שנה א' בתואר הראשון. לצורך זה תוקנו ציוני יע"ל לקיצוץ תחום לפי נתוני ציון הסכם. בדומה למחקרים אחרים בעולם, נמצא כי המתאם בין ציון יע"ל ובין הציון בסוף שנה א' חיובי ונמוך. בנוסף נמצא כי ציון יע"ל אינו מוסיף לניבוי ציון שנה א' מעל מה שמנבא ציון הסכם, וכן שהחלק הפתוח ביע"ל אינו מוסיף על החלק הסגור בניבוי ציון שנה א'. בחלוקה לפקולטות, תוקף הניבוי של יע"ל היה גבוה יותר בתחומי הלימוד המסתמכים בעיקר על יכולות מילוליות, והדבר בלט בייחוד בתחום המשפטים.

באמצעות ניתוחי רגרסיה, נבדק האם ליע"ל תוספת בניבוי הצלחה בבחינת מו"ר (מיון לרפואה) ובבחינת מתא"ם (מיון לתארים מתקדמים בפסיכולוגיה) מעבר למה שמנבא ציון מכפ"ל. נמצא כי בניכוי ציון מכפ"ל, ציון יע"ל מוסיף להסבר השונות בציוני מו"ר וכמעט שאינו מוסיף להסבר השונות בציוני מתא"ם.

לבסוף נבדקו קשרים בין יע"ל ומדדים אחרים של ידע בעברית (תוקף מתכנס) ונמצאו מתאמים חיוביים בעוצמה בינונית-חזקה.



האם היבחנות חוזרת באותו נוסח במבחן אדפטיבי באנגלית נותנת יתרון לנבחן?

טוני גוטנטג, אבי אללוף, מיכל באומר, נדב פודולר, מרינה פרונטון
מרכז ארצי לבחינות ולהערכה (מאל"ו)

הציונים במבחנים רבי-ברירה לינאריים משתפרים בהיבחנות חוזרת, ואפקט זה מועצם כשמשמשים באותו נוסח בחינה לעומת נוסחים מקבילים (Kulik, & Bangert, 1984). אך בעוד היבחנות חוזרת באותו נוסח במבחן לינארי משמעה שהנבחן מקבל בדיוק את אותם הפריטים שקיבל בבחינה הקודמת, אין כך הדבר בבחינה אדפטיבית, שבה רק חלק מהפריטים עשוי לחזור, מאחר והם נדגמים ממאגר הפריטים המרכיב את הנוסח. במחקר הנוכחי בדקנו את שכיחות המקרים שבהם נבחנים חוזרים בבחינה אדפטיבית קיבלו את אותו נוסח בהשוואה לחזרה על נוסח מקביל. שאלה זו היא בעלת השלכות מעשיות על העברת הבחינה. חידושו של המחקר הנוכחי הינו עיסוקו בהיבחנות חוזרת בבחינה אדפטיבית, להבדיל מבחינה לינארית, נושא שלא נחקר רבות עד כה, בנוסף לבקרת איכות שגרתית.

המחקר נערך על קבוצת הנבחנים החוזרים בבחינת ידע באנגלית, בחינת אמיר"ם, בחינה ממוחשבת ואדפטיבית המבוססת על מודל IRT תלת פרמטרי. בין השנים 2013-1992 אירעו 199,674 היבחנות ב-11 נוסחים. מתוכן, 152,952 היו בחינות ראשונות והיתר (23.4%) בחינות חוזרות – בהן התמקדנו במחקר הנוכחי; סך-הכול 35,709 נבחנים חוזרים (גיל ממוצע 26.6 שנים, ס.ת. 7.7; 57% נשים). נבדק השיפור בציון בקרב נבחנים שניגשו לבחינה חוזרת עד ארבע פעמים (99.6% מהנבחנים). הפרש הימים החציוני בין העברות עוקבות עומד על כחצי שנה. ככל שמספר החזרות על הבחינה עולה – ממוצע ציוני הנבחנים החוזרים הולך ועולה. נמצא שיש שיפור ציון אצל הנבחנים החוזרים על אותו נוסח בשתי העברות עוקבות (אך לא בהעברות שאינן עוקבות) הגבוה במובהק מזה של מקבלי הנוסח המקביל, אך האפקט הוא קטן מאוד (הפרש מתוקנן = 0.02).

למרות שהסיכוי לקבלת אותו נוסח אינו גבוה, שגם שכשמדובר באותו נוסח הסיכוי לדגימת אותם פריטים בדיוק ממאגר הפריטים הינו נמוך ושההשפעה של קבלת אותו נוסח על הציון קטנה יחסית – יש למצוא דרכים לצמצמה עוד, או אף למנוע חזרה על אותו הנוסח. מחקר עתידי יעסוק בביצועי נבחנים חוזרים בבחינות ידע באנגלית תוך השוואה בין בחינות שונות (אמיר"ם, אמיר"ר וחלק האנגלית של הבחינה הפסיכומטרית) ובין בחינות אדפטיביות ללינאריות. שאלה מעניינת נוספת הנבדקת עתה היא מהו אחוז הפריטים החופף אצל נבחנים החוזרים על נוסח ספציפי בבחינת אמיר"ם, מהו סוגם והאם יש קשר בין אחוז זה לציוני הנבחנים.

בדיקת הקשר בין סולם CEFR לבין סולם אמ"ר

רות פורטס, נעמי גפני
מרכז ארצי לבחינות ולהערכה (מאל"ו)

המחקר נערך במסגרת פרויקט ECOSTAR, פרויקט שממומן ע"י הנציבות האירופית. מטרתו למסד תכנית לימודים כלל-ארצית להוראת אנגלית כשפה זרה במוסדות להשכלה הגבוהה בארץ, שתתבסס על רמות המיומנות של ה-CEFR*, רמות שהפכו לסטנדרט מוכר ברמה בינלאומית. ה-CEFR הוא מסמך שמגדיר בסיס משותף לתכנון תכניות להוראת שפה (כל שפה שהיא), למתן תעודות הסמכה בשפה, וללימוד עצמי של שפה. המסמך מפרט מכלול המלצות והנחיות שאפשר להתאים לצרכי המשתמש. יחד עם זאת, אחת ממטרות המסמך היא לספק תיאור מפורט וסטנדרטי של רמות מיומנות, ובכך לאפשר השוואה בין מערכות הסמכה שונות.

השמה לרמות לימוד באנגלית במוסדות להשכלה גבוהה בארץ נעשית כיום באמצעות ציון באחת משלוש בחינות של המרכז הארצי לבחינות ולהערכה (מאל"ו): הציון ברכיב האנגלית בבחינה הפסיכומטרית, בבחינת אמ"ר או בבחינת אמיר"ם. הציונים בשלוש הבחינות שקולים זה לזה (וייקראו להלן "ציון אמיר"ר"). לעומת זאת, השמה לרמות לימוד המבוססות על סולם CEFR נעשית באמצעות מגוון רחב של מבחנים, שאינם בהכרח שקולים מבחינת תוכנם או רמת הקושי שלהם.

מטרת המחקר להעריך הערכה ראשונית את הקשר בין רמות המיומנות הנקבעות בעזרת הציון בבחינת אמיר"ר לבין רמות המיומנות הנקבעות בעזרת הציון במבחנים רחבי היקף שציונם מקושר לרמות CEFR (מבחן TELC). מבחן ה-TELTC, שהציונים שלו מדווחים בסולם CEFR, הועבר לסטודנטים ישראלים שציון אמיר"ר שלהם ידוע; כמו כן, הועברה בחינת אמיר"ר לסטודנטים אירופאים, שרמת ה-CEFR שלהם ידועה.

בהצגה נפרט את אופן ההתמודדות עם ההבדלים בין מבחן אמיר"ר ומבחני TELC: הבדלים ביכולות הנמדדות בשני המבחנים, והבדלים הנובעים מן העובדה שלבחינת אמיר"ר יש גרסה אחת, המועברת לנבחנים בכל הרמות, בעוד שלמבחני TELC יש כמה גרסאות, כל אחת מותאמת לאחת משש רמות המיומנות של ה-CEFR.

נאספו נתונים מ-1,174 סטודנטים הלומדים בארבע אוניברסיטאות ומכללות בארץ, ומ-353 סטודנטים הלומדים בשלוש אוניברסיטאות באירופה. המתאם בין הציונים באמיר"ר ורמת CEFR היה 0.77 עבור הישראלים ו-0.48 עבור האירופאים. המתאם עבור המדגם כולו היה 0.17. בהתבסס על התוצאות, נבנתה טבלת קישור ראשונית בין שני סולמות הציונים.

הסיבה העיקרית למתאם הנמוך שנמצא במדגם כולו היא העובדה שהשיבוץ לרמות CEFR של המשתתפים לקוח משני מקורות שונים: אצל הסטודנטים הישראלים השיבוץ נגזר מהישגיהם במבחני TELC, ואצל הסטודנטים האירופאים השיבוץ נקבע ע"י המוסדות האירופאים שבהם למדו על סמך מגוון מבחנים לא סטנדרטיים ולא מכילים זה לזה, ואשר מבוססים על פרשנות סובייקטיבית של ה-CEFR.

* CEFR - Common European Framework of Reference for Languages: Learning, Teaching, Assessment

השפעת הניסוח של הוראות במטלות למידה ומבחנים על ביצועי תלמידים

איתמר שץ
אוניברסיטת תל-אביב

רקע: קיימות שתי אוריינטציות מטרה עיקריות שמניעות אנשים במהלך התמודדות עם משימות: מטרת למידה שגורמת לאנשים לשאוף לשפר את היכולות שלהם, ומטרת ביצוע, שגורמת לאנשים לשאוף לקבל משוב חיובי על ביצועיהם (Bell and Kozlowski, 2002; Zimmerman, 2002). אוריינטציית המטרה מושפעת מגורמים חיצוניים, ולכן ניסוח של הוראות במטלות למידה עשוי להשפיע על האופן שבו תלמידים מתמודדים עם מטלות (Steele-Johnson et al., 2000). בנוסף, במידה וההשפעה של הניסוח תלויה בגורמי רקע של התלמידים כגון מין, אז עשויה להיווצר אפליה מגדרית בעקבות השימוש בהוראות שאינן מנוסחות באופן מיטבי.

מתודולוגיה: הנבדקים (N = 526) השתתפו במטלת למידה של אוצר מילים בשפה זרה. במהלך המטלה הנבדקים שינו את הפירוש באנגלית של מילים בשפה זרה (פינית) בתוך פרק זמן קצוב. לאחר מכן, הנבדקים קיבלו את הרשימה של המילים החדשות שהם למדו, והתבקשו לכתוב את התרגום באנגלית של המילים שזכרו, כאשר כל תשובה נכונה זיכתה אותם בנקודה אחת. סדר המילים היה אקראי, ושונה בשלב הלמידה ובשלב הבחינה.

לפני המטלה, הנבדקים חולקו באופן רנדומלי בין שלוש קבוצות. הקבוצה הראשונה קיבלה הוראות בעלות אוריינטציית למידה, שעודדו את הנבדקים להשתדל לשפר את יכולותיהם. הקבוצה השנייה קיבלה הוראות בעלות אוריינטציית ביצוע, שעודדו את הנבדקים להשיג משוב חיובי על הישגיהם. קבוצת הביקורת קיבלה הוראות נייטרליות, ללא תוספת של אוריינטציית מטרה. לאחר סיום המטלה, הנבדקים מילאו שאלון שמדד משתנים שרלוונטיים להליך הלמידה. אלו היו: חרדה, ביטחון עצמי, מוטיבציה (לפני ואחרי קבלת פידבק), ומוכנות לקחת סיכונים (Cervantes, 2013; Liu and Jackson, 2008).

תוצאות: MANOVA הראתה שינוי הבדל מובהק סטטיסטית בין הקבוצות במשתנים שנמדדו. ANOVAs הראו שההבדל מובהק עבור ציון המטלה הממוצע ($p = .021$) ועבור המוכנות לקחת סיכונים ($p = .037$). אנליזת פוסט הוק הראתה שהציון הממוצע במטלה עבור הקבוצה עם אוריינטציית הלמידה היה נמוך ב-8.57% ($p = .013$) מהציון הממוצע של קבוצת הביקורת, ושנטילת הסיכונים הממוצעת של הקבוצה עם אוריינטציית הביצוע הייתה נמוכה ב-6.66% ($p = .011$) מאשר נטילת הסיכונים הממוצעת של קבוצת הביקורת. לא הייתה אינטראקציה בין השפעת ההוראות לבין המין של הנבדקים (כפי שהוראה על ידי two-way MANOVA ו-ANOVAs עוקבות).

דיון: ממצאים אלו מראים שבשני המקרים בהם היה שימוש בהוראות שכללו אוריינטציית מטרה, הייתה פגיעה ביכולת הלמידה של הנבדקים לעומת הנבדקים בקבוצת הביקורת שקיבלו הוראות נייטרליות, ללא התייחסות לאוריינטציית מטרה. בהשוואה לנבדקים בשתי הקבוצות האחרות, לנבדקים בקבוצת הביקורת היו סך הכל את הביצועים הטובים ביותר, מאחר והם לא נפגעו מההשפעה השלילית של ההוראות בעלות אוריינטציית המטרה. ממצאים אלו מדגישים את החשיבות של שימוש בהוראות שמנוסחות באופן נייטרלי, ושאין מדגישות אוריינטציית מטרה ספציפית, אלא מתמקדות במידע שרלוונטי לביצוע המטלה בלבד.

תכנית אבחון מותאם תרבות (אמ"ת)

נועם פישמן
מאירס-ג'וינט מכון ברוקדייל
טלי סמני
בסיס להערכה ומכון שקד
ליאת בסיס
בסיס להערכה

תכנית אבחון מותאם תרבות (אמ"ת) של ג'וינט-תבת הוקמה בשנת 2010 בעקבות דיווחים של מעסיקים על קשיי הקבלה של יוצאי אתיופיה לתפקידים המותנים בהצלחה במבחני מיון. מטרתה העיקרית היו להקטין את חסמי התעסוקה של יוצאי אתיופיה ולהגדיל את האופק התעסוקתי שלהם באמצעות עידוד השימוש של מעסיקים ישראלים בכלי מיון מותאמים-תרבותית לאוכלוסיות מיעוט. התכנית לוותה במחקר מקיף שנערך על ידי מאירס-ג'וינט ברוקדייל בשני שלבים.

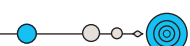
המטרות העיקריות בשלב הראשון של המחקר היו לסקור את השיטות לצמצום פערים בין קבוצות בתהליכי מיון לתעסוקה על פי הספרות המקצועית, ולבחון אם מבחני המיון הקיימים מקשים על מועמדים ממוצא אתיופי בקבלה לעבודה. השלב הזה של המחקר כלל סקירת ספרות בין-לאומית ורכיבים איכותניים. בנוסף, השוונו את הביצועים של 186 מועמדים ממוצא אתיופי לאלה של מועמדים מקבוצת הרוב שהיו דומים להם במגדר, בתפקיד המיועד ובתאריך האבחון, בארבעה כלים קיימים: מבחני כישורים שכליים, מבחן מצב קבוצתי (דינמיקה קבוצתית), ריאיון חצי מובנה וחוות הדעת של המעריך:

- הפערים במבחני הכישורים השכליים בין נבחנים ממוצא אתיופי לנבחנים מקבוצת הרוב היו עמוקים ומשמעותיים. באומדן כישורים שכליים כלליים הפער בין שתי הקבוצות עמד על 1.19 סטיות תקן.
- נמצאו פערים מתונים יותר, אם כי משמעותיים, במבחן מצב קבוצתי ובמדדים שהיו מבוססים על חוות הדעת של המעריך. לא נמצאו הבדלים משמעותיים בריאיון חצי מובנה.

בעקבות ממצאי השלב הראשון, תבת החליט לפתח כלים חדשים רגישי-תרבות כדי שיסייעו להקטין את החסמים העומדים בפני קבוצות מיעוט. בסיוע צוות רב-תרבותי ואנשי מקצוע, פותחו שמונה כלים חדשים, הבוחנים מגוון תחומים: כישורים מספריים, בין-אישיים ועוד. בשלב השני של המחקר ערכנו מחקר תיקוף, כדי לבחון את איכות הכלים החדשים. המטרה הייתה לאתר אילו כלים מאזנים טוב יותר בין יכולת הניבוי של איכות התפקוד בעבודה לבין הקטנת הפערים בין הקבוצות. בסיוע 13 ארגונים, 144 אנשים עובדים השתתפו במחקר: 69 ממוצא אתיופי, 32 מהחברה הערבית ו-43 מקבוצת הרוב. כל המשתתפים עברו יום אבחון שכלל כלי מיון חדשים וכלי מיון קיימים. בנוסף, קבלנו חוות דעת ממנהליהם של המשתתפים לגבי תפקודם בעבודה.

- נמצאו מספר כלים חדשים בעלי יכולת טובה בניבוי תפקוד בעבודה. הפערים שהם מראים קטנים בהרבה מאלה שנמצאו בכלים הקיימים הבוחנים כישורים שכליים, שהם נפוצים ביותר בתהליכי מיון.
- הכלים החדשים עשויים לסייע למעסיקים להגביר את הגיוון בהעסקה, תוך כדי בחירת מועמדים איכותיים.

יש לבחון ממצאים אלה בתשומת לב רבה, כי הם מבוססים על מדגם קטן יחסית שאינו מייצג את האוכלוסייה הכללית של מחפשי עבודה - ממוצא אתיופי ומקבוצת הרוב. כמו כן, יש צורך בבדיקה מעמיקה של הפוטנציאל של כלי מיון נבחרים במחקרים עתידיים עם מדגמים גדולים יותר, ועם משתתפים ממקצועות מגוונים יותר. עם זאת, מחקר זה הוא הראשון מסוגו בישראל, ולמצאיו עשויות להיות השלכות חשובות להרחבת אפשרויות התעסוקה של קבוצות מיעוט ולעידוד הגיוון בהעסקה בישראל.



העדפות שפה בקרב נבחני ערבית בבחינה הפסיכומטרית

כרמל אורן, איילת גולדצויג
מרכז ארצי לבחינות ולהערכה (מאל"ו)

מטרת המחקר: בדיקת הקשר בין מידת החשיפה, הרגלי השימוש, והשליטה בשפה העברית בקרב נבחני ערבית במבחן הכניסה הפסיכומטרי לאוניברסיטאות (מכפ"ל), לבין המשתנים: ציונים במכפ"ל, ציונים בבחינת ידע בעברית (יע"ל), ומאפיינים דמוגרפיים. השערת המחקר הייתה כי ימצא קשר חיובי בין נטייה להעדיף את השפה העברית על פני ערבית בתפקודי יומיום שונים לבין הציון במכפ"ל ובבחינת יע"ל, וכן קשר חיובי עם המעמד החברתי-כלכלי.

כלים: בתום בחינת מכפ"ל, מועבר שאלון משוב שממנו ניתן ללמוד על עמדותיהם של הנבחנים בערבית ועל מידת השימוש שלהם בעברית ובערבית. השאלות שנכללו במשוב היו: באיזה שפה למדת בבית הספר התיכון? באיזה שפה היית מעוניין שילדיך ילמדו בבית הספר? מהי השפה העיקרית שבה כתובים העיתונים ואתרי האינטרנט שהורך קוראים? מהי השפה העיקרית שבה כתובים העיתונים ואתרי האינטרנט שאתה קורא?

מדגם: 1404 נבחנים בערבית שנבחנו לראשונה במכפ"ל ממועד יולי 2013 ועד מועד אפריל 2014, הזדהו והשיבו על השאלות הרלוונטיות בשאלון המשוב. הצלבת התשובות לשאלות אלה במשוב יצרה ארבע קבוצות: נבחנים שדיווחו כי הם למדו בעברית ומעדיפים שילדיהם ילמדו בעברית, למדו בערבית ומעדיפים שילדיהם ילמדו בערבית, למדו בעברית ומעדיפים שילדיהם ילמדו בערבית, למדו בערבית ומעדיפים שילדיהם ילמדו בערבית. ארבע הקבוצות הללו הושאו ואופיינו על פי המשתנים הבאים: ציוני מכפ"ל, ציוני יע"ל, השכלת אם, מצב כלכלי, ואחוז הבנים בקבוצה.

תוצאות: הממצא המרכזי הוא כי רוב הנבחנים במדגם למדו בערבית והיו רוצים שילדיהם ילמדו בערבית. רק מעטים דיווחו כי הם למדו בעברית, וכי היו רוצים שילדיהם ילמדו בעברית (ובהתאם גם יכולת ההכללה מקבוצה קטנה זו).

מי שלמדו בעברית קיבלו ציונים גבוהים יותר ביע"ל, ולא נמצא אפקט לשפה שבה היו מעדיפים שילדיהם ילמדו בבית הספר על ציוני יע"ל. זאת בהתאם להשערות המחקר.

כאשר נבדקו ציוני יע"ל בניכוי ציוני מכפ"ל נמצא גם אפקט אינטראקציה: בקרב מי שלמדו בעברית, לא היה הבדל מובהק בציונים כתלות בשפת הלימוד המועדפת עבור הילדים, ואילו בקרב מי שלמדו בערבית, הציונים היו גבוהים יותר בקרב המעדיפים שילדיהם ילמדו בעברית. במידה מסוימת גם ממצא זה עולה בקנה אחד עם השערות המחקר. לעומת זאת, בבדיקת ציוני מכפ"ל והשכלת אם נמצאו הבדלים עקביים בין נבחנים שהעדיפו שילדיהם ילמדו בעברית ונבחנים שהעדיפו שילדיהם ילמדו בערבית: לקבוצה שהעדיפה שילדיהם ילמדו בערבית ציונים גבוהים יותר במכפ"ל, וממוצע השכלת האם בה היה גבוה יותר. ההבדל בהשכלת האם לא הסביר את כל ההפרש בין הקבוצות בציוני מכפ"ל, והפרש זה גם לא הוסבר על ידי המצב הכלכלי ואחוז הבנים בקבוצה. בבדיקת משתנה המצב הכלכלי נמצאה אינטראקציה: בקרב מי שלמדו בעצמם בעברית, המצב הכלכלי של נבחנים שהעדיפו שילדיהם ילמדו בערבית גבוה בממוצע משל אלו שהעדיפו שילדיהם ילמדו בעברית. בקרב מי שלמדו בעצמם בערבית, המצב הכלכלי של הנבחנים שהעדיפו שילדיהם ילמדו בעברית דומה לשל אלה שהעדיפו שילדיהם ילמדו בערבית. ממצאים אלה אינם תואמים את השערות המחקר.

קבלה להשכלה גבוהה וצדק חברתי - דוברי ערבית ועולים חדשים והבחינה הפסיכומטרית

נעמי גפני, אבי אללוף, ראובן סטולר
מרכז ארצי לבחינות ולהערכה (מאל"ו)

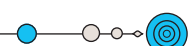
הבחינה הפסיכומטרית יחד עם הבגרות משמשת מכשיר למיון מועמדים למוסדות להשכלה גבוהה בישראל. הבחינה הפסיכומטרית היא בחינת יכולת לימוד והיא כוללת מטלת כתיבה ושמונה פרקים רבי בררה מתוך שלושה תחומים: חשיבה מילולית, חשיבה כמותית ואנגלית כשפה זרה. בבחינה שישה פרקים תפעוליים, שניים מכל תחום ושני פרקים שמועברים לצרכים סטטיסטיים. לציון בבחינה חשיבות רבה גם עבור מועמדים מחו"ל שאינם מחזיקים בתעודת בגרות ישראלית.

בישראל, כידוע, מיעוט ערבי גדול (21%) ובנוסף המדינה היא ארץ קולטת עלייה. כבר בשנות ה-70 באוניברסיטה העברית ובשנות השמונים עם הקמתו של מאל"ו הוחלט לבחון דוברי ערבית בשפה הערבית ולבחון עולים חדשים בשפתם וזאת מתוך תפיסה כי בחינה בשפת האם משקפת באופן המדויק ביותר את היכולת הנמדדת. עם זאת, הלימודים במוסדות מתנהלים בעברית ולכן נדרשת גם שליטה בשפה העברית אשר נבדקת בעזרת מבחן נפרד - בחינת עברית לדוברי שפות זרות (בחינת יעל).

בשנת 2014 תורגמה והותאמה הבחינה הפסיכומטרית לשפות הבאות: ערבית, רוסית, צרפתית, ספרדית, איטלקית ופורטוגלית. בנוסף פותח במאל"ו נוסח משולב שכולל זה לצד זה נוסח בעברית ונוסח באנגלית ולקסיקון מילים קשות מתורגמות לשמונה שפות. נוסח זה מיועד למועמדים ששפת אימם אינה אף אחת מהשפות לעיל אך הם שולטים במידה מסוימת באנגלית או בעברית. בשנת 2014 נבחנו כ-65,000 נבחנים מתוכם 36% נבחנו בשפה שאיננה עברית. המועמדים בוחרים את שפת הבחינה המועדפת עליהם.

קיימים הבדלים ניכרים בין קבוצות הנבחנים בביצוע בבחינה הפסיכומטרית. מטרות המחקר הנוכחי הן: (1) לתאר את האמצעים בהם מאלו נוקט על מנת להפוך את הבחינה להוגנת ככל האפשר כלפי קבוצות השפה השונות; (2) לבדוק באיזו מידה משתני רקע שונים ובלתי רלבנטיים ליכולת הנמדדת עשויים להסביר הבדלי ביצוע בין הקבוצות; ו- (3) להציג קריטריונים לשיפוט מידת ההוגנות ותוצאות של בדיקות הוגנות כלפי קבוצות השפה. מטבע הדברים הדגש מושם על קבוצות הנבחנים בעברית ובערבית המהווים את הרוב הגדול של נבחני הבחינה הפסיכומטרית.

קבוצות השפה, ובמיוחד נבחנים בעברית ובערבית, שונים מאד בהישגיהם. נשאלת השאלה למה ניתן לייחס הבדלים אלה. באיזו מידה הם משקפים הבדלים אמיתיים בין הקבוצות, והאם הם גם פונקציה של הטיית וגורמים אחרים שאינם רלבנטיים. על מנת לענות על שאלה זו נבדקו כמה ממשתני הרקע (הלא רלבנטיים למדידת יכולת למידה) שעשויים להסביר את ההבדלים בממוצע הישגים בין הנבחנים בשפות השונות, כגון שיעור הנבחנים החוזרים בכל קבוצה, הבדלים בדפוס ההכנה לבחינה בין הקבוצות, שיעור הנשים והגברים בכל קבוצה, והבדלים בין הקבוצות בשיעור מבקשי ומקבלי התאמות. על מנת לענות על שאלת ההוגנות נבדקו נתוני מהימנות ותוקף בקרב הקבוצות השונות. נמצא כי המהימנות הפנימית הן של הציון הכללי בבחינה ווהן של מרכיביה גבוהה - מעל 0.90 - עבור הנוסחים בכל השפות. כפי שנמצא בקרב השחורים והלבנים בארה"ב עבור בחינת ה-SAT, גם הבחינה הפסיכומטרית נמצאה תקפה יותר עבור הנבחנים בעברית ופחות תקפה עבור הנבחנים בערבית. כמו כן, נמצא כי הפערים בין הקבוצות בבחינה הפסיכומטרית הם באותו סדר גודל כמו ההבדלים בין קבוצות בהישגים בלימודים אקדמיים ובמבחנים אחרים במערכת החינוך כולל מבחנים בינלאומיים. נראה כי אפשר להסיק שלפחות רובו המכריע של הפער בין קבוצות ניתן לייחס להבדלים אמיתיים בתכונה הנמדדת, ורק מקצתו נובע ממשתנים לא רלבנטיים הקשורים לביצוע.



היבטים פסיכומטריים בהכשרה להוראה וקריירה בהוראה

דוד מעגן הלשכה המרכזית לסטטיסטיקה

במכללות האקדמיות לחינוך, הוחלה חובת הבחינה הפסיכומטרית משנת 2006 ואילך כתנאי לקבלה ללימודי תואר ראשון "בוגר בהוראה" (B.Ed). כתוצאה מתהליכים אלו, בשנים האחרונות נעשה שימוש הולך וגדל בנתונים על הבחינה הפסיכומטרית בסטטיסטיקה של כוחות הוראה, הכשרה וכניסה להוראה ונשירה מהמקצוע.

הבחינה הפסיכומטרית כהגדרתה הינה כלי לחיזוי סיכויי ההצלחה בלימודים במוסדות להשכלה גבוהה. ממצאי מחקרים מעלים כי להישג בבחינה הפסיכומטרית יכולת ניבוי ההצלחה בלימודי שנה א' (קנת-כהן ושות', 2013; 2014) והצלחה בלימודים וקבלת תואר ראשון (קלפר ושות', 2014; מעגן ודדש, 2012).

בקרב חוקרים קיימת הסכמה רחבה כי לאיכות המורים רלוונטיות גבוהה באשר להישגי התלמידים. יחד עם זאת, הקושי נעוץ בניסיון להגדיר מהו מדד טוב לאיכות המורים. במחקרים שעוסקים בשאלות על איכות מורים, נעשה שימוש במספר מאפייני המורים על מנת לאמוד את איכותם, ביניהם השכיחים ביותר הינם, מאפייני רקע השכלתיים ויכולת אקדמית במבחנים סטנדרטיים.

מחקרים שבחנו את הכניסה להוראה וההתמדה במקצוע בקרב בוגרים ומורים חדשים על-פי הישגיהם במבחני קבלה למוסדות להשכלה גבוהה (ACT/SAT), נמצא כי לבעלי הישגים גבוהים סיכוי נמוך יותר להיכנס למקצוע ומבין אלו שנכנסו להוראה, לבעלי ההישגים הגבוהים סיכוי גבוה יותר לעזוב את המקצוע. עוד נמצא כי למורים חדשים בעלי כישורים גבוהים סיכוי גדול יותר לנשור מהמקצוע (Podgursky, Monroe and Watson, 2004; Lankford, Lobe and Wyckoff, 2002; Stinebrickner, 2001).

בהתאם לכך לניתוח המוצע מספרת מטרות, להלן:

- בשלב ראשון, נרצה לבחון האם ועד כמה להישג הפסיכומטרי יכולת ניבוי לקבלת תואר ראשון במכללות לחינוך בישראל, זאת בהשוואה ליכולת הניבוי לקבלת תואר ראשון באוניברסיטאות ובמכללות האקדמיות הכלליות.
- בהמשך, נרצה לבחון האם ועד כמה להישג הפסיכומטרי קשר עם הסיכוי להיכנס למקצוע ההוראה, להתמיד בו במהלך השנים הראשונות לקריירה בהוראה וכן להתקדם לתפקידי ניהול בת-ספר.

מתודולוגיה: מערך הנתונים למחקר כלל מספר קבצים מהעשור האחרון, על סטודנטים ומקבלי תארים במוסדות להשכלה גבוהה, נבחנו בבחינה הפסיכומטרית ובבחינות הבגרות וכוחות הוראה במערכת החינוך.

ממצאים עיקריים:

מעבר למסגרות השכלה גבוהה אחרות: כ-15% במוצע מבין הסטודנטים שנה א' במכללות לחינוך, עברו למסגרות השכלה גבוהה אחרות. לסטודנטים בעלי הישגים גבוהים בבחינה הפסיכומטרית, סיכוי גבוה משמעותית ($se=0.42$) לעשות כן.

קבלת תואר: ממצאי המודל על הסיכוי לקבלת תואר ראשון בקרב סטודנטים שנה א' מעלים כי בשונה ממסגרות השכלה גבוהה אחרות, במכללות לחינוך להישג הפסיכומטרי תרומה שולית ($se=-0.019$) ברמת מובהקות נמוכה יחסית ($p=0.040$) לסיכוי לקבלת תואר. עוד עולה ממצאי המודל, כי בכל המסגרות להשכלה

גבוהה, לציון הבגרות תרומה גדולה יותר משמעותית לסיכוי לקבלת תואר ביחס לציון הפסיכומטרי ($se=0.10-0.24$ לעומת $se<0.06$). ממצאים דומים נמצאו גם במחקר קודם, על יכולת גבוהה של הישגי הבגרות ביחס להישג הפסיכומטרי בניבוי הישגים לתואר ראשון באוניברסיטאות (מעגן ודדש, 2012).

להישג הפסיכומטרי קשר שלילי משמעותי ($se=-0.161$) ומובהק ($p < 0.0001$) עם הסיכוי להיכנס להוראה. קשר חיובי ($se=0.103$) ומובהק ($p < 0.0001$) עם הסיכוי לנשור מהמקצוע. וקשר שלילי ($se=-0.084$) ומובהק ($p=0.001$) עם הסיכוי לקידום לתפקידי ניהול בתי-ספר (מנהלים/סגני מנהלים).

לסיכום, על-פי ממצאי הניתוח, גם בישראל בדומה לממצאי מחקרים בעולם, ניתן לזהות מגמת פחות מתמשכת בפרופיל הכישורים של האוכלוסייה הפונה להכשרה להוראה, לאורך השלבים השונים שבין שלב ההכשרה להוראה דרך הכניסה להוראה ועד לקריירה בהוראה.

ממצאי הניתוח תורמים להרחבת ההבנה והמיגבלות בתהליכי המיון והקבלה למוסדות להשכלה גבוהה בישראל בכלל ולתהליכי ההכשרה וקריירה בהוראה בפרט.

פיתוח והשמה של מערכת מיון מבוססת מרכז הערכה עבור מכון "אבני ראשה" למיון המועמדים לתוכנית הכשרת פרחי ניהול בתי ספר

דוד ציגלר, אביטל מושינסקי
מרכז ארצי לבחינות ולהערכה (מאל"ו)

ליסה לוי, איתי סודרי
אבני ראשה

הלנה קימרון, חני שלטון
הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

החל משנת תשע"ד מפעיל "אבני ראשה" – המכון הישראלי למנהיגות בית ספרית – מערכת חדשה למיון פרחי ניהול לתכניות הכשרת מנהלי בתי ספר. ההחלטה לפתח מערכת מיון והערכה נבעה מחוסר שביעות רצון מהליך המיון שנעשה בו שימוש קודם לכן הן בשל שונות נמוכה של ציוני המועמדים והן בשל המספר הנמוך של בוגרי תוכניות ההכשרה שהגישו את מועמדותם למכרזי ניהול בתי ספר.

מערכת המיון פותחה בשיתוף פעולה של המרכז הארצי לבחינות ולהערכה (מאל"ו), מכון "אבני ראשה", הראשות הארצית למדידה והערכה (ראמ"ה) ומשרד החינוך.

לצורך פיתוח המערכת הוגדרו המשתנים שיימדדו במערכת המיון. משתנים אלו נסמכו על תפיסת תפקיד מנהל בית הספר כפי שנוסחה על ידי מכון "אבני ראשה" ועל ידי משרד החינוך. היכולות אותן הוחלט להעריך היו יכולות תקשורת בין אישית, הובלה, תפיסת מורכבות, מודעות וביקורת עצמית. לאור זאת הוחלט על פיתוח שלושה כלי מדידה שבאמצעותם אפשר למדוד את התכונות הללו: שאלון חוות דעת מובנה, שאלון אישי-ביוגרפי, ומרכז הערכה.

מערכות מיון הכוללות שילוב של מרכזי הערכה ושאלונים נפוצות בשנים האחרונות למיון מועמדים ללימודים ולהשמה לעבודה, בעיקר לתפקידי ניהול. מערכות מיון דומות שפותחו על ידי מאל"ו ואשר נמצאות בשימוש למעלה מעשר שנים, כגון, מערכות מיון מועמדים לבתי הספר לרפואה בישראל, הוצגו בעבר בכנסי אפי.

בהרצאה תוצג מערכת המיון, הרציונאל שעומד בבסיסה, מרכיבי השונים ואופן פיתוחה והפעלתה. בנוסף, יוצגו נתונים סטיסטיים בנוגע למהימנות מערכת המיון בהשוואה למערכות אחרות שנועדו להעריך משתנים לא קוגניטיביים, וכן נתונים אודות היבחנות חוזרת במערכת המיון ואודות השפעה אפשרית של הכנה לבחינה ושל היכרות מוקדמת עם פריטיה. לסיכום, יוצגו אינדיקציות ראשוניות ליעילות המערכת ותוכניות עתידיות לבדיקת תוקף רחבה יותר.

פיתוח כלי למדידת מעמד המורה בחברה הישראלית

זהבית כהן, יצחק גילת
מכללת לוינסקי לחינוך

רקע ומטרה: לנוכח העיסוק המוגבר בסוגית מעמד המורה בחברה, המחקר הנוכחי הציב לעצמו כמטרה לפתח כלי כמותי למדידת מעמדו של המורה בחברה הישראלית.

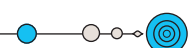
שיטת המחקר: המדגם כלל 500 משיבים המהווים מדגם מייצג של המגזר היהודי. הייצוג המגדרי היה מאוזן, גיל המשיבים הממוצע הינו כ-43 ($SD=15.41$), ממוצע שנות הלימוד של הנבדקים הינו כ-14 שנים ($SD=2.61$). כלי המדידה היה שאלון עמדות שפותח במיוחד לצורך המחקר. השאלון נבנה כך, שניתן יהיה למדוד באמצעותו מעמד של כל מקצוע בחברה הישראלית. המעמד הוגדר באמצעות שבעה קריטריונים שהוסכמו בקבוצת מיקוד של אנשי מקצוע במדעי החברה:

א. תרומתו של בעל המקצוע לחברה ב. רמת המקצוענות הנדרשת במקצוע ג. המידה בה המקצוע מאפשר התפתחות אישית ד. המידה בה בעל המקצוע זוכה לשכר ראוי ה. המידה בה המשיב היה רוצה שילדו יעסוק במקצוע ו. מעמד המקצוע באפן כללי בעיני המשיב ז. מעמד המקצוע באופן כללי, בעיני החברה. כדי שניתן יהיה למקם את מעמדו של המורה ביחס למקצועות אחרים, נקבעו חמישה מקצועות המהווים קבוצת השוואה רלוונטית למדידת מעמדו של המורה: רופא, עורך דין, איש הייטק, עובד סוציאלי ואיש צבא קבע.

המשיבים התבקשו להעריך כל אחד מן המקצועות האלה על כל אחד מן הקריטריונים באמצעות סולם ליקרט בן שש דרגות (1 = נמוך מאד, 6 = גבוה מאד). השאלון כלל גם פרטים על מאפיינים סוציו-דמוגרפים של המשיבים.

שיטות ניתוח הנתונים: חושבו ממוצעים וסטיות תקן של כל אחד מן הקריטריונים, לגבי כל אחד מן המקצועות הכלולים בשאלון. חושב מדד כללי של מעמד המקצוע באמצעות הציונים בכל הקריטריונים. נערכו ניתוחי שונות עם מדידות חוזרות, לגבי כל אחד מן הקריטריונים, כדי לבחון את מובהקות האפקט של המקצוע על הציון בקריטריון ונערכו השוואות מרובות לאיתור מקורות ההבדלים בין המקצועות השונים.

תוצאות: הניתוח הניב שני מצאים עיקריים. האחד - מעמדו של המורה הוא בינוני: נמצאו הבדלים מובהקים בעמדות כלפי מעמד המקצועות השונים, כך שמעמד הרופאים הוא הגבוה ביותר ($M=5.11, SD=.63$), לאחריו מעמד אנשי הייטק ($M=4.88, SD=.65$) המייצג עמדה הקרובה לגבוהה, לאחריו מעמד עורכי-הדין ($M=4.28, SD=.81$) המייצג עמדה די גבוהה, לאחריו מעמד המורים ($M=3.86, SD=.71$) ואנשי צבא הקבע ($M=3.89, SD=.87$) המייצגים עמדה בינונית, ולבסוף מעמדם של העובדים הסוציאליים ($M=3.61, SD=.78$) המייצג עמדה די נמוכה עד בינונית. הממצא השני: הפרופיל של המורה על פני הקריטריונים השונים הוא תנודתי יותר מאשר המקצועות האחרים, כפי שמראה הלוח הבא:



המידה שבה השכר שמקבל בעל המקצוע הוא ראוי	המידה שבה המשיבים הביעו רצון שהילד שלהם יעסוק במקצוע	המידה שבה המקצוע מאפשר התפתחות אישית	מעמד בעל המקצוע בעיני המשיבים	רמת המקצוענות הדרושה מבעל המקצוע	תרומתו של בעל המקצוע לחברה	מעמד בעל המקצוע בחברה הישראלית
איש הייטק	איש הייטק	רופא	רופא	רופא	רופא	רופא
רופא	רופא	איש הייטק	איש הייטק	איש הייטק	מורה	איש הייטק
עורך דין	עורך דין	עורך דין	עורך דין	מורה	עובד סוציאלי	עורך דין
איש צבא קבע	מורה	מורה	מורה	עורך דין	איש הייטק	איש צבא קבע
מורה	איש צבא קבע	עובד סוציאלי	עובד סוציאלי	עובד סוציאלי	איש צבא קבע	מורה
עובד סוציאלי	עובד סוציאלי	איש צבא קבע	איש צבא קבע	איש צבא קבע	עורך דין	עובד סוציאלי

חשיבות המחקר: הכלי למדידת מעמד חברתי עשוי להעשיר את הידע על אודות המעמד החברתי של מורים ושל בעלי מקצועות אחרים באמצעות מחקרים עתידיים.

פיתוח מדד יחיד לאיכות מעריכים של מבחני ביצוע

אבי אללוף, דביר קלפר, האשם ניג'ם, ג'ני שמולביץ
מרכז ארצי לבחינות ולהערכה (מאל"ו)

מבוא ומטרות

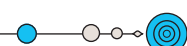
שמירה על מהימנות ואובייקטיביות של ציונים על מבחני ביצוע דורשת מחוון ברור ומעריכים מיומנים. עם זאת, לא ניתן לזהות מראש מי יהיה מעריך איכותי (Leckie and Baird, 2011). בקרה במהלך ההערכה, או עם סיומה, עשויה לאפשר את איתור המעריכים האיכותיים. הספרות מתארת מדדים סטטיסטיים לאיכות מעריכים (למשל, Myford and Wolfe, 2009) אך לא נעשה ניסיון ליצור מדד יחיד. מחקר זה עסק באפיון מעריכים של מבחני ביצוע, בהסתמך על מדדי איכות ויעילות.

שיטה

מבחן: המחקר נעשה על מטלת הכתיבה המהווה חלק מהתחום המילולי בבחינה הפסיכומטרית של מאל"ו המשמשת למיון מועמדים להשכלה גבוהה וכוללת גם חלקים מילולי וכמותי. החיבור מוערך על פני שני ממדים – תוכן ולשון, והציון הוא סכום ציוני שני מעריכים. **נתונים:** מאגר נתונים ייחודי (Cohen, 2014) – 500 חיבורים שנדגמו מקרית, חולקו ל-2 קבוצות של 250 חיבורים בכל אחת, ונבדקו על ידי 14 ו-15 מעריכים, בהתאמה. ההערכות המרובות לחיבור שימשו לאמידת הציונים האמיתיים לכל חיבור. **מדדי איכות:** שבעה מדדי איכות חושבו לכל מעריך – (1-3) מתאמים בין ציוניו לציונים האמיתיים, (4-6) ממוצעי פערים בין ציוניו לציונים האמיתיים, (7) מספר חיבורים למעריך בשעה (מדד יעילות). ניתוח: PCA (Principal Component Analysis) מבטא מאגר נתונים מורכב באמצעות מספר משתנים קטן. במחקר זה, הניתוח ביטא את מדדי האיכות – כולם או חלקם – ע"י מדד יחיד. הניתוח מספק את המשקלים היחסיים של כל מדד, ומחשב את אחוז השונות המוסברת (POV) על ידי המדד היחיד, ובכך מסייע בבחירת המדדים הרלוונטיים והמודל לחישובם.

תוצאות והשלכות מעשיות

שמונה מודלים הכוללים צירוף שונה של מדדים הושוּו באמצעות שלושה קריטריונים: אחוז שונות מוסברת (POV), העלאה (Lift), היחס בין POV לבין ממוצע השונות למדד) ועוצמה (POV בהפחתת ההעלאה). כמקובל ב-PCA, לא נותחו מודלים שבהם אחד המדדים מהווה סכום של שני מדדים אחרים. שניים מהמודלים נמצאו עדיפים, אחד מהם כלל את מדד היעילות. התוצאות שימשו לדרוג איכות מעריכי המחקר, והשוו לדירוגם על סמך נתונים דומים שחושבו מהערכות תפעוליות של מעריכים אלה. המתאם שנמצא היה גבוה – 0.84. בנוסף, המודלים סייעו לדרג 127 איש ממאגר המעריכים הכללי, כך שניתן יהיה בעתיד לבצע החלטות מושכלות באשר לזימונם להערכה תפעולית. המתודולוגיה שפותחה במחקר זה מאפשרת בקרה ודירוג של מעריכים גם במבחני ביצוע אחרים.



שיפור המהימנות של מערכי חיבורים: בחינת כלי מהספרות על חוכמת ההמונים

מאיר ברנרון, אילן יניב
האוניברסיטה העברית בירושלים
אבי אללוך
מרכז ארצי לבחינות ולהערכה (מאל"ו)

College candidates taking the Psychometric Entrance Test (the Israeli SAT equivalent) are required to write a short essay. This task tests the candidate's academic writing skills and is a central component of the test. The essays are traditionally evaluated and graded by raters who are well-trained for this task.

Given the importance of obtaining reliable and accurate evaluations, the common practice is to average the evaluations of two independent raters. This practice is known to improve the reliability in performance assessment tests.

The National Institute for Testing and Evaluation, in charge of the PET, accepts essays written in a dozen of foreign languages. The rationale is to make the entry into institutions of higher education accessible to candidates from various backgrounds and is based on the assumption that an examinee expresses his/her writing skills better in her mother tongue. Yet, in some languages (e.g., Amharic) multiple well-trained raters are hard to find. This raises the question of whether there is a method to improve the accuracy of grades based on a single rater.

Recent research in the field of judgment and decision-making suggests that judgments accuracy could be improved by eliciting multiple judgments from the same individual (at different times), rather than by eliciting single judgments from multiple individuals. This "wisdom-of-crowd effect" within the mind of a single individual implies that essay evaluations made by the same rater at two different occasions should be more accurate than a grade based on a single evaluation.

Our study used professional raters and real essays. We found robust evidence for benefits of this method. The project is unique in that it incorporates ideas from the judgment and decision-making literature into the field of assessment and evaluation, suggesting a noteworthy application that should be considered in attempt to improve the reliability of evaluations.

התנהגויות לא נאותות בקרב עובדים בעלי כישורי יתר

סול פיין

מידות

מיכל אדווארד

אוניברסיטת חיפה

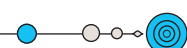
Past research has found that overqualified workers tend to maintain relatively high levels of task performance, but that this performance is often accompanied by negative affect, such as feelings of boredom and job dissatisfaction. And, while negative affect has been found to be independently associated with counterproductive work behaviors (CWB), very little is known about counterproductivity among overqualified workers.

The present study investigated the relationship between two types of overqualification, and employees' involvement in various forms of CWB. Specifically, we looked at both objective overqualification (OOQ; i.e., having extremely high test scores) and perceived overqualification (POQ; i.e., the subjective belief of being overqualified), and their effects on combinations of four categories of CWB: minor CWB, serious CWB, organization-oriented CWB, and interpersonal-oriented CWB. It was hypothesized that overqualification would be related to minor CWBs that target the organization (e.g., taking long breaks, arriving late) as opposed to serious CWBs that target the organization (e.g., theft, fraud) or other employees (e.g., harassment, bullying).

In addition, since job dissatisfaction is the underlying antecedent for CWB, it was further hypothesized that job satisfaction will mediate the relationship between overqualification and CWB. Finally, even among dissatisfied overqualified employees, personality traits may ultimately dictate who decides to engage in CWB in response to their situations. Therefore, the personality trait of conscientiousness, which is one of the strongest individual predictors of CWB, was expected to moderate the secondary relationship between job satisfaction and CWB.

This study included a sample of 166 customer service and sales representatives. OOQ was measured as the gap between participants' general mental ability (GMA) and the required level of GMA for their jobs, as defined by job analyses. POQ, job satisfaction, and conscientiousness were measured using well-known self-report questionnaires. CWB was measured by self-admissions of past behaviors.

OOQ was found to be associated with subsequent POQ ($r = .25$), and POQ was in-turn associated with job dissatisfaction ($r = -.39$), as well as with minor ($r = .26$) and organization-oriented CWB ($r = .29$), but not with serious ($r = .10$) or interpersonal-oriented CWB ($r = .08$), as hypothesized. The direct relationship between OOQ and organizational CWB was weaker ($r \leq .16$). Job satisfaction partially mediated the relationship between POQ and CWB, and conscientiousness fully mediated the relationship between job satisfaction and CWB, while POQ remained significant even after controlling for both mediators.



This is one of the first studies to offer a predictive model for explaining how OOQ can predict POQ, and the subsequent involvement in specific types of CWB. Overall, the present study facilitates a better understanding of the potential risks associated with hiring overqualified job applicants, such that overqualified employees may tend to break organizations' minor rules, but are unlikely to engage in serious or illegal acts. In addition, organizational interventions aimed at minimizing overqualified employees' negative affect may help to reduce their involvement in CWB.

הערכת מוכנות ללימודים אקדמיים

מיכל פיליפס-ברנשטיין, איתמר גתי
האוניברסיטה העברית בירושלים

נועה סקה
מרכז ארצי לבחינות ולהערכה (מאל"ו)

רקע. נשירה מלימודים אקדמיים היא בעיה חמורה בהיקף ומשמעות. לנשירה מהלימודים השלכות שליליות כלכליות, פסיכולוגיות, וחברתיות על הסטודנט, אך הן נוגעות גם למוסד הלימודים ולהיבטים חברתיים רחבים יותר. הכניסה ללימודים אקדמיים מהווה שינוי משמעותי עבור הסטודנט החדש. ההשתלבות בסביבה החדשה כרוכה לא רק באתגרים אקדמיים אלא גם במגוון אתגרים אישיים. סטודנטים רבים אינם מצליחים להתמודד עם האתגרים האלה בצורה מוצלחת ונושרים מהלימודים, כבר בתחילת השנה, במהלך או בסוף הסמסטר הראשון, ואף בסוף השנה הראשונה ללימודיהם האקדמיים.

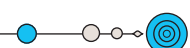
נשירה נובעת בין היתר מהגורמים הבאים: יכולת קוגניטיבית-אקדמית, תהליך בחירה לא מוצלח בתחום הלימודים, נסיבות אישיות (כלכליות, אישיות, משפחתיות וכו') ומשאבים אישיים פסיכו-חברתיים. אלה האחרונים, מבטאים את מידת המוכנות של הסטודנט ללימודים אקדמיים, וקשורים לרמת הסיכון לנשירה, הם מוקד המחקר הנוכחי.

המטרה. כחלק מהמאמצים המתמשכים של מערכת ההשכלה הגבוהה לצמצם את אחוז הנשירה, פותח שאלון אבחון לדיווח עצמי שבאמצעותו ניתן יהיה לזהות סטודנטים שנמצאים בסכנת נשירה, הנובעת מחוסר מוכנות, מעבר ליכולת האקדמית הבאה לידי ביטוי בעמידה בדרישות הקבלה. שאלון ה-מוכנות ללימודים אקדמיים (RAS - Readiness for Academic Studies) מתמקד בהערכת גורמים פסיכו-סוציאליים אשר קשורים בהתמדה בלימודים אקדמיים. השאלון מתייחס לששה מימדים: מסוגלות עצמית אקדמית, מחויבות לימודית, נינוחות חברתית, מעורבות בחיים בקמפוס, משמעת עצמית וחוסן רגשי. מימדים אלה מתבססים על מחקרים שנערכו בארה"ב.

השאלון מאפשר למפות חוזקות וחולשות בתוך הפרט, באופן יחסי לקבוצת ההשתייכות שלו ויחסית לעמיתיו במוסד האקדמי בו הוא לומד. איתור מוקדם עוד לפני תחילת הלימודים של סטודנטים בסיכון לנשירה הנובעת מחוסר מוכנות אישית, מאפשר התערבות שתתמקד בחיזוק הסטודנט בתחומים הנדרשים, ובכך לצמצם את אחוזי הנשירה בשנה א' בלימודים האקדמיים.

השערת מחקר. (1) לשאלון המוכנות RAS ובפרט לששת הסולמות שלו, תכונות פסיכומטריות טובות. (2) השאלון יכול לנבא את הסיכויים לנשירה.

שיטה. נאספו 4,339 שאלונים בשבועיים הראשונים של הלימודים מסטודנטים ממגוון מוסדות אקדמיים בשני מחזורים (2013-2014).



ממצאים. נמצאה מהימנות פנימית טובה של ששת סולמות השאלון ($r > .78$), ניתוח הפריטים הראה שהמתאמים בין פריט לסולם שלו (כשהפריט לא נכלל בסולם) גבוהים תמיד ממתאמיו עם הסולמות האחרים). ניתוח גורמים על מחזור 2013 מצביע על מבנה של שישה גורמים, אחוז השונות המוסברת לאוניברסיטה הוא 67.52% (N=1246) ולמכללות 65.35% (N=1092). ניתוח גורמים מאשש תמך במבנה של ששה מימדים, אך לא בהצדקת האגרזיה שלהם לציון כללי אחד (RMSEA = .041, GFI = .94, CFI = .97).

נתוני פיילוט שבו הוצבה הנשירה כקריטריון, מראים כי סטודנטים מתמידים הם בעלי ציוני RAS גבוהים יותר מציוני RAS של סטודנטים שנשרו וכן כי הסיכוי לנשירה של סטודנט בעל ציון RAS נמוך, גבוה פי 2.5 לעומת סטודנט בעל ציון RAS גבוה.

מסקנות. ממצאים אלה תומכים בטענה שניתן לזהות סטודנטים בסיכון לנשירה באמצעות ה-RAS עוד לפני תחילת הלימודים. בנוסף, ממצאי מחקר פיילוט מראים שלפעולות התערבות והעצמה לפני תחילת הלימודים, או מיד עם תחילתם, המתמקדים בתלמידים בסיכון בהתאם למוקדי חוסר המוכנות שלהם כפי שבאו לידי ביטוי בשאלון ה-RAS, יש תרומה בצמצום היקף הנשירה.

הנשירה הסמויה בישראל: ניתוק בקרב תלמידי בתי הספר

רות ברוך-קוברסקי, דליה בן רבי, וצ'סלב קונסטנטינוב
מאירס-ג'וינט-מכון ברוקדייל

רקע: חוזר מנכ"ל משרד החינוך (1994) מגדיר את ה"נשירה הסמויה" כמצב בו התלמידים נוכחים פיסית בכיתה אך אינם מתפקדים לימודית, מתנתקים מהמסגרות הקיימות וחשים ניכור כלפי הכיתה ובית הספר. בספרות המחקרית מקובל להתייחס כיום למונח "engagement", ולהיפוכו, "Disengagement" (חוסר חיבור, ניתוק), המבטא את התנהגותו של התלמיד ותחושותיו כלפי בית הספר. מחקרים שונים מדברים על מידת ההתקשרות של התלמיד לבית הספר כגורם המוביל לנשירה מוחלטת מבית הספר. עם זאת, ישנה גם עלייה במודעות לכך שניתוק, מעבר להשפעתו על הסיכויים לנשירה, הוא מצב המשפיע לרעה על רווחת התלמיד ועל המידה בה הוא מפיק את המירב מתהליך הלמידה.

למרות המודעות הגוברת לתופעת הנשירה הסמויה והניתוק, חסרים נתונים על היקפה ונראה כי הסיבה לכך היא הקושי המתודולוגי בניטור התופעה, והעובדה שמעצם אופייה, מדידתה כרוכה בבחינת תחושות סובייקטיביות של התלמידים ותפיסותיהם, ולא רק בדיווח חיצוני על התנהגותם.

מטרת המחקר: גיבוש הגדרה אופרציונלית של תופעת הנשירה הסמויה, שתאפשר לספק לבעלי עניין במערכת החינוך תמונה עדכנית על התופעה ועל מאפייניה, כבסיס לפיתוח מדיניות ודרכי התערבות ולהמשך ניטור התופעה בעתיד.

שיטת המחקר: המדד גובש על בסיס נתונים מבחן פיז"ה שהועבר בשנת 2012 לתלמידים בני 15 במדגם מייצג של בתי ספר בישראל וב-34 מדינות נוספות. בשלב הראשון, ועל סמך הספרות המחקרית המתארת את התופעה, נבנו שלושה מדדים במייצגים שלושה ממדים של התופעה: (1) היעדרויות מרובות ואיחורים, (2) תחושת שייכות לבית הספר, ו- (3) יחס התלמיד לבית הספר. על בסיס שיטה בה נעשה שימוש ביחס לנתונים דומים במחקר פיז"ה 2000 (Williims, 2003), חשובה עבור כל ממד נקודת חתך המבוססת על הצטברות של כמה היגדים בעייתיים. מדד הניתוק הוגדר כציון מעל החתך בשניים מתוך שלושת הממדים לפחות.

בהרצאה תוצג שיטת בניית המדד ויוצגו האתגרים שבבניית מדד לשימוש יישומי על סמך נתונים שלא נאספו לצורך זה מלכתחילה. בנוסף, יוצגו בקצרה הממצאים העיקריים.

קבלת החלטות עבר/נכשל, שאפשר להגן עליהן בעזרת השימוש ביכולת הנבחנים ובקושי הפריטים

בועז שולרוך, פיל ג'ונס

Office of Medical Education, University of New South Wales, Sydney, Australia

רולף טרנר

Department of Statistics, University of Auckland, Auckland, New Zealand

The determination of Pass/Fail decisions over Borderline grades, (i.e., grades which do not clearly distinguish between the competent and incompetent examinees) has been an ongoing challenge for academic institutions.. The objective borderline method (OBM), employing a probability approach to reclassify borderline grades, has been recently described. This study describes a modification of the OBM that uses two parameters (examinee ability and item difficulty) to determine whether t a Borderline grade should be a Fail or Pass.

Examinees' Borderline grades from a clinical examination were reclassified into Pass or Fail based on the probability, derived by the OBM2 method The examination outcomes (original and reclassified grades) were compared to each other and to subsequent clinical examination results.

The distribution of Pass/Fail decisions by the OBM2 method was significantly different to the original method ($p < .0001$). The predictive validity was studied using students' performance in a subsequent clinical examination. It was found that the OBM2 model was more stringent and yielded higher specificity and lower sensitivity than the original decisions.

The OBM2 is a simple method that uses a probabilistic approach to address decision making when dealing with Borderline grades in clinical examinations. The current study introduces a modification of the OBM encompassing both examinee ability and item difficulty which can be applied to categorical grades and can be adapted depending on the preferred level of stringency or leniency.

מודלים של כריית נתונים לחקירת סוללת המבחנים הממוחשבים של מת"ל

דביר קלפר, ענת בן-סימון, אפי לוי, נועה סקה
מרכז ארצי לבחינות ולהערכה (מאל"ו)

מת"ל (מערכת תפקודי למידה) היא מערכת ממוחשבת של מבחנים ושאלונים שפותחה לצורך אבחון מועמדים להשכלה הגבוהה ותלמידים הלומדים במסגרתה, המבקשים התאמות בבחינות וסיוע בלימודים בגין לקות למידה. המערכת כוללת שני שאלונים ו-20 מבחנים בתחומים הבאים: תפקודי שפה (עברית ואנגלית), תפקודים כמותיים, תפקודי קשב, זיכרון מילולי שמיעתי ותפיסה חזותית (Ben-Simon & Inbar-Wiess, 2012). המערכת נועדה לאבחן שלוש לקויות למידה: לקות קריאה (דיסלקציה), לקות כתיבה (דיסגרפיה), לקות בתפקוד כמותי (דיסקלקוליה) וכן הפרעת קשב וריכוז. בהרצאה זו יוצגו התוצאות שהתקבלו עבור דיסלקציה בלבד.

באבחון מבוסס מת"ל, האבחנה הקלינית של לקות נעשית על ידי מאבחנת מומחית ונשענת על נתונים ממגוון רחב של מקורות מידע: תוצאות מבחני מת"ל, תצפית בוחנת, שאלון אישי, מסמכי רקע וראיון אישי עם הפונה. אבחנה זו, המדווחת לפי שלוש רמות חומרה: (0) ללא לקות; (1) לקות קלה-בינונית; (2) לקות בינונית-חמורה, תשמש במחקר זה כקריטריון אותו ננסה לחזות. בהתאם לכך, נתייחס לקריטריון כמשתנה סדר (ordinal).

מטרתו של המחקר הנוכחי היא לבחון את דיוק הסיווג לרמות לקות המתקבל מיישום שלושה מודלים של סיווג (classification models) מעולם למידת המכונה (machine learning). המודלים שנבדקו הם:

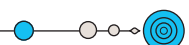
- מודל עץ החלטה טרינומי
- מודל רשת נוירונים
- מודל הרגסיה לוגיסטית מצטברת

המדגם כלל 562 משתתפים שפנו לאבחון מבוסס מת"ל ושעבורם קיימות תוצאות מלאות של 47 מדדי הביצוע. מדדים אלו קובצו ל-10 גורמים על פי תוצאות ניתוח גורמים שבוצע על מדגם של 2,312 נבחני מת"ל (בן-סימון, קלפר וסקה, בכתביה). עשרת הגורמים הם: שליפה לקסיקלית; פיענוח פונולוגי; קשב; מוסחות; עיבוד כמותי בסיסי; קשב; תסמינים התנהגותיים; קצב עבודה; תפקודים גבוהים; הבנת טקסט; דיוק; עיבוד אורתוגרפי; אימפולסיביות; ושטף מילולי.

כמקובל בסוג כזה של מחקר, בסיס הנתונים חולק לשלושה מדגמים באופן אקראי: 50% לאימון, 25% לבחירת המודל ו-25% לבדיקת המודל. חלוקה זו בוצעה ארבע פעמים וזאת על מנת לקבל תוצאות יציבות עבור מטריצת האמת (מטריצה שמתארת את מידת ההתאמה בין הניבוי לבין המצב האמיתי).

ממודל הרגסיה ובעיקר ממודל העץ, עולה שהגורמים שמשפיעים על האבחנה הקלינית של דיסלקציה הם: עיבוד אורתוגרפי (המפריד המרכזי) ולאחריו פיענוח פונולוגי (מפריד משני). עוד עולה מן המחקר שמבין שלושת המודלים שנבדקו, המודל הטוב ביותר לחיזוי דיסלקציה הינו מודל הרגסיה, מודל המצליח לנבא נכונה כ-75% מהמקרים.

רב המחקרים בתחום המדידה ובפרט בתחום הדיאגנוסטיקה משתמשים במודלים מעולם הסטטיסטיקה הקלאסית (רגרסיות/ניתוחי שונות וכו'). מחקר זה, הבוחן שלושה מודלים מקובלים מעולם למידת המכונה, הינו חדשני ומאפשר לקבל תובנות פנימיות נוספות על תוקף המבנה של התפקודים העומדים בבסיס כל אחת מהלקויות השונות. בנוסף, המחקר הנוכחי משווה בין טיב הניבוי של שלושה מודלים שונים, והמודל הטוב ביותר עבור כל אחת מן הלקויות יכול להחליף את המודל הנוכחי עליו מבוסס האלגוריתם הממוחשב במערכת מת"ל.



השוואה בין שתי שיטות לחישוב טעויות תקן ורווחי סמך למתאמים שתוקנו לקיצוץ תחום עקיף

תמר קנת-כהן, דביר קלפר, אליוט טורוול
מרכז ארצי לבחינות ולהערכה (מאל"ו)

במחקרים להערכת יעילות כלי מיון מחשבים את מקדם המתאם בין חזאי (X) לקריטריון (Y). מחקרים אלה מתבצעים לעתים קרובות במדגמים של אנשים שהתקבלו על בסיס משתנה שלישי (Z). במידה שבה X ו-Y מתואמים עם Z, אזי במדגם המתקבלים יש צמצום בטווח הערכים שלהם (קיצוץ תחום עקיף) והמתאם המחושב מהווה הערכת חסר למתאם ביניהם באוכלוסיית המועמדים. נוסחה ידועה לתיקון המתאם עקב קיצוץ התחום העקיף (Thorndike, 1949) מפיקה אומד נקודתי ("מתאם מתוקן לקיצוץ תחום") למתאם באוכלוסייה.

מעבר לאומד הנקודתי, חשוב לספק מדדים על מידת אי-הוודאות באמידה, והמדדים המומלצים בספרות המקצועית הם טעות התקן של האומד ורווח סמך לפרמטר (AERA, APA, & NCME, 2014). אלא שבשל מורכבות נוסחת התיקון לקיצוץ תחום, לא קיימת נוסחה סגורה לחישוב טעות התקן של המתאם המתוקן, ובנוסף, מאחר שהתפלגות הדגימה של המתאם המתוקן אינה ידועה, לא ניתן לחשב רווח סמך למתאם באוכלוסייה. לאחרונה הוצעה שיטה של Bootstrap לחישוב טעות התקן ורווח הסמך (Li, Chan, & Cui, 2011). שיטה זו נבדקה בהדמיה של התפלגות נורמלית ונמצא שהיא מתפקדת באופן משביע רצון.

מאחר ושיטת ה-Bootstrap צורכת משאבי מחשב רבים, יש יתרון לפתרון אנליטי לחישוב טעות התקן ורווח הסמך. במחקר זה פותח ונבדק פתרון כזה, ברוח פתרון שהוצג בעבר עבור המקרה של קיצוץ תחום ישיר (Chan & Chan, 2004). השיטה האנליטית הושוותה לשיטת ה-Bootstrap באמצעות הדמיה עם נתונים מהתפלגויות שונות של משתנה הבררה, Z, באוכלוסייה (נורמלית, אחידה, אסימטרית חיובית ואסימטרית שלילית), ועם רמות שונות של גודל מדגם המתקבלים (בין 20 ל-100), של יחס הבררה (בין 0.10 ל-0.50) ושל המתאמים באוכלוסייה בין שלושת המשתנים - X, Y ו-Z (בין 0.20 ל-0.80).

ממצאי המחקר הראו שמידת הדיוק של האומד הנקודתי תלויה בהתפלגות של משתנה הבררה: המתאמים המתוקנים היו מדויקים למדי כאשר Z התפלג נורמלית או אסימטרית חיובית ומוטים בשתי ההתפלגויות האחרות, ובעיקר כש-Z התפלג אסימטרית שלילית. ברוב המקרים כיוון ההטיה היה שלילי: המתאמים המתוקנים סיפקו אומדן חסר של המתאם באוכלוסייה.

אשר לשני המדדים של מידת אי-הדיוק באמידה - טעות התקן של המתאם המתוקן ורווח הסמך למתאם באוכלוסייה - שיטת ה-Bootstrap הפגינה עדיפות על פני השיטה האנליטית: בשיטה זו מידת הדיוק של אומדן טעות התקן הייתה רבה יותר והסתברות הכיסוי של רווח הסמך הייתה קרובה יותר לערך הנקוב (95%). עם זאת, תחת מרבית סוגי ההתפלגויות (להוציא את זו שבה Z התפלג אסימטרית שלילית) נמצאו צירופי תנאים - למשל, כאשר Z התפלג נורמלית ומדגם המתקבלים היה גדול - שבהם תפקוד שתי השיטות היה דומה.

המסקנה המעשית שעולה מן הממצאים היא שבנסיבות שבהן סוג ההתפלגות של משתנה הבררה ותנאים רלוונטיים אחרים אינם ידועים, שיטת ה-Bootstrap לאמידת טעות התקן של מתאם מתוקן לבררה עקיפה ולחישוב רווח סמך למתאם באוכלוסייה היא עדיפה. עם זאת, בהינתן ששיטת ה-Bootstrap נזקקת למשאבי מחשב רבים, אזי כאשר ידוע שמתקיימים התנאים שבהם תפקודן של שתי השיטות הוא דומה, עשוי להימצא יתרון לשימוש בשיטה האנליטית.

האם אנחנו מוכנים לחיות ביחד? המשגה ומדידה של קיום משותף בין יהודים וערבים בישראל

מאור שני, קלאוס בואנקה, מרגריט שרייר

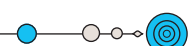
בית-הספר הבינלאומי ללימודים גבוהים במדעי החברה בברמן, גרמניה

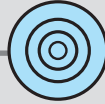
המונח קיום משותף (Coexistence) נמצא בשימוש תדיר בשיח הפוליטי והחברתי בישראל בהתייחס ליחסים בין אזרחים יהודים וערבים. בספרות הפסיכולוגית-חברתית, מונח זה מתייחס לשלב ראשוני של יחסים חיוביים בין קבוצות לאחר סכסוך מתמשך, שאף נתפס כשלב הכרחי בדרך ליחסים חיוביים יותר כגון פיוס, שלום חיובי והרמוניה. למרות העניין התיאורטי הרב במושג זה בשנים האחרונות, עד כה לא נעשה ניסיון להגדירו באופן שיטתי ולפתח כלי למדידתו. מחקר זה מנסה לגשר על פער זה על-ידי בניית כלי למדידה ובחינה אמפירית של מידת האוריינטציה לקיום משותף בקרב יהודים וערבים בישראל. אוריינטציה לקיום משותף נתפסת כמורכבת מחמישה תחומים, וכל תחום ניתן לבחינה באמצעות מספר אינדיקטורים: רגשות כגון אמפתיה ותקווה עשויים להצביע על מידת האוריינטציה בתחום הרגשי; תפיסות האיום הקולקטיבי ודימויים שליליים עשויים להצביע על מידת האוריינטציה הקוגניטיבית; עניין באחר ובתרבותו ופתיחות לפרספקטיבה של האחר עשויים להצביע על מידת האוריינטציה המוטיבציונית; מוכנות למגע עם האחר עשויה להצביע על מידת האוריינטציה בתחום ההתנהגותי; ולבסוף, תמיכה בשוויון זכויות וסובלנות פוליטית עשויות להראות את מידת האוריינטציה הפוליטית לקיום משותף בקרב יהודים.

עקב אפיונו של המחקר כמחקר גישוש ותיקוף ראשוני, יושמה גישה מרוכזת-אדם (person-centered approach), במסגרתה זוהו אשכולות של משתתפים המתאפיינים בדפוסים דומים לאורך תחומי האוריינטציה. הנתונים נותחו באמצעות ניתוח פרופילים חבויים (Latent Profile Analysis, LPA), טכניקה אינדוקטיבית מבוססת-מודל לסיווג מקרים לתת-קבוצות שאינן גלויות, המאפשרת לזהות את מספר האשכולות האופטימלי על-פי טיב ההתאמה. LPA בוצע לכל קבוצה לאומית בנפרד, ונבחנה מידת ההתאמה של 1 עד 10 פרופילים על בסיס חמשת (ליהודים) או ארבעת (לערבים) משתני התחומים. לאחר בחירת הפתרון האופטימלי וסיווג המשתתפים לאשכולות, בוצע תיקוף חיצוני לפרופילים על-ידי בחינת הקשר בינם לבין משתנים סוציו-דמוגרפיים, אישיותיים וחברתיים-פוליטיים.

עבור המשתתפים היהודים, הפתרון האופטימלי כלל חמישה פרופילים (BIC = -1910.97, misclassification error = 8.99%): בעלי אוריינטציה בינונית (36.2%), בעלי אוריינטציה נייטרלית (26.0%), בעלי אוריינטציה חזקה (19.4%), חסרי אוריינטציה (11.7%), ובעלי אוריינטציה חזקה מאוד (6.6%). עבור המשתתפים הערבים, נמצא כי קיימים ארבעה אשכולות חבויים (BIC = -1038.87, misclassification error = 14.46%): בעלי אוריינטציה חזקה (35.5%), בעלי אוריינטציה בינונית (29.9%), בעלי אוריינטציה חלקית (19.8%), וחסרי אוריינטציה (14.8%). עוד נמצא כי סיווג המשתתפים הפלסטינים לארבעת הפרופילים כמעט ולא הושפע מהאוריינטציה הקוגניטיבית, אשר היתה נמוכה למדי בשלושה פרופילים (ממוצע בין 0.30 ל-0.33 בסולם של 0 עד 1). מבחני התיקוף החיצוני חשפו הבדלים מובהקים בין הפרופילים בכל קבוצה במאפייני רקע (מגדר, דתיות), משתני אישיות (לקיחת הפרספקטיבה של האחר, אינטרס עצמי היררכי), משתנים פוליטיים (הזדהות תוך-קבוצתית, אוריינטציה פוליטית ולאומנית), ואף משתנים התנהגותיים (מגע בין-קבוצתי).

תיקוף הפרופילים באמצעות משתנים חיצוניים הראה את יעילותו של האינדקס בחשיפת המידה בה תתי-קבוצות בכל קבוצה לאומית מוכנים לקיום משותף עם בני הקבוצה האחרת. לדעתנו, כלי זה יכול לשמש למעקב לאורך זמן אחר דפוסי האוריינטציה לקיום משותף בישראל, וכן להערכת האפקטיביות של תכניות התערבות. בנוסף, הממצא לפיו ריבוי תפיסות איום קולקטיבי וסטריאוטיפים כלפי יהודים משותף מאפיינים את כל הפרופילים בקרב פלסטינים מחייב התייחסות חינוכית וחברתית.





ידיעון אפי לובש צורה חדשה:

ה"בלוגמטרי"

באתר אפי www.ispa.org.il

הבלוגמטרי הינו פלטפורמה נגישה להחלפת מידע ודעה בתחום המדידה וההערכה.

בבלוג מתפרסמים עדכונים, חידושים, דעות ומידע חשוב לעוסקים בתחום ולמתעניינים בנעשה בתחום הפסיכומטריקה בישראל.

אתם מוזמנים להתארח בבלוג, להגיב, ולשתף במחקרים, פרויקטים, מבחנים חדשים, כנסים מעניינים ועוד.

אנא העבירו כל חומר רלבנטי **למנחת הבלוג, ד"ר ליאת בסיס** liat@basisle.co.il

תאובות ופצות
בנושא האגודה הישראלית לפסיכומטריקה
אפילו הפסנות אל:

שרה שפירא, האגודה הישראלית לפסיכומטריקה ispaisrael@gmail.com

אתר אפי www.ispa.org.il