

הכינוס העטירי טל האגודה היטראלית לפסיכומטריקה

2014

כ"ב בשבט תשע"ד 23.1.2014
וואהל סנטר, רמת גן

בתמיכת:

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

<p>התכנסות וארוחת בוקר קלה דברי פתיחה יו"ר אפי: ענת בן-סימון, יו"רי הכינוס: ליאת בסיס וחנן גולדשמידט</p>	<p>8:45 – 9:15 9:15 – 9:30</p>
<p>מושב א' יו"ר: יואב כהן</p> <ul style="list-style-type: none"> ■ על הקשר בין חודש הלידה ומחוננות בישראל אלעד שגב, סורל קאהן ■ על מבנים וזהויות – ניתוחים מבניים בשירות הערכת תכניות: המקרה של תכניות "מסע" ענת קדם, דניאל בן אליעזר, יעל אבירם, טלי פרוינד ■ פיתוח מדדי בקרת איכות לציוני מבחנים המועברים תכופות מיכל באומר, אבי אללוף, טוני גוטנטג, מרינה פרונטון ■ הבדלים בין המינים בהישגים בבית הספר: פרספקטיבה תוך-כיתתית מאיר ברנרן, סורל קאהן, סוהאד זועבי 	<p>11:00 – 9:30</p>
<p>הפסקה</p>	<p>11:00 – 11:30</p>
<p>מושב ב' 1 (מושב מקביל) יו"ר: אלי וקיל</p> <ul style="list-style-type: none"> ■ מדידת זיוף תוך נבדקית באמצעות זמני תגובה מירב פירק, סול פיין ■ התפתחות כלי המיון במערכת המיון לצוות אויר: ממבחני כשרים בודדים לסימולטורי הטסה ענבל רויטר ■ סקירת מערכות מיון להשכלה גבוהה ברחבי העולם איילת גולדצויג ■ על הקשר בין מבנה הבחינה המסכמת בקורסים אקדמיים לבין סגנון הלמידה בקורס נעמה סיימונס, ענת בן-סימון 	<p>11:30 – 13:00</p>
<p>מושב ב' 2 (מושב מקביל) יו"ר: פאדיה נאסר</p> <ul style="list-style-type: none"> ■ ברירה ללימודים גבוהים לפי בגרות בלבד: אמידת השינוי בסף הבגרות הנדרש לקבלה למגוון חוגים מתוך נתוני הלומדים יונתן סער ■ המחשת התוקף של שלושה ציוני מכפ"ל כלליים בניבוי ציוני שנה א' באוניברסיטאות כרמל אורן, תמר קנת-כהן, אליוט טורוול, אבי אללוף ■ האם ההכנה למבחני הקבלה לרפואה מסייעת למועמדים? אביטל מושינסקי, דוד ציגלר ■ ניתוח מבנה הבחינה הפסיכומטרית החדשה – ניתוח גורמים מגשש ומאשש בשלושה חלקים דביר קלפר 	<p>11:30 – 13:00</p>
<p>ארוחת צהריים</p>	<p>13:00 – 14:15</p>

מושב ג' יו"ר: חנן גולדשמידט וליאת בסיס	15:00 – 14:15
■ פאנל אלטרנטיבי	14:35 – 14:15
■ חלוקת פרסי אפי לעבודות הזכות:	15:00 – 14:35
עדי לוי-ורד האומנם מוכנים מורי העתיד לאתגר של הערכת לומדים? לירון צוברי לתקן או לא לתקן? על היתרונות שבשימוש ציוני תקן של פריטי שאלונים ומבחנים בחישוב ציון כולל	
■ האומנם מוכנים מורי העתיד לאתגר של הערכת לומדים?	
עדי לוי-ורד	
מושב ד' 1 (מושב מקביל) סימפוזיון יו"ר: נעמי גפני	16:00 – 15:00
■ סיכום של שנה לרפורמה בבחינה הפסיכומטרית – הוספת מטלת	
כתיבה ועוד...	
נעמי גפני	
■ הכנות לקראת יישום הרפורמה בבחינה הפסיכומטרית	
נעמי גפני	
■ סוגיות בפיתוח מבחן כתיבה: פיתוח הדגם המתאים למטלת הכתיבה,	
שיקולים בבחירת סוג המחווה, מיון והכשרת המעריכים	
גלעד סופר, ספי פומפיאן	
■ הכללת מטלת כתיבה בבחינה הפסיכומטרית – היערכות לוגיסטית ונתונים	
סטטיסטיים	
אבי אללוף	
מושב ד' 2 (מושב מקביל) סימפוזיון יו"ר: יואל רפ	16:00 – 15:00
■ פיתוח כלי להערכת תפקוד של תלמידים בעלי צרכים מיוחדים	
במערכת החינוך בישראל – היבטים פסיכומטריים ותיקוף הכלי	
יואל רפ	
■ הליך פיתוח הכלי להערכת תפקוד של תלמידים בעלי צרכים מיוחדים	
לאורה קורטן	
■ מחקר התיקוף של הכלי להערכת תפקוד של תלמידים עם צרכים מיוחדים	
בכיתות א'-ו'	
נוגה גוטספלד	
■ כלי להערכת תפקוד של תלמידים עם צרכים מיוחדים בחטיבת הביניים	
אסטלה מלמד	
הפסקה	16:20 – 16:00
מושב ה' יו"ר: יצחק פרידמן	17:30 – 16:20
■ טעות מדידה בשיפוטי חיבורים: בדיקה אמפירית של "כשל המעריך השלישי"	
יואב כהן	
■ הטיה אופטימית בשיפוט מתאם פירסון	
אייל גמליאל, רואי גור	
■ יישום גישות שונות לחישוב רווח סמך ולאמידת מתאם מתוקן לקיצוץ תחום	
אילה כהן, אטי דובא, טטיאנה אומנסקי, תמר קנת-כהן	
אסיפת חברי אפי	18:00 – 17:30

על הקשר בין חודש הלידה ומחוננות בישראל

סורל קאהן

האוניברסיטה העברית

אלעד שגב

המכון הטכנולוגי חולון

רקע

בישראל קיימות כ-50 תכניות אזוריות לטיפול תלמידים מחוננים. בכל אחת מהתכניות, בחירת המועמדים מתבצעת בכיתה ב', בשני שלבים. בשלב הראשון, נבחרים בכל בית ספר 15% התלמידים בעלי הציונים הגבוהים ביותר במבחן הישגים אחיד. תלמידים אלה עוברים לשלב השני, שבו מועבר להם מבחן כושר אחיד, ובכל אזור נבחרים לתכנית כ-10%-20% בעלי הציונים הגבוהים ביותר במבחן (כלומר, 1.5% - 3% מכלל התלמידים באזור). בהתאם למספר המקומות בתכנית.

הבעיה

בשני השלבים הבחירה מתבססת על ציוני הגלם של התלמידים ומתעלמת מאפקט הגיל הכרונולוגי על ציונים אלה בתוך אותה דרגת כיתה. כתוצאה מכך, בחירת המועמדים לתכניות טיפוח מחוננים מוטה לטובת הילדים המבוגרים בכיתה ונגד הילדים הצעירים.

המחקר

מטרת המחקר הנוכחי הייתה לאמוד את גודל האפקט של הגיל הכרונולוגי על ציוני הגלם במבחנים המשמשים לאבחון מחוננים בישראל ואת גודל ההטיה הנובעת מהתעלמות מאפקט זה בתהליך הבחירה של מועמדים לתכניות טיפוח מחוננים. בהסתמך על אוכלוסיית כל תלמידי כיתה ב' שנולדו ב-2002, המחקר בדק את הקשר הקיים בישראל בין גיל כרונולוגי (חודש לידה) לבין ההסתברות להיבחר לתכנית המחוננים. התוצאות מצביעות על קשר כמעט מלא ($r^2=0.92$) בין השניים. הסתברות הבחירה לתכנית מחוננים היא פונקציה ליניארית חיובית של הגיל הכרונולוגי: הבדל של חודש אחד בגיל הכרונולוגי מעלה ב-11% בממוצע את סיכויי ההבחרות לתכנית. כתוצאה מכך, הסיכוי של הילדים המבוגרים ביותר בכיתה (ילידי ינואר) להיות מאובחנים כ"מחוננים" ולהיבחר לתכנית גדול פי 3.5 מזה של הילדים הצעירים ביותר (ילידי דצמבר של אותה שנה).

מסקנה

ביסוס הבחירה לתכניות לטיפול מחוננים בישראל על ציוני גלם במבחני הישגים ובמבחני כושר, תוך התעלמות מהאפקט החיובי הניכר של הגיל הכרונולוגי על ציונים אלה בתוך אותה דרגת כיתה, גורם להטיה בבחירה - אבחון יתר בקרב התלמידים המבוגרים ואבחון חסר בקרב הצעירים - ופוגע לכן בתוקף תהליך הבחירה ובהוגנותו.

על מבנים וזהויות - ניתוחים מבניים בשירות הערכת תכניות: המקרה של תכניות "מסע"

ענת קדם, דניאל בן אליעזר, יעל אבירם, טלי פרוינד

היחידה להערכת תכניות - המרכז לטכנולוגיה חינוכית

פרויקט "מסע" הוא יוזמה משותפת של ממשלת ישראל והסוכנות היהודית לעידוד צעירים יהודים מהתפוצות להשתתף במגוון תכניות ממושכות בישראל (6-12 חודשים). "מסע" מתאם ומפקח על כ-150 גופים מארגנים המפעילים כ-200 תכניות שבהן משתתפים כ-10,000 צעירים מדי שנה. תיאורית הפעולה של "מסע" מבוססת על ההנחה כי שהות ממושכת בישראל, המלווה בפעילות חינוכית מובנית, תניב את התוצאות האסטרטגיות האלה: העמקת הקשר והמחויבות לישראל, חיזוק הזיקה ליהדות, והגברת הנכונות למעורבות בקהילה היהודית.

החל משנת תש"ע (2010), מלווה מטח את הפרויקט בפעולת הערכה מקיפה. במסגרת זו פיתח צוות ההערכה של מטח, בסיוע מומחי תוכן, שאלון שנועד לבחון את המידה בה משיג "מסע" את התוצאות האסטרטגיות המצופות, במונחים של שינויים בעמדות ובכוונות ההתנהגות של המשתתפים ביחס לישראל, ליהדות ולקהילה היהודית. השאלון הועבר למדגם נרחב של משתתפים (N=2,438) כסקר בתחילת התכנית ובסיומה בשנים תשע"א-תשע"ב.

ניתוח גורמים שבוצע על ממצאי הסקר הציע ארבעה גורמים, שהומשגו כ: "קשר רגשי לישראל", "עניין בתרבות ובחברה הישראלית", "זיקה למסורת ולערכים יהודיים" ו"כוונה למעורבות ולהנהגה בקהילה היהודית". גורמים אלה הסבירו 58% מהשונות ושימשו כמשתנים לפיהם נבדקה השפעת ההשתתפות בתכניות "מסע". נמצאה עלייה מובהקת ($p < .001$) בכל ארבעת הגורמים בין תחילת התכנית לסיומה (באמצעות מבחן t למדגמים מזווגים), כאשר עלייה חזקה יחסית נרשמה ב"קשר רגשי לישראל" ו"כוונה למעורבות ולהנהגה בקהילה היהודית" (Cohen's $d = 0.31, 0.32$).

לעיון נוסף במבנה השאלון והמושגים שהוא מייצג בוצע ניתוח MDS. ניתוח ה-MDS מפקח מעין מפה המאורגנת לפי שני צירים ועליה מפוזרים הפרטים, כאשר המרחק בין הפרטים מבטא את עוצמת הקשר ביניהם. ממצאי הניתוח איששו את המבנה שנמצא בניתוח הגורמים בכך שפרטים ששויכו לאותו גורם, הצטופפו במידה רבה זה לצד זה, על פני מערכת הצירים. לצד אישוש זה הציע הניתוח מבט נוסף על המבנים הנמדדים באמצעות השאלון, זאת באמצעות המשגת הצירים כקטבים המארגנים את זהות המשתתף. ציר אחד נע בין "תרבות ישראלית" ל"תרבות אוניברסלית", והשני בין "הוויה ישראלית ל"הוויה בתפוצות". השוואה ויזואלית בין ממצאי ניתוח MDS שנערך על נתוני סקר תחילת התכנית לבין ממצאים של הסקר בסיום התכנית הראתה כי חלה התקרבות בין הפרטים של שלושה גורמים: "זיקה למסורת ולערכים יהודיים", "קשר רגשי לישראל" ו"כוונה למעורבות ולהנהגה בקהילה". ממצא זה רימז לאפשרות כי לצד העלייה בעוצמת הקשר הרגשי לישראל והכוונה למעורבות בקהילה, ישנו אפקט נוסף להשתתפות בתכניות "מסע" והוא אפקט של התגבשות בין שלושה מרכיבים בזהות המשתתף: ישראל, יהדות וקהילה יהודית. אישוש אמפירי להשערה זו התקבל לאור ממצאים המלמדים על עלייה מובהקת ($p < .001$) בעוצמת המתאמים המתקבלים בין שלושת הגורמים בסוף התכנית בהשוואה לאלה המתקבלים בתחילתה (מבחן Pearson-Filon).

למחקר ההערכה הנוכחי, אשר ניסה להתחקות אחר מושג מופשט ומורכב של זהות יהודית, זימן ה-MDS תובנות חדשות ומרתקות הן באשר לממדים שמארגנים אותו והן באשר לתהליכי השינוי שהוא עובר בעקבות התערבות חינוכית מובנית. חשיבותו של המחקר הנוכחי היא בכך שהוא מדגים את התועלת של שימוש מושכל בניתוח ה-MDS בהקשר של הערכת תכניות, ומזמן הרחבת היישום שלו במסגרת מחקרים דומים.

פיתוח מדדי בקרת איכות לציוני מבחנים המועברים תכופות

מיכל באומר, אבי אללוף, טוני גוטנטג, מרינה פרונטון

מרכז ארצי לבחינות ולהערכה

בניית מבחנים, העברתם, ניתוח תוצאותיהם וחישוב ודיווח ציונים, הם תהליכים מורכבים רב-שלביים – כאשר בכל שלב עלולה להתרחש טעות. מכיוון שבמבחנים עתירי סיכון (High-stakes testing) מחיר טעויות הוא גבוה, יש צורך בבקרת איכות קפדנית על תהליכי המדידה בכלל ועל הציונים בפרט. ניתן לחלק הליכי בקרת איכות על ציונים לשני סוגים עיקריים: בקרת איכות על בחינות המועברות לקבוצות אוכלוסיה גדולות בכמה מועדים מוגדרים תוך שימוש במספר נוסחים קטן, ובקרת איכות על בחינות המועברות לקבוצות אוכלוסיה קטנות במועדים רבים ותכופים תוך שימוש במספר רב של נוסחי בחינה. מחקר זה עוסק בסוג השני, שהוא מאתגר וחדשני, שכן עולם הבחינות משתנה במהירות לכיוון זה.

מטרת המחקר הנוכחי הינה ביצוע בקרת איכות על מבחנים המועברים תכופות באמצעות שיטות סטטיסטיות שבהן נעשה שימוש בשנים האחרונות תוך פיתוח שיטות בקרת איכות חדשות המתאימות הן למבחן הנבדק והן למבחנים נוספים.

מחקר זה השתמש בנתונים מ-23,548 נבחנים בבחינת מימ"ד במהלך ארבע שנים (2008 – 2012). בחינת מימ"ד היא בחינה ממוחשבת המיועדת למועמדים למכינות הקדם אקדמיות במכללות. הבחינה מודדת ידע בשלושה תחומים: עברית, מתמטיקה ואנגלית.

בדיקות האיכות התבצעו בשלבים השונים של הבחינה. נבדקו היבטים הקשורים בהעברת הבחינה ונמצא כי דגימת הנוסחים במשקולות שנקבעו מראש עבדה כמתוכנן וכן שסדר הפרקים בבחינה היה רנדומאלי כמתבקש. היבטים הקשורים בתוכן והמאפיינים הפסיכומטריים של הבחינה כגון רמת הקושי של הפריטים נבדקו אף הם ונמצאו יציבים לאורך זמן, כמה פריטים שנמצאו פחות יציבים נבדקו, אך לא נמצאה סיבה עקבית לכך. באשר לשקילות הנוסחים, נמצא הבדל מובהק בציונים בנוסחים השונים, אך גודל האפקט היה קטן. מבחני המשך איששו החלטה קודמת להוריד נוסח אחד ממחזור ההעברות מאחר שהניב ציונים נמוכים יותר באופן עקבי. כמו כן, בעקבות בדיקה זו אותר נוסח בחינה נוסף, שבו נדרש שינוי באחד מתחומי הבחינה. ניטור יציבות הציונים לאורך זמן נעשה באמצעות תרשימי בקרה מסוג Shewhart המאפשרים מעקב אחר ציוני הבחינה לאורך זמן ואיתור תקלות במהירות. כדי לנטרל תנודתיות מחזורית שנצפתה בתרשימי הבקרה הנובעת מעונתיות מסוימת בציוני בחינת מימ"ד, הבדלים בין-אישיים וגורמים נוספים, נערכה גרסיה רב משתנית לניבוי ציון הבחינה הכללי ושורטטו תרשימי בקרה עבור השאריות מהגרסיה (השונות הבלתי-מוסברת). השימוש בתרשימים מסוג זה לבקרה על ציוני מבחנים הוא החידוש המרכזי במחקר הנוכחי ומציאת הגורמים המשפיעים על השונות האופיינית בציונים לאורך השנה ונטרולם אפשרה שיפור של הבקרה באמצעות תרשימי Shewhart, צמצום התראות השווא, ואיתור תקלות יעיל יותר.

למחקר זה השלכות תיאורטיות ומעשיות. בפן התיאורטי, במחקר זה פותחו ויושמו שיטות בקרת איכות למבחנים המועברים תכופות לקבוצות נבחנים קטנות. במובן המעשי, שקורם עור וגידים בימים אלו במרכז הארצי לבחינות ולהערכה, יישום של טכניקות בקרת איכות באופן שגרתי ומתמשך יסייע בזיהוי מהיר ומדויק של תקלות בבחינות באופנות העברה מתמשכת.

הבדלים בין המינים בהישגים בבית הספר: פרספקטיבה תוך-כיתתית

סורל קאהן, סוהאד זועבי

בית הספר לחינוך, האוניברסיטה העברית

מאיר ברנרן

המחלקה לפסיכולוגיה, האוניברסיטה העברית

המחקרים על הבדלים בין המינים בהישגים התמקדו בפער בין המינים ברמת האוכלוסייה והתעלמו מהסביבה שבה מתרחש תהליך הלמידה - הכיתה - ומהשונות האפשרית בין כיתות בכיוון הפער ובגודלו. המחקר הנוכחי הוא הניסיון הראשון לאמוד את ההבדלים מבחינה זאת בין כיתות. המחקר מבוסס על תוצאות מבחני המיצ"ב במתמטיקה, מדעים, שפה אם ואנגלית, שהועברו ל-1430 כיתות ה' מעורבות (662 בתי ספר ו-37,861 ילדים) בשנת 2006. עבור כל כיתה ומקצוע, חושב ההפרש המתוקן (d של Cohen) בין ממוצע הבנים לממוצע הבנות,

התקן המשותפת לבנים ולבנות בתוך הכיתה).

$$d_j = \frac{M_j^b - M_j^g}{S_{wj}}$$
 כאשר M_j^b הינו ממוצע הבנים בכיתה j, M_j^g הינו ממוצע הבנות בכיתה j, ו- S_{wj} הינה סטיית

התוצאות העיקריות הן:

- יציבות גבוהה של ההפרש המתוקן התוך-כיתתי בין המינים מעבר למקצועות.
- הבדלים ניכרים בין הכיתות בסימן ההפרש המתוקן התוך-כיתתי ובגודלו: ילדים הלומדים בכיתות שונות נחשפים להבדלים שונים בין המינים, מבחינת סימנם וגודלם.
- רוב ההבדלים בין הכיתות בסימן ההבדל התוך-כיתתי בין המינים ובגודלו הם הבדלים בתוך בתי הספר ולא בין בתי הספר.

תוצאות אלו מצביעות על חשיבות הפרספקטיבה התוך-כיתתית בחקר ההבדלים בין המינים בהישגים לימודיים במערכות חינוך, ועל הצורך בהבחנה מושגית ואמפירית בין ההבדלים התוך-כיתתיים להבדל בין הממוצעים באוכלוסייה כולה. כפי שמדגים מחקר זה, ההבדל בין הממוצעים של שני המינים בכלל האוכלוסייה המדווח בספרות, מסתיר מציאות תוך-כיתתית מורכבת ומגוונת. לכן, ההיסק מתוך הרמה האגרגטיבית לרמה התוך-כיתתית הינו בבחינת טעות מושגית ואמפירית חמורה ביותר (the ecological fallacy). הפרספקטיבה התוך-כיתתית בחקר ההבדלים בין מינים (כמו גם בין קבוצות המוגדרות על ידי מאפיינים אחרים) בקרב תלמידים הינה קריטית להבנת התופעה.

מדידת זיוף תוך נבדקית באמצעות זמני תגובה

סול פיין
חברת מידות

מירב פירק
אוניברסיטת חיפה

זיוף במבחנים לא קוגניטיביים בדיווח עצמי מהווה בעיה מרכזית במיון תעסוקתי. שיעור לא מבוטל של מועמדים מזייפים במבחני מיון (Griffith, Chmielowski, & Yoshita., 2007), דבר המשפיע ישירות על דירוג המועמדים והחלטות המיון. בשל הקושי לזהות זיוף באופן מדויק (Saxe, 1994), הוצעו שיטות חדשות למדידת זיוף באמצעות זמני תגובה ארוכים מהצפוי. אולם, שיטות אלו נבחנו בעיקר ברמת הקבוצה, בין מבחנים חוזרים, וללא יכולת להערכה מהימנה של זיוף ברמת הפרט או למדידת מידת הזיוף באופן רציף.

לאחרונה, הציעו Fine & Komar (2011) שיטה חדשה למדידת זיוף באמצעות זמני תגובה, שנועדה לספק מענה למגבלות אלו בתוך אותו מבחן, אולם היא טרם נבדקה אמפירית. השיטה, Difference Score Index (DSI), מתבססת על יצירת מדד base-line של זמני תגובה, המורכב מסולם פריטי "ביקורת", ומאפיין תגובות כנות בלבד. לאחר מכן מושווה מדד זה לזמני התגובה בפריטים בסולם "המטרה" המודדים את התכונה הרלוונטית שבה עולה חשש לזיוף. מאחר שהשיטה החדשה מביאה בחשבון הבדלים אינדיבידואליים טבעיים של זמני תגובה, שאינם בהכרח מקושרים לזיוף, שוער כי שיעור הדיוק במדידת זיוף באמצעות השיטה החדשה יהיה גבוה לעומת שיטות נורמטיביות המודדות זמני תגובה ביחס לנורמות הקבוצה.

המחקר הנוכחי בחן את יעילות שיטות אלו (נורמטיבית, DSI) במערך ניסויי, שבו 104 נבדקים נבחנו בתנאי זיוף ו-102 נבחנו בתנאי כנות. ההשערה נבחנה בשני תנאים אלו ובקרב שני סוגים של פריטים: פריטי שאלונים "גלויים" הנחשבים חשופים יותר לזיוף, ופריטי שאלונים "סמויים" הנחשבים פחות חשופים לזיוף.

תוצאות המחקר, תומכות בטענה כי נבדקים מסוגלים להעלות את ציוניהם בשאלוני אישיות, כאשר הם מתבקשים לעשות כן. במחקר הנוכחי נבדקים העלו את ציוניהם בכסטיית תקן, הן בפריטים גלויים והן בפריטים סמויים. כלומר שהמידה שבה הצליחו הנבדקים להעלות ציוניהם (לזיוף) נמצאה דומה בין שני סוגי הפריטים. בנוסף, בשני סוגי הפריטים התקבלו זמני תגובה ארוכים יותר באופן מובהק בעת זיוף לעומת מענה כנה. בכך, תומכות התוצאות בטענה הרווחת כי למענה שקרי נדרש יותר זמן לעומת מענה כנה, ככל הנראה כתוצאה מכך שהראשון מערב תהליכים קוגניטיביים מורכבים יותר היוצרים עומס קוגניטיבי רב יותר מאשר השני. בנוסף, זמני התגובה בפריטים הסמויים לעומת גלויים נמצאו ארוכים יותר.

ממצא עיקרי במחקר הנוכחי הוא כי שיטת ה-DSI הניבה אחוזי דיוק גבוהים יותר (68.5% בפריטים גלויים ו-65.5% בפריטים סמויים) לעומת שיטות נורמטיביות למדידת זיוף באמצעות זמני תגובה, המודדות זמני תגובה בהשוואה לממוצע קבוצתי (56.6% בפריטים גלויים ו-55.8% בפריטים סמויים). כמו כן, בקרב שני סוגי הפריטים, מתוך כלל הנבדקים שאובחנו נכונה כמזייפים באמצעות שיטת ה-DSI, כ-60% בלבד אובחנו גם על ידי השיטה הנורמטיבית. כלומר, ה-DSI מאפשרת לזהות סוג ייחודי של מזייפים שלא היה מתגלה בשיטה הנורמטיבית. לבסוף, שיטת ה-DSI נמצאה יעילה במדידת זיוף במידה דומה עבור סוגים שונים של פריטים. על כן, נראה כי היא יכולה להיות מיושמת בתהליך מיון הכולל מגוון של מבחנים.

שיטת ה-DSI מאפשרת מדידה של הבדלים אינדיבידואליים בזמני תגובה, עובדה ההופכת את השיטה החדשה למדויקת יותר והוגנת יותר כלפי אנשים איטיים מטבעם, ביחס לנורמה. למרות התוצאות המעודדות, מומלץ כי מחקרי המשך ימשיכו לבחון את תועלת השיטה במצבים נוספים, בתרבויות שונות, ובקרב מועמדים אמיתיים לעבודה. התקווה היא, שהמחקר הנוכחי יחד עם מחקרי המשך בנושא, יוביל לשיפור ביכולתם של חוקרים ויועצים לזהות זייפנים במבחנים פסיכולוגיים ובכך לשפר את תועלתם בתהליכי מיון ובכלל.

התפתחות כלי המיין במערכת המיין לצוות אויר: ממבחי כשרים בודדים לסימולטורי הטסה

ענבל רויטר

צה"ל

פרופיל 'איש צוות אויר' מורכב ממגוון של יכולות ומאפיינים, חלקם בהיבט האישי והאופי וחלקם בהיבט היכולות והכישורים. מורכבות מערכת המיין נעוצה בכך שמדובר בתמהיל יכולות, אשר חלקן נדירות באוכלוסייה ברמתן הנדרשת. עבודה זו תתמקד באותם כלי מיין הבוחנים את יכולותיו הקוגניטיביות של המועמד, בדגש על היכולות הספציפיות הנדרשות לטיסה.

מטרת העבודה להציג ולבחון את התפתחות כלי המיין למדידת כושר ויכולת טיסה לאורך השנים, ממבחינים הבודקים כשרים ספציפיים למבחיני סימולציה אינטגרטיביים המתייחסים למכלול המשימה ובוחנים את הביצוע השלם. בנוסף, נעמוד על ההיבטים הפסיכומטריים התומכים במעבר ממדידת יכולות בודדות למדידה אינטגרטיבית.

נבדקו הכלים במערכת המיין, בחלוקה ל'דור' התפתחות. מבחיני דור א' הינם המבחינים הקלאסיים, מבחיני נייר ועיפרון אשר הותאמו להעברה על גבי מחשב, כמו גם מבחיני ביצוע מכניים. מבחיני דור ב' הינו מבחיני סימולציה הטסה, אשר נבנה לבדיקת יכולת ההטסה וה'שימוש בידים' ודורש שליטה בידית היגוי (סטיק) ומצערת באופן הדומה לשימוש הנעשה בהם בהטסה בפועל במטוס. המשימה הינה פשוטה אך הביצוע נבחן באופן אינטגרטיבי. מבחיני דור ג' פותחו על מנת לדמות משימות מורכבות בתחום ההטסה. מעבר לשליטה בהטסת המטוס, נדרש איש צוות האוויר לניהול המשימה והתמודדות עם מגוון של מטלות וקבלת החלטות מורכבות בו-זמנית.

תוקף הניבוי של כלי המיין השונים, בחלוקה ל'דור' התפתחות' נבדק אל מול קריטריון ביצוע מבית הספר לטיסה. בהתאם למצופה, נמצא כי ככל שכלי המיין מורכבים ודומים יותר ל'דבר עצמו', כך יכולת הניבוי טובה יותר. יחד עם זאת, גם לחלק מכלי המיין ה'קלאסיים' עדין יש תרומה לניבוי (כפי שניתן לראות בכניסתם לנוסחאות רגרסיה לניבוי הקריטריון), מאחר ודוגמים יכולת מעט אחרת. בנוסף לכך, חלק מכלי המיין הפשוטים נדרשים לסינון גם בשלבי המיין המוקדמים.

עבודה זו מציגה מספר סוגיות מעולם המדידה. ראשית, מדידה של כשרים נפרדים ביחס למדידת הביצוע השלם כמשימה אינטגרטיבית אחת והתרומה שלהם לניבוי.

סוגיה שנייה קשורה למיין המבוסס טכנולוגיה. הטכנולוגיה מתפתחת ובהתאם פוטנציאל ההתפתחות של כלי המיין. היתרונות של מיין רווי טכנולוגיה ברורים. עם זאת קיימים קשיים ואתגרים לא פשוטים, החל מהעברת המבחינים, תיחזוקם, זמן הבחינה פר מועמד והעלות הגבוהה, וכלה באתגרים הפסיכומטריים: המעבר ממספר תשובות רבות ברירה לאלפי נתונים הנאספים כל דקת מבחן ועושר המידע שיש להפיק ולזקק. בנוסף, מפאת כוח הניבוי הברור והקירבה לקריטריון נשאלות שאלות של הוגנות ואתיקה במיין של מגוון אוכלוסיות וחיפוש בעלי פוטנציאל ההצלחה.

סוגיה שלישית קשורה בהכנסה של כלים מורכבים למערכת מיין מרובת כלים ושלבים והיכולת לשמר יציבות ותקפות של מערכת מיין הפועלת לאורך שנים לצד חידוש והכנסה של כלי מיין חדשים ומתקדמים ללא הוצאה בהכרח של הכלים המסורתיים.

סקירת מערכות מיון להשכלה גבוהה ברחבי העולם

איילת גולדצויג

מרכז ארצי לבחינות ולהערכה

בשנים האחרונות ישנה התעניינות גוברת בתהליך מיון וברירת מועמדים להשכלה גבוהה ברחבי העולם, בצד מגמה גלובלית של גידול במספר הסטודנטים בהשכלה הגבוהה. מטרת סקירה זו היא בחינה השוואתית של שיטות המשמשות למיון וקבלה להשכלה גבוהה במדינות שונות, ובכלל זה ישראל, מדינות ה-OECD, ומדינות נוספות מקבוצת G20.

כאשר מתבוננים בתיאור מערכות מיון שלמות במדינות שונות בעולם, נראה כי הן שונות מאד זו מזו. עם זאת, כאשר משווים היבטים שונים של מערכות מיון בעולם, עולה כי במרבית המדינות שנסקרו נעשה שימוש בציוני התיכון ו/או בתעודת בגרות כתנאי לקבלה להשכלה גבוהה. כמו כן, כמעט בכל מדינות ה-OECD נעשה שימוש במבחנים נוספים פרט לבגרות כתנאי קבלה, לפחות בחלק מתחומי הלימוד. אלה כוללים מבחני הישגים ארציים (35% מהמדינות), מבחני יכולת אחידים (14% מהמדינות), ומבחני קבלה ספציפיים המחוברים ומועברים בכל מוסד לימוד בנפרד (26% מהמדינות). ב-88% מהמדינות נעשה שימוש בקריטריונים נוספים לקבלה, וביותר ממחציתן קיימים גם אפיקי קבלה אלטרנטיביים.

אחד ההיבטים המעניינים להשוואה בין מדינות בהקשר של מיון להשכלה גבוהה הוא מידת הנגישות להשכלה גבוהה. דרך אחת לבחון זאת היא מידת התחרותיות בקבלה להשכלה גבוהה, שמושפעת מהיחס בין ההיצע לביקוש. ההיצע תלוי בקיבולת מוסדות הלימוד ובמספר המקומות הפנויים בכל מסלול לימודים, ואילו הביקוש להשכלה גבוהה מצד המועמדים מושפע מגורמים רבים, כגון מוטיבציה, הליך הקבלה להשכלה גבוהה עצמו, ואף מעצם קיומה של תחרותיות בהליך זה (OECD, 2012).

כדי ללמוד על מידת התחרותיות בקבלה להשכלה גבוהה, חושב היחס בין שיעור הכניסה להשכלה גבוהה ובין שיעור בוגרי התיכון הזכאים להירשם, כלומר המועמדים הפוטנציאליים. לא נמצא קשר עקבי בין מדד זה ובין מאפיינים אחרים של מערכות המיון במדינות שנסקרו. נראה כי גורמים שונים משפיעים על היחס בין שיעור הזכאים לשיעור המתקבלים ללימודים גבוהים, ובכלל זה הסללה בשלבים מוקדמים יותר של מערכת החינוך, קבלת סטודנטים זרים, וקיומן של אלטרנטיבות תעסוקתיות טובות שאינן מצריכות לימודים לתואר אקדמי. בדומה לכך, לא נמצא יחס פשוט בין שיטת המיון להשכלה גבוהה ובין שיעור בוגרי תואר ראשון או שיעור הנושרים מתואר ראשון. כלומר, היחס בין שיטות המיון ובין השלכותיהן מורכב, ומושפע מגורמים תרבותיים, חברתיים וכלכליים רבים.

לסיכום, נראה כי על אף שתהליך הרחבת ההשתתפות בהשכלה גבוהה הוא תהליך גלובלי, ישנה שונות רבה בין מערכות המיון הנמצאות בשימוש ברחבי העולם. מערכות אלה צמחו על רקע ההקשר החברתי, התרבותי, הכלכלי והפוליטי הקיים בכל מדינה, ולא ניתן להעריך את יעילותן במנותק מן ההקשר שבו הן נטועות. על כן, קיים קושי בהסקת מסקנות באשר ליעילותה האוניברסלית של מערכת מיון כזו או אחרת.

על הקשר בין מבנה הבחינה המסכמת בקורסים אקדמיים לבין סגנון הלמידה בקורס

ענת בן-סימון

מרכז ארצי לבחינות ולהערכה

נעמה סיימונס

המח' לפסיכולוגיה, האוניברסיטה העברית

מבחן, בהקשר האקדמי, נחשב להליך שמטרתו להעריך תוצרי למידה. תשומת לב מועטה מוקדשת לאופן שבו מבחנים מבנים ומעצבים את תהליך הלמידה באקדמיה. ספרות המחקר העוסקת בהיבט זה מלמדת שלאופן ההערכה של הלמידה (מבנה המבחן) יש השפעה לא מבוטלת על תהליך הלמידה עצמו בכלל ועל אסטרטגיות ההכנה של הלומדים למבחן בפרט¹.

מטרתו העיקרית של מחקר זה היא לבדוק את ההשפעה של מבנה השאלות במבחן מסכם בקורס אקדמי על סגנון הלמידה בקורס ועל אופן ההכנה למבחן. באופן ספציפי יותר מתמקד המחקר בהשוואה בין המוטיבציות והאסטרטגיות שסטודנטים מיישמים בתהליך ההכנה למבחן המורכב משאלות ברירה לבין אלה המיושמות בהכנה למבחן המורכב משאלות פתוחות.

המשתתפים במחקר היו 126 סטודנטים מחמש אוניברסיטאות, ממגוון פקולטות (חברה, רוח וטבע) ולכל התארים (ראשון, שני ושלישי). 88 מהם היו נשים ו-38 גברים. טווח הגילים נע בין 19 ל-32 (ממוצע=24.7, ס.ת.=2.4). הציון הממוצע של המשתתפים בחוג הלימודים הראשי (לפי דיווח עצמי) נע בין 59.2 ל-100, (ממוצע = 88.7, סטיית תקן = 6.3).

הכלי המרכזי שנעשה בו שימוש במחקר הוא שאלון ה-R-SPQ-2F². השאלון כולל 20 פריטים ומודד שתי גישות ללמידה, גישה שטחית וגישה עמוקה, כאשר לכל גישה שני רכיבים, רכיב מוטיבציוני ורכיב אסטרטגי. לצורך המחקר תורגם השאלון והותאם לשפה העברית.

מערך המחקר היה תוך נבדקי וכלל שלושה שלבים. בשלב הראשון הוצג למשתתפים קורס היפותטי בפילוסופיה של המדע אשר הציון בו נקבע באמצעות מבחן המורכב משאלות ברירה. לאחר התיאור התבקשו המשתתפים לענות על שאלון ה-R-SPQ-2F. בשלב השני הוצג למשתתפים תיאור של אותו קורס אשר הציון הסופי בו מורכב משאלות פתוחות והם התבקשו לענות שוב על שאלון ה-R-SPQ-2F. בשלב השלישי מילאו המשתתפים שאלון פריטים אישיים. למחצית מהמשתתפים הוצג שלב ב' לפני שלב א'.

הנבדקים הוקצו באופן אקראי לאחת משתי הגרסאות של כלי המחקר. בתום תהליך הפיתוח של כלי המחקר הוא הועבר לגרסה מקוונת. ההזמנה להשתתפות במחקר הופצה באמצעות הפייסבוק.

1 Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35(4), 453-472.

2 Biggs, J., Kember, D., & Leung, D. Y. (2001). The revised two factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71(1), 133-149.

תוצאות המחקר מצביעות על הבדלים משמעותיים בין מבחנים פתוחים למבחנים סגורים באסטרטגיות ובמוטיבציות שסטודנטים מדווחים שיעשו בהם שימוש במסגרת הלמידה וההכנה למבחן סוף קורס. הממצאים מלמדים שבמבחנים פתוחים יעשה שימוש רב יותר בגישה עמוקה, המזוהה עם למידה איכותית יותר, הישגים גבוהים יותר ותחושות חיוביות יותר כלפי הלמידה³. בהתאמה, במבחנים סגורים ייעשה שימוש רב יותר בגישה השטחית, המזוהה עם הישגים נמוכים יותר ותחושות של חוסר שביעות רצון, שעמום ושנאה⁴. כיוון שמבנה ההערכה משפיע על בחירת הגישה ללמידה, נראה כי מומלץ לקחת בחשבון השלכות אלה בבחירת שיטת ההערכה.

3,4 Biggs, J. (1987). Study Process Questionnaire Manual. Student Approaches to Learning and Studying. Australian Council for Educational Research Ltd., Radford House, Frederick St., Hawthorn 3122, Australia..

ברירה ללימודים גבוהים לפי בגרות בלבד: אמידת השינוי בסף הבגרות הנדרש לקבלה למגוון חוגים מתוך נתוני הלומדים

יונתן סער

מרכז ארצי לבחינות ולהערכה

מפעם לפעם מועלות בציבור הצעות לשינוי תהליך הברירה ללימודים באוניברסיטאות בישראל. מחקר זה בא לבחון את אחת מהשפעותיו האפשריות של שינוי תיאורטי מהסוג שהוזכר לעיל, ובפרט, מנסה לאמוד את השינוי שעשוי לחול בסף הבגרות הנדרש לקבלה למגוון חוגים אקדמיים בתרחיש בו שיטת הברירה ללימודים תסתמך על ממוצע הבגרות בלבד, ולא על ציון סכם – שקלול ממוצע הבגרות וציון המכפ"ל (מבחן כניסה פסיכומטרי לאוניברסיטאות), כנהוג כיום. בעוד שהאמידה פשוטה יחסית אם ידועים נתוני המועמדים לחוג, הרי שהיא נעשית מורכבת יותר כשידועים רק נתוני הלומדים בפועל בחוג. העבודה הנוכחית מציגה שיטה לאמידת השינוי הצפוי בסף הבגרות בהתבסס על נתוני הלומדים בלבד, ומדגימה אותה על נתוני הלומדים במגוון חוגים ומוסדות בשנת 2009.

תחת מספר הנחות באשר להתפלגות המשותפת של ציוני המכפ"ל וממוצעי הבגרות בקרב המועמדים לחוג ובאשר לאופן הברירה לקבלה ללימודים בחוג, ניתן לתאר בצורה כללית את השיטה שעליה נסמך המחקר הנוכחי באופן הבא:

- א. מתבוננים בדיאגרמת הפיזור של ציוני המכפ"ל וממוצעי הבגרות של הלומדים בחוג, ומסננים מתוכה תצפיות קיצוניות שניכר עליהן בעליל שלא עברו ברירה במסלול המקובל על-פי ציון סכם.
- ב. מבין התצפיות הנותרות, מזהים את קבוצת בעלי ציון הסכם המינימאלי בקרב הלומדים בחוג.
- ג. בקרב הקבוצה שתוארה בסעיף ב', מזהים את בעלי ממוצע הבגרות הנמוך ביותר. זהו סף הבגרות הנוכחי.
- ד. בקרב הקבוצה שתוארה בסעיף ב', מזהים את בעלי ממוצע הבגרות הגבוה ביותר.
- ה. סף הבגרות העתידי המשוער הוא אמצע הטווח שבין ממוצע הבגרות הנמוך ביותר (שחושב בסעיף ג') לבין ממוצע הבגרות הגבוה ביותר (שחושב בסעיף ד') בקרב בעלי ציון הסכם הנמוך ביותר מבין הלומדים.
- ו. מחשבים את ההפרש בין הסף העתידי (שחושב בסעיף ה') לסף הנוכחי (שחושב בסעיף ד'). זוהי העלייה הצפויה בסף הבגרות לקבלה לחוג.

מהתוצאות עולה כי ברירה על סמך בגרות בלבד צפויה להעלות את סף הבגרות המינימלי לקבלה לחוג בקרוב ל-6 נקודות בממוצע, המהוות 0.55 – 0.85 סטיית תקן (כתלות במוסד) של ממוצע הבגרות בקרב כלל הלומדים במוסד בשנת 2009. זהו אפקט משמעותי, במיוחד אם לוקחים בחשבון שבתנאים מציאותיים, שינוי מדיניות הברירה צפוי לגרום אחריו שינוי התנהגותי בקרב המועמדים; המשאבים שהשקיעו בעבר לטובת המכפ"ל יופנו לטובת שיפור ממוצע הבגרות. שינוי התנהגותי כזה עשוי להוביל לעלייה חדה עוד יותר בסף הבגרות מזו שחזה המודל המוצע.

תוצאות המחקר ממחישות כי, בניגוד לתפיסה המקבלת ביטוי בציבור, לפיה תהליך הקבלה לאוניברסיטאות צריך להישען על מספר מצומצם ככל האפשר של כלים ומדדים, הרי שבשימוש בכמה כלים לצרכי ברירה ישנם גם יתרונות, ולא רק חסרונות. במחקר הנוכחי הודגמה התופעה על-ידי מקרה שבו הברירה נעשת לפי בגרות בלבד; כמובן שגם ברירה לפי מכפ"ל בלבד היתה גוררת עלייה בסף המכפ"ל הנדרש לקבלה ללימודים. היתרון המומחש כאן בריבוי כלים ראוי שיעמוד מול עיניהם של מקבלי החלטות בתחום המדיניות החינוכית בבואם לעסוק בנושאים מעין אלו.

המחשת התוקף של שלושה ציוני מכפ"ל כלליים בניבוי ציוני שנה א' באוניברסיטאות

נרמל אורן, תמר קנת-כהן, אליוט טורוול, אבי אללוף

מרכז ארצי לבחינות ולהערכה

בשנת 2011 נוספו למבחן הכניסה הפסיכומטרי לאוניברסיטאות (מכפ"ל) שני ציונים כלליים מדווחים, שונים זה מזה במשקלות היחסיים של מרכיבי המבחן: באחד ניתן דגש למרכיב המילולי, בשני ניתן דגש למרכיב הכמותי. הרציונל להוספת הציונים החדשים היה ששימוש בהם צפוי להגדיל, ולו במעט, את תוקף הניבוי של מערכת המיון בתחומים שיש בהם דגשים בהתאמה. מעבר לכך, הציונים הללו יעניקו יתרון למי שציוני היחסיים (לעצמו) בתחומי מכפ"ל נמצאים בהתאמה עם בחירתו המקצועית וכנראה גם עם נטיותיו המקצועיות. **מטרות המחקר:** לבדוק, להשוות ולדווח לראשונה את תוקף הניבוי של ציונים כלליים אלה ושל הציון הכללי המקורי – בתוך קבוצות שונות של תחומי לימוד. מטרה נוספת היא להמחיש את התועלת בשימוש במבחן במונחים שאינם מתאם: במונחי שיפור בציון הקריטריון, ובמונחי שיפור בהשמות הנכונות של בעלי רמות הישגים שונות בלימודים, למשל מצטיינים, על פי המנובא להם מכלי המיון. **שיטה:** נתוני המחקר מתבססים על חמישה מחזורים של תלמידי שנה א' שהחלו לימודיהם בשנים 2005–2009, בחמש האוניברסיטאות ובטכניון, סה"כ 100,863 תצפיות. יחידת העיבוד הבסיסית לכל העיבודים היא חוג-בשנה-במוסד, סה"כ 1,095 חוגים. התוצאות מקובצות, תוך מיצוע משוקלל בגודל החוגים, בתוך הפקולטות הבאות: מדעי הרוח, מדעי החברה-מילוליים, משפטים, מדעי החברה-כמותיים, מדעי הטבע, הנדסות, רפואה, מקצועות פארא-רפואיים. **משתני המחקר:** חזאים: בגרות, שלושת מרכיבי מכפ"ל, שלושה ציוני מכפ"ל הכלליים, שלושה ציוני סכם של בגרות פלוס ציון מכפ"ל כללי במשקלות שווים. הקריטריון: ממוצע ציוני שנה א'. **עיבודים:** בנוסף להצגת הסטטיסטיים הפשוטים (שכיחויות, ממוצעים, ס.ת.), מוצגים הסטטיסטיים הבאים: א. משקלות מתוקננים של החזאים על הקריטריון ברגרסיה לינארית – לאומדן משקלם היחסי בניבוי, ב. מתאמים רגילים ומרובים, נצפים ומתוקננים לקיצוץ תחום בחזאים, של החזאים עם הקריטריון, לאומדן תוקף הניבוי שלהם. בנוסף, נערכה סימולציה המשתמשת בנתוני התוקף האמפיריים שהתקבלו, במדגם היפותטי של מועמדים שאינו מקוצץ על טווח ציוני הקריטריון. תוצאות הסימולציה משמשות להמחשת הערך המוסף שיש לשימוש בכלי מיון בעל תוקף נתון, במונחי ציוני הקריטריון, ובמונחי שיעור השמות נכונות לקבוצות יכולת. **תוצאות ומסקנות:** ניתוחי הרגרסיה מראים כי תחומי הלימוד נחלקים לשני גושים: אלה שבהם משקל התחום הכמותי בניבוי גדול משמעותית (פי ארבעה לפחות) מזה של התחום המילולי: טבע, חברה-כמותי, הנדסות, רפואה, ואלה שבהם משקל התחום המילולי (ואנגלית) גדול משל התחום הכמותי: חברה-מילולי, רוח, משפטים, ומקצועות פארא-רפואיים. המתאמים מראים כי בפקולטות בעלות דגש מילולי, מסתמן יתרון לציון הכללי בדגש מילולי, ובאלו בעלות דגש כמותי – יתרון לציון מכפ"ל הכללי בדגש כמותי. בממוצע מעבר לתחומי הלימוד, תוקף ציון הסכם גבוה מזה של ציון מכפ"ל, וזה בתורו גבוה מהתוקף של הבגרות: 0.46, 0.43 ו-0.36. בהתאמה. תוקף ארבעת המרכיבים במשקלות אופטימליים (המתאם המרובה) הוא 0.54. ברוב הפקולטות, צירוף מכפ"ל ובגרות מניב תוקף גבוה יותר מאשר כל אחד מהם בנפרד. זאת – למעט משפטים, חברה מילולי ומקצועות פארא-רפואיים, שבהם הוספת הבגרות במשקל שווה למכפ"ל, אינה תורמת לתוקף מעבר למכפ"ל לבדו.

האם ההכנה למבחני הקבלה לרפואה מסייעת למועמדים?

אביטל מושינסקי, דוד ציגלר

מרכז ארצי לבחינות ולהערכה

החל משנת 2004 מפעילה הפקולטה לרפואה של אוניברסיטת תל-אביב מרכז הערכה למועמדים ללימודי רפואה (מור) שבו מוערכים משתנים שאינם קוגניטיביים, וזאת כתחליף לריאיון הקבלה שהיה נהוג עד אז. בשנת 2006 החל גם בית הספר לרפואה של הטכניון למיין את המועמדים לרפואה באמצעות מערכת מור, ובשנת 2011 הצטרפה הפקולטה לרפואת שיניים של אוניברסיטת תל-אביב. הפקולטה לרפואה של האוניברסיטה העברית בירושלים מפעילה החל משנת 2006 מרכז הערכה בשם מרקם – מערכת ראיונות קצרים ומובנים, שבו מוערכים אותם המשתנים.

מערכות מור ומרקם כוללות שני שאלונים – שאלון ביוגרפי ושאלון שיפוט והחלטה, ושמונה תחנות התנהגותיות. מרכיב השאלונים משותף לשתי המערכות.

בכינוסי אפי קודמים הוצגו מערכות מור ומרקם וכן נתונים הנוגעים למהימנות ולתוקף שלהן. ההרצאה הנוכחית תתמקד בנושא ההכנה של המועמדים לבחינה ובניסיון להשיב לשאלה האם הדבר משפיע על ביצועיהם במרכז ההערכה.

בבחינת מור נבחנים 96 מועמדים מדי יום ובבחינת מרקם נבחנים 16 – 32 מועמדים מדי יום. מטעמי עלות אי אפשר לפתח נוסח בחינה נפרד לכל יום בחינה, וכתוצאה מכך פריטי בחינה חוזרים על עצמם במהלך ימי הבחינה של אותה שנה ואף בין השנים.

החל מהשנה הראשונה שבה הופעל מרכז ההערכה מור, המועמדים העלו מדי יום את כל פריטי הבחינה לרשת האינטרנט מיד עם תום הבחינה. בעקבות זאת מועמדים שנבחנו בימי הבחינה הראשונים, התלוננו שלמועמדים הנבחנים בימי הבחינה המאוחרים יותר ואשר נחשפו לחלק מפריטי הבחינה, יש יתרון, מכיוון שהם יכולים להתכונן טוב יותר לבחינה.

בהרצאה יוצגו נתונים הבודקים את השאלה האם תלונות אלו מוצדקות, ויפורטו דרכי ההתמודדות עם תלונות אלו. כמו כן, יוצגו נתונים הבודקים אם להיבחנות חוזרת יש השפעה על הציונים, ואם השיפור בציון היבחנות חוזרת נובע מעצם ההתנסות בבחינה. בנוסף, יוצגו נתונים אודות אחוז המועמדים המשתתפים בקורסי הכנה לבחינה, ואודות הערכתם את טיב קורס ההכנה שהם עוברים.

ניתוח מבנה הבחינה הפסיכומטרית החדשה - ניתוח גורמים מגשש ומאשש בשלושה חלקים

דביר קלפר

מרכז ארצי לבחינות ולהערכה

מטרת מחקר זה היא לתאר ולסכם תוצאות של מספר ניתוחים, אשר נערכו על מנת לאפיין את המבנה של כלי המיון המרכזיים לאוניברסיטאות בכלל, ובפרט של הבחינה הפסיכומטרית במתכונתה החדשה הנוהגת החל מספטמבר 2012.

מחקרים ברוח דומה נעשו במרכז הארצי לבחינות ולהערכה בעבר. למשל, בודסקו (1985) בדק את מבנה הבחינה הפסיכומטרית במתכונתה הראשונה (כאשר כללה חמישה תחומים: צורות, מתמטיקה, הבנה מילולית, אנגלית וידע כללי). במחקרו נמצאו שני גורמים – גורם של ידע (שכלל את ידע כללי ואנגלית) וגורם של חשיבה כמותית (שכלל את התחום הכמותי וצורות), כאשר התחום המילולי נטען על שניהם. מחקר נוסף נעשה ע"י בלר (1990) וכלל ניתוח גורמים כאחת משלוש שיטות ניתוח מבנה – מודל עצים חיבורי (ADDTREE) (Sattath & Tversky, 1977), אנליזת מרחב מינימלי (SSA - Smallest Space Analysis) (Guttman, 1968), וניתוח גורמים. גם במחקר זה נמצאו שני גורמים מרכזיים: ידע ופתרון בעיות.

העבודה הנוכחית חוקרת את המבנה הפנימי של המבחן הפסיכומטרי במתכונתו החדשה.

החל מספטמבר 2012 כוללת הבחינה הפסיכומטרית חלק נוסף – מטלת כתיבה – המהווה 25% מן הציון בחלק המילולי. מטרת עבודה זו לראות כיצד משתלב חלק זה בשאר החלקים של הבחינה הפסיכומטרית בפרט, ולחקור את המבנה של הבחינה הפסיכומטרית בכלל.

בעבודה זו התקבלו גורמים דומים לאלו שהתקבלו בעבר, אלא שבמחקר הנוכחי, בנוסף לשני הגורמים האלו (ידע ופתרון בעיות), נבדק מבנה שבו נכפה פתרון של שלושה גורמים.

עבודה זו מחולקת לשלושה חלקים:

בחלק הראשון ביצענו ניתוח גורמים מגשש (Exploratory Factor Analysis) עבור אחד הנוסחים במתכונת החדשה. הניתוח בחלק זה נעשה לפי ציונים של מקבצי פריטים לפי סוג הפריט בהתאם לאופן שבו הם מוצגים בבחינה עצמה. חלק זה בוצע בשלושה שלבים: בשלב הראשון הניתוח התבצע על הבחינה הפסיכומטרית בלבד, בשלב השני הוכנס גם ציון הבגרות כרכיב נפרד, ובשלב השלישי חזרנו על הניתוח כמו בשלב השני רק שהפעם המתאמים תוקנו למהימנויות.

בחלק השני בוצע ניתוח מרכיבים ראשי מגשש (Exploratory Principal Component Analysis) ברמת הפריטים הבודדים. הניתוח נעשה בשני שלבים: על כל הבחינה ביחד (131 שאלות) ועל כל אחד מהתחומים בנפרד: כמותי (40 שאלות), מילולי (47 שאלות) והאנגלית (44 שאלות). לצורך ביצוע חלק זה חושבו מתאמים טטרכוריים בין השאלות רבות הברירה לבין עצמן ומתאם בי-סריאלי בין ציון מטלת הכתיבה לבין שאלות הברירה.

לבסוף, בחלק השלישי, בוצע ניתוח גורמים מאשש (Confirmatory Factor Analysis) ברמת סוג פריט וזאת על מנת לבדוק את המודל שהתקבל בחלק הראשון (ניתוח מגשש ברמת סוג פריט). נבדק מודל סטנדרטי ולבסוף הוצע מודל לא סטנדרטי.

האומנם מוכנים מורי העתיד לאתגר של הערכת לומדים?

עדי לוי-ורד

אוניברסיטת תל-אביב

הנושא של הערכת לומדים ניצב בעשור האחרון במוקד דיון סוער בשדה החינוך (Chen, 2005). על אף שמורים מקדישים זמן רב (בין שליש לחצי מזמנם המקצועי) לפעולות הקשורות בהערכה, נשאלת השאלה האם אמנם יש בידיהם את הידע והכלים לביצוע הערכה ראויה, והאם מוכנים הם לאתגרי ההערכה הניצבים בפניהם בעידן הפוסט-מודרני?

המחקר בא להציג "תמונת מצב" ביחס לאוריינות ההערכה בקרב מורים מתחילים בישראל, תוך בחינת מידת עמידתם בסטנדרטים של אוריינות בהערכה, ותוך בחינת אופן הכשרת המורים בתחום של הערכת הישגי לומדים, במכללות להוראה בישראל. בנוסף, נבנה ונבחן מודל מבני הקושר בין האוריינות בהערכה לבין מאפייני ההכשרה בהערכה, החוללות העצמית של המורים המתחילים ותפיסותיהם ביחס להערכה.

המדגם כלל 327 נחקרים, מתוכם 187 מורים מתמחים ו-140 מורים המלמדים בשנה הראשונה שלאחר ההתמחות. שאלוני המחקר הועברו בעשר מכללות להוראה בישראל (מכללות עבריות, מהמגזר הממלכתי והממלכתי-דתי וממחוזות שונים בארץ). בנוסף, רואיינו 11 מורים ונותחו 18 סילבוסים של קורסי "מדידה והערכה" שנלמדו ב-15 מכללות להוראה בארץ. הנתונים הכמותיים נאספו באמצעות גרסא עברית של מבחן "אוריינות בהערכה" (Mertler & Campbell, 2005) ומספר שאלונים מובנים.

ממצאי המחקר הצביעו על מידה דלה ולא מספקת של אוריינות בהערכת לומדים בקרב מורים (בהתייחס למגוון תחומים, פעולות וסטנדרטים שונים). בממוצע הצליחו המורים להשיב נכון על 47% בלבד מכלל שאלות המבחן. במקביל, דיווחו הנחקרים על מידה מועטה של ידע שרכשו בפעולות הערכה שונות במסגרת ההכשרה, אם כי, ניתוח הסילבוסים הצביע באופן כללי על הימצאות תחומי תוכן אלה בקורסי "מדידה והערכה" במכללות. שליש מהנחקרים כלל לא למדו קורסים בנושאי הערכת לומדים במסגרת הכשרתם. עוד נמצא כי 85% מהנחקרים לא התנסו כלל, או התנסו במידה מועטה בלבד, בהערכת לומדים במסגרת הכשרתם במכללה. הממצאים האיכותניים העלו תמונה דומה, על פיה חשים הנחקרים צורך ורצון עז ללמוד ולהתנסות באופן משמעותי יותר בהערכה במסגרת ההכשרה. בחינת מרכיב המודלינג העלתה כי ההערכה המסורתית היא השכיחה במסגרת ההכשרה להוראה וכי מתן ציון מספרי נותר כשיטת המשוב העיקרית שניתן לסטודנטים על ידי המרצים.

המחקר הוביל לעשר המלצות ביחס לפיתוח האוריינות בהערכה ולשיפור איכות ההכשרה של המורים בתחום ההערכה: א. הגדרת לימודי יסוד באוריינות הערכה המשלבים בין התיאוריה למעשה, ב. קביעת מדיניות לפיה קורס "אוריינות בהערכת לומדים" הינו קורס חובה, ג. מתן הזדמנויות להתנסות משמעותית ואפקטיבית בהערכת לומדים, ד. הדגמת פעולות הערכה ראויות תוך שיתוף הסטודנטים בתהליכי הערכה (מודלינג), ה. זיהוי, דיון ועיצוב תפיסות הסטודנטים להוראה ביחס להערכת לומדים, ו. בניית ערוצי תקשורת ושיתוף פעולה בין קובעי מדיניות במשרד החינוך, מוסדות להכשרת מורים, מחוזות ובתי הספר, ז. התפתחות רציפה בתחום ההערכה לאורך כל החיים המקצועיים, ח. שימוש בטכנולוגיה ככלי מסייע ומעשיר בהערכת לומדים, ט. יצירת מנגנוני בקרה על איכות ותרומת ההכשרה בהערכה, י. התייחסות אינטרדיסציפלינרית לתחום ההערכה והשתמעויות להכשרה אפקטיבית.

סימפוזיון סיכום של שנה לרפורמה בבחינה הפסיכומטרית - הוספת מטלת כתיבה ועוד...

יו"ר: נעמי גפני

מרכז ארצי לבחינות ולהערכה

הקדמה

מ-1990 ועד 2012 לא חלו שינויים מהותיים במבנה הבחינה הפסיכומטרית. לאורך כל התקופה כללה הבחינה שמונה פרקים שבהם שאלות בִּרְרָה משלושה תחומים: חשיבה מילולית, חשיבה כמותית, ואנגלית. שישה מתוך הפרקים שימשו לחישוב הציון, ושניים לבדיקת שאלות חדשות ולכיול הציונים. במהלך השנים נחקרה הבחינה על היבטיה השונים ונמצא כי מהימנותה גבוהה ביותר וכי תוקף הניבוי שלה גבוה יחסית, ודומה לזה שנמצא בעולם עבור בחינות דומות. כמו כן נמצא כי השימוש בבחינה בצירוף ציון הבגרות הוגן לשני המינים, לנבחנים ששפת אימם אינה עברית (ובפרט לנבחנים בערבית), וכי הוא לא מושפע ממצבו החברתי-כלכלי של הנבחן. למרות הנתונים שהצביעו על איכותה של הבחינה, סבלה הבחינה מביקורת ציבורית על כך שאינה רלבנטית לתכניות הלימוד השונות, שהיא מוטה כנגד קבוצות שונות, ושההכנה אליה יקרה ומיותרת. ביקורת זאת הושמעה גם ע"י חברי כנסת שבעשרים השנה האחרונות הגישו למעלה מעשר הצעות חוק לביטול הבחינה הפסיכומטרית.

המרכז הארצי לבחינות ולהערכה (מאל"ו) הכיר בצורך לשינוי הבחינה כך שיינתן מענה לחלק מהביקורות שהופנו כלפיה. אחד מהשינויים שנשקלו היה הוספה של מטלת כתיבה לתחום החשיבה המילולית. כבר בשנות התשעים פיתח מאל"ו מבחן שכלל שלוש מטלות כתיבה ואשר הוקצבו לו 90 דקות. מבחן זה לא תוכנן כחלק מהבחינה הפסיכומטרית אלא כמבחן העומד בפני עצמו. למרות שהמוסדות גילו בו עניין לא נעשה בו שימוש.

ב-2008 הוחל בתכנון רפורמה בבחינה הפסיכומטרית שכללה כמה אפשרויות להוספת מטלת כתיבה. הרציונל לתוספת של מטלת הכתיבה נבע משיקולי תוקף – הערכת יכולת כתיבה תחזק את הקשר בין המיומנויות הנמדדות ע"י הבחינה הפסיכומטרית לבין אלו הנלמדות ונדרשות הן בבתי הספר הן במוסדות להשכלה גבוהה. כמו כן היה מקום לקוות כי לשינוי תהיה השפעה על הוראת הכתיבה בבתי ספר ועל הגברת המודעות לחשיבותה.

ב-2010 התכנון של המבנה החדש של הבחינה הפסיכומטרית על כלל השינויים הכלולים, בו היה מוכן להצגה בפני נציגי מערכת החינוך, הציבור וחברי הכנסת. השינויים שהוכנסו הם:

דיווח ציונים כלליים נוספים

החל ממועד אוקטובר 2011 חושבו לכל נבחן עוד שני ציונים, בנוסף לציונים הניתנים בנפרד לתחומי הבחינה (מילולי, כמותי ואנגלית) ולציון הכללי, שחושבו עד אז:

- ציון כללי בדגש הומני, שבו משקל הציון בחשיבה מילולית 60%, משקל הציון בחשיבה כמותית 20%, ומשקל הציון באנגלית 20%.
- ציון כללי בדגש ריאלי, שבו משקל הציון בחשיבה כמותית 60%, משקל הציון בחשיבה מילולית 20%, ומשקל הציון באנגלית 20%.

שינויים במתכונת הבחינה

החל ממועד אוקטובר 2012, לתחום החשיבה המילולית התווסף פרק ובו מטלת כתיבה. לפרק זה מוקצבות כ-30 דקות. כדי להימנע מהארכת הבחינה, קוצר הזמן המוקצב לכל אחד משמונת הפרקים האחרים, והוא עומד על כ-20 דקות.

במקביל צומצם מספר השאלות בכל פרק, ובתוך כך:

- א) מפרקי החשיבה המילולית הוצאו השאלות מהסוגים "מילים וביטויים" ו"החלפת אותיות".
- ב) מפרקי החשיבה הכמותית הוצאו השאלות מסוג "השוואה כמותית".
- ג) בפרקי האנגלית הוקטן מספר השאלות תוך שמירה על כל סוגי השאלות.

בסימפוזיון תוצג הרפורמה על שלביה השונים. נעמי גפני תציג את השינויים שעליהם הוחלט ואת הרציונל המונח בבסיסם, כמו גם את ההכנות שנדרשו לצורך יישום הרפורמה. גלעד סופר יציג את תהליך פיתוח מטלות הכתיבה והמחווון בשפה העברית, ואת תהליך הגיוס והמיון של המעריכים, ואבי אללוף יציג נתונים שונים שמתבססים על ביצועי הנבחנים במהלך שנת המתכונת החדשה הראשונה.

הכנות לקראת יישום הרפורמה בבחינה הפסיכומטרית

נעמי גפני

מרכז ארצי לבחינות ולהערכה

בעקבות ההחלטה על השינויים, נערך מאל"ו על כל מדוריו לקראת יישומם באוקטובר 2011 (דיווח הציונים) ובאוקטובר 2012 (מתכונת חדשה של הבחינה).

הכנות לקראת יישום הרפורמה

ההכנות לקראת יישום הרפורמה כללו: פיתוח מטלות כתיבה, פיתוח פרקי בחינה במתכונת חדשה, פיתוח מחווון, תכנון סדנאות למעריכים, עיצוב כתובון, פיתוח חומרי הסבר לנבחנים. כמו כן היה צורך להיערך להעברת הבחינה על שני חלקיה. מדור ציונים נערך לקראת גיוס ומיונם של מאות מעריכים בחמש שפות, היה עליו לפתח שיטות לניהול תהליך הערכת הכתיבה והתאמת תכניות חישוב הציונים למתכונת החדשה.

מטלת הכתיבה

לאחר קבלת ההחלטה על תוספת מטלת הכתיבה ננקטו הצעדים הבאים:

1. סקירת ספרות מקיפה של מבחני מיון להשכלה גבוהה שכללו מטלת כתיבה
2. אמידת המהימנות הצפויה של כל אחד מתחומי הבחינה ושל הבחינה כולה
3. אמידת התוקף הצפוי של הבחינה במתכונתה החדשה
4. הגדרת אפיוני מטלת הכתיבה
5. החלטה על הזמן המוקצב למטלה
6. פיתוח מאגר מטלות כתיבה.

ניסויים

הוחלט לערוך מספר ניסויים במטרה לאסוף מידע שיסייע בקבלת החלטות בנושאים הבאים:

1. מספר השאלות בפרקים המקוצרים
2. מהימנות הפרקים הקצרים
3. שיטת ההערכה – הוליסטית או אנליטית
4. סולם הציונים במטלת הכתיבה
5. איכות הערכת מטלות הכתיבה
6. ניסוח מחוון להערכת מטלת הכתיבה
7. ניסוח הוראות לנבחנים ולמשגיחים
8. עיצוב הכתובן

נערכו שני ניסויים. בניסוי הראשון הועברו 12 מטלות בעברית ובערבית, והשתתפו בו 745 נבחנים בעברית, ו-628 בערבית. בניסוי השני השתתפו 929 נבחנים בעברית ו-325 בערבית, והועברו בו שמונה מטלות בעברית וארבע בערבית. בעקבות הניסויים התקבלו ההחלטות הבאות:

1. החיבורים יוערכו בשיטת ההערכה האנליטית על שני ממדים: תוכן ולשון.
2. סולם ציונים לכל ממד יהיה 1-6, כך שהציון של כל מעריך ינוע בין 2 ל-12 וציון המטלה המבוסס על סכום ציוני המעריכים בין 4 ל-24.
3. הזמן שיוקצה לכתיבה יהיה 30 דקות בעברית ו-35 בירת השפות.
4. היקף הכתיבה שיידרש יהיה בין 25 ל-50 שורות. לחיבורים שאורכם קצר מעשר שורות יינתן ציון מינימאלי.
5. נקבע קריטריון להערכה שלישית – פער בין מעריכים של יותר 4 נקודות
6. באפריל 2012 פורסם המחוון בלויית דוגמאות.

מספר הפריטים בפרקי הבררה

בעקבות הניסויים הוחלט כי פרק חשיבה מילולית יכלול 23 פריטים בעברית ו-20 בכל אחת מהשפות האחרות; פרק בחשיבה כמותית יכלול 22 פריטים וכך גם פרקי האנגלית. עם זאת מאל"ו שומר לעצמו את הזכות לשנות את מספר הפריטים בפרק.

התאמות

כדי לספק פתרון לקבוצת המתקשים בכתיבה ננקטו הפעולות הבאות: פותחו גרסאות ממוחשבות של המתכונת החדשה של מכפ"ל (מפע"ם), ניתנה האפשרות לכתוב חיבור במחשב, לנבחנים במפע"ם ניתנה האפשרות לכתוב בנייר ועיפרון, ולנבחנים בנייר ועיפרון ניתנה האפשרות לכתוב במחשב.

הזמן המוקצב למטלת הכתיבה

נערך ניסוי שמטרתו לבדוק את ההשפעה של תוספת זמן על הציון במטלה. בניסוי השתתפו 482 נרשמים לבחינה הפסיכומטרית. המשתתפים חולקו באופן מקרי לארבע קבוצות לפי מספר הדקות שהוקצבו לכתיבה: 30, 35, 40, ו-45 דקות. התוצאות הראו שלזמן המוקצב לא הייתה השפעה על הציון הכללי במטלה. אפקט האינטראקציה שבין הזמן המוקצב לבין המטלה לא נמצא מובהק. כמו כן נראה כי המתאמים בין המשתנים (בין הציון במטלה לבין הציון בפרקי הבררה) לא היו שונים עבור הקבוצות השונות.

סוגיות בפיתוח מבחן כתיבה: פיתוח הדגם המתאים למטלת הכתיבה, שיקולים בבחירת סוג המחון, מיון והכשרת המעריכים

גלעד סופר, ספי פומפיאן

מרכז ארצי לבחינות ולהערכה

פיתוח מטלת הכתיבה במכפ"ל

הצעד הראשון בפיתוח היה בחינה של כמה סוגיות עקרוניות הנוגעות לאפיון מטלות כתיבה. אלה כוללות: מהימנות, תוקף, אותנטיות, בחינה עקיפה מול ישירה, מידת ההסתמכות על קריאה, מידת ההסתמכות על ידע קודם, תפקיד הגריין במטלה ומידת המיקוד של המטלה. בנוסף לבחינת סוגיות אלה, נסקרו מטלות הערכת הכתיבה במבחנים מקבילים למכפ"ל, וכן מבחן כתיבה שפותח במאל"ו בשלהי שנות התשעים ומבחינה בגרות בהבעה.

בחינת הסוגיות העקרוניות, מסגרת האילוצים המעשיים וסקירת המבחנים, הביאו לידי פיתוח ראשוני של ארבעה דגמי מטלות: מטלה המבוססת על גריין ארוך אחד; מטלה המבוססת על שילוב של שני גריינים; מטלה המבוססת על גריין קצר אחד; שתי מטלות המבוססות על גריין קצר.

לצורך הערכת הדגמים השונים נערכו שני ניסויים רחבי היקף. בניסוי הראשון נבדקו 12 מטלות, שהועברו כחלק מניסוי שבדק את הפורמט החדש של הבחינה כולה. הניסוי השני, שבו נבדקו 8 מטלות, התבסס על תוצאות הניסוי הראשון והוא נועד לאפיון מפורט ומתקדם יותר של מטלת הכתיבה.

בהסתמך על שילוב הנתונים הסטטיסטיים שהתקבלו בשני הניסויים, על משובי הנבחנים, ועל עמדתם המקצועית של חברי צוות הפיתוח, גובש דגם המטלה התפעולי: מטלת כתיבה ישירה, שבה נדרשת כתיבה עיונית-טיעונית, המשלבת בין הדגם הקצר והארוך.

שיקולים בבחירת סוג המחון (הוליסטי או אנליטי)

שיטת ההערכה ההוליסטית היא הנפוצה ביותר בפרויקטים רחבי היקף של הערכת יכולות כתיבה. ההערכה ההוליסטית מבוססת על תפיסת הכתיבה כיכולת אחת, שאי אפשר, או לא כדאי, להעריכה בדרך של פירוקה להיבטים שונים. בהערכה בגישה זו החיבור מקבל מהמעריך ציון אחד.

ההערכה האנליטית מבוססת על התפיסה שיכולת הכתיבה מורכבת מכמה מרכיבים נפרדים, שיש להעריך באופן בלתי תלוי. נקידה אנליטית מבוססת על זיהוי הממדים החשובים בכתיבה, אופרציונליזציה שלהם, ומתן משקלות יחסיים לכל אחד מהם.

לאחר שקילת היתרונות והחסרונות של כל שיטה, ובחינת המסקנות מפיתוח מבחן הכתיבה במאל"ו בסוף שנות ה-90, הוחלט להעריך את מטלות הכתיבה במכפ"ל בשיטה האנליטית. לצורך זה גובש מחון אנליטי שכולל שני ממדים: תוכן ולשון.

מיון והכשרת המעריכים

תוספת של מטלה פתוחה למבחן חייבה גיוס והכשרה של מאות מעריכים בעברית, בערבית וביתר שפות הבחינה. בשלב הראשון המועמדים עוברים סינון ראשוני בהתאם להמלצות ולקורות חיים. בשלב השני הם מוזמנים לראיון קצר ולהתנסות מעשית הכוללת כתיבת חיבור והערכה של כמה חיבורים בנושא נתון. על בסיס תוצאות התנסות זו מתקיים סינון שני. כל המעריכים עוברים סדנת הכשרה לפני כל מחזור בחינה. לפני הסדנה נבחרים חיבורים מתוך כלל החיבורים שנכתבו בתגובה למטלה שבה מתמקדת הסדנה, ואלה מנותחים ומוערכים על ידי חברים

מצוות פיתוח הבחינה. היעד המעשי של הסדנה הוא לכייל את הערכותיהם של המעריכים המתלמדים להערכותיהם של חברי צוות הפיתוח. החלק הראשון של הסדנה מוקדש להכשרה כללית. מוצגים בו בפרוטרוט אופי מטלת הכתיבה, הגישה הנכונה להערכה אנליטית, אופן העבודה עם המחונן והיבטים טכניים שונים של העבודה. עיקר הסדנה מוקדש למטלה הספציפית, שאת החיבורים שנכתבו בתגובה לה יעריכו המשתתפים. חלק זה כולל בעיקר עבודה מעשית: הערכה של חיבורים וניתוח שלהם.

הכללת מטלת כתיבה בבחינה הפסיכומטרית - היערכות לוגיסטית ונתונים סטטיסטיים

אבי אללוף

מרכז ארצי לבחינות ולהערכה

הכללת מטלת הכתיבה בבחינה הפסיכומטרית הייתה ועודנה תהליך מורכב. האתגרים העיקריים היו גיוס והכשרת מאות מעריכים בפרק זמן קצר, הגדרת תהליך כולל לטיפול במטלת הכתיבה, פיתוח מערכת ממוחשבת מקיפה לסריקת המטלות, הפצתן למעריכים, ניהול ההערכה והזנת הציונים לבסיס הנתונים, פיתוח שיטות לחישוב ציונים וניהול בקרת איכות על עבודת המעריכים. המערכת הממוחשבת שפותחה לצורך הטיפול במטלת הכתיבה מאפשרת: (1) פיזור המטלות למעריכים לפי הגדרות קבועות מראש, (2) הערכה מול מגוון מסכי מחשב תוך אפשרות רישום הערות והתייעצות שוטפת, (3) ניטור איכות עבודת המעריכים כבר מתחילתה, (4) צפייה בהתקדמות תהליך ההערכה לפי פרמטרים שונים, (5) התעדכנות וטיפול בתקלות בזמן אמת.

מתאם בין ממדים - כל מעריך מספק ציון לכל ממד, תוכן ולשון, לכל מטלה. יוצגו נתונים על המתאם הנצפה בין הממדים בתוך מעריכים, לפי נוסח, בשפות עברית וערבית. המתאמים שהתקבלו גבוהים למדי. **מתאמים בין מעריכים** - מאפשרים לחשב מהימנות בין מעריכים. המהימנות שהתקבלה היא 0.70 - 0.75.

בקרת איכות - בקרה על תהליך ההערכה חשובה כדי לשפר את טיב ההערכה, והיא יכולה להתקיים במשך ההערכה ובסיומה. נתוני בקרת האיכות יכולים לשמש למשוב לשיפור ההערכה וכן במקרים מסוימים לניפוי מעריכים שאינם עומדים בדרישות. הנתונים מבוססים בעיקר על השוואת כל מעריך לשאר המעריכים שהעריכו את אותם החיבורים. ההנחה היא שעל המעריכים להסכים זה עם זה בהערכותיהם. הנתונים העיקריים המשמשים לבקרה על איכות ההערכה הם: (1) ממוצע המתאמים בין ציוני מעריך לבין ציוני המעריכים האחרים, (2) השוואת התפלגות ציוני מעריך לציוני המעריכים האחרים, (3) אחוז המטלות שבהם מעריך נמצא חורג בציון ממעריך שני, ומעריך שלישי נמצא קרוב יותר למעריך השני, (4) המתאם שבין ציוני המעריך לציון בחלק המילולי הסגור של הבחינה.

חישוב ציונים, קשר לבחינה הפסיכומטרית ומהימנות - מאחר שהחיבורים שונים זה מזה ברמת הקושי שלהם, הוחלט לסלם את הציון במטלת הכתיבה לציון בחלק הסגור של התחום המילולי. השימוש בשיטה זו מתבסס על ההנחה שרמת יכולת הכתיבה של קבוצה זהה לרמת החשיבה המילולית, המתבטאת בציוני החלק הסגור. התוצאה של השימוש בשיטה זו היא שהתפלגות הציונים בקבוצה לא תשתנה בעקבות הכנסת מטלת הכתיבה לבחינה הפסיכומטרית. כמו כן לא תימצא השפעה של קושי המטלה על ציוני קבוצה. חושבו **מתאמים עם ציוני הבחינה הפסיכומטרית**: המתאם עם ציוני החלק המילולי הסגור הוא הגבוה ביותר, לאחר מכן המתאם עם האנגלית, ולבסוף, המתאם עם הציון הכמותי. כמו כן, נמצא כי מהימנות הבחינה הפסיכומטרית במתכונתה החדשה נשארה ברמה דומה לזו שבמתכונת הקודמת (0.95), ומכאן נצפה שתוקף הניבוי להצלחה בלימודים גבוהים יישאר כשהיה.

סימפוזיון

פיתוח כלי להערכת תפקוד של תלמידים בעלי צרכים מיוחדים במערכת החינוך בישראל – היבטים פסיכומטריים ותיקוף הכלי

יו"ר: יואל רפ

הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

בספטמבר 2007 מינתה שרת החינוך דאז, פרופ' יולי תמיר, את "הוועדה הציבורית לבחינת מערכת החינוך המיוחד בישראל", בראשות השופטת בדימוס דליה דורנר. ועדה זו הוקמה כדי לבחון את מכלול ההיבטים הכרוכים בשירותי החינוך המיוחד, הן בבתי ספר של החינוך המיוחד והן בשילוב תלמידים עם צרכים מיוחדים בבתי ספר רגילים. בינואר 2009 הגישה הוועדה דוח, ובו פירטה את המלצותיה ואת מסקנותיה. בין השינויים המוצעים המליצה הוועדה שהתקצוב לתלמיד בחינוך המיוחד ייקבע לא רק על פי סוג הלקות/הקושי של התלמיד (כפי שנעשה כיום) אלא גם על פי רמת התפקוד בפועל של התלמיד בבית הספר.

לשם יישום המודל התקציבי החדש ביקש האגף לחינוך מיוחד במשרד החינוך מראמ"ה, הרשות הארצית למדידה והערכה בחינוך, לפתח כלי שימדוד את רמת התפקוד הבית-ספרי של תלמידים עם צרכים מיוחדים. מדד זה אמור לשמש את משרד החינוך כאחד הנתונים שעל פיהם ייקבע גובה התקציב שיקצה המשרד לתלמיד בחינוך המיוחד.

עקרונות פיתוח הכלי

המאפיינים המרכזיים שהוגדרו בוועדת דורנר לכלי המדידה הנדרש:

- כלי אחד גנרי שישמש להערכת תפקוד של תלמידים בעלי לקויות מסוגים שונים ובדרגות חומרה שונות, הלומדים במסגרות חינוכיות שונות.
- התפקוד יתייחס להיבטים בית-ספריים בלבד, הרלוונטיים לתקצוב.
- הגורם המעריך (שישתמש בכלי) יהיה גורם אובייקטיבי ונטול פניות ככל האפשר, שמכיר את התלמיד ואת תפקודו בבית הספר, ולא גורם אבחוני/טיפולי/רפואי חוץ בית-ספרי.
- הערכת התפקוד על פי ממדים שנקבעו מראש.
- ההערכה תתבסס ככל האפשר על התנהגות נצפית ולא על פרשנות של ההתנהגות, סיבותיה ומקורותיה.
- הערכת התנהגות עכשווית בלבד, ולא הערכת הפוטנציאל להתנהגות עתידית.
- תוצרי ההערכה ישמשו ככלי תומך החלטה להליך קביעת התקציב בלבד ולא לאבחון, לקביעת עצם הזכאות לשירותי חינוך מיוחד או/ו להשמה.

עם תחילת פיתוח הכלי הוחלט להוסיף את העקרונות האלה:

- תפוחת ערכה נפרדת לכל שכבת גיל: גן טרום חובה, גן חובה, בי"ס יסודי, חט"ב, חטיבה עליונה ומבוגרים (18-21).
- הערכת התנהגות כפי שבאה לידי ביטוי בפועל, כלומר הכוללת כל תיווך ותמיכה קיימים, ובסביבה שבה מתרחשת ההתנהגות.
- הערכת התנהגות עקבית.
- הערכה באמצעות "רשת גסה" – התייחסות לתפקוד באופן שיאפשר לזהות הבדלים גדולים יחסית בין תלמידים שונים.
- הכלי לא יעריך הישגים לימודיים לפי תכנית לימודים של דרגת כיתה כזאת או אחרת.

הליך פיתוח הכלי להערכת תפקוד של תלמידים בעלי צרכים מיוחדים

לאורה קורטן

מרכז לטכנולוגיה חינוכית (מט"ח)

תהליך הפיתוח של הכלי למדידת התפקוד של תלמידים בעלי צרכים מיוחדים לווה על ידי ועדת היגוי שבה השתתפו מומחים לחינוך המיוחד, מומחים ללקויות שונות ומומחים למדידה והערכה, ונערך על פי השלבים האלה. הועדה ליוותה את צוות הפיתוח ואישרה את הכלי בשלבים הקריטיים בפיתוח.

הגדרת טקסונומיה ← ניסוח היגדים ← ביקורת היגדים ← בדיקה באמצעות "מיני-פיילוט" ← מקצה שיפורים ← בדיקה במחקר תיקוף ← עיצוב וגיבוש סופי של הכלי.

ההצגה הנוכחית מתמקדת בתיאור הכלי (היגדים, סולמות, אופן ההצגה למעריכים וכיו"ב), תהליך פיתוח הטקסונומיה ותיאור הקשיים הייחודיים בפיתוח כלי מדידה זה.

הטקסונומיה הראשונית של התכונה "תפקוד בית-ספרי של תלמידים עם צרכים מיוחדים" גובשה על ידי ועדות ייעודיות שמונו על ידי לשכת המדען הראשי במשרד החינוך. טקסונומיה זו כללה שישה ממדים: תפקוד חברתי, תפקוד רגשי, תפקוד קוגניטיבי-לימודי, תפקוד שפתי, תפקוד בתחום התקשורת ועצמאות תפקודית והתארגנות. צוות הפיתוח גיבש את הנושאים שהרכיבו כל ממד והגדיר את ההיבטים הרלוונטיים להערכה ולמטרת הכלי.

צוות הפיתוח מנה מנהלת פרויקט, שני פסיכולוגים (חינוכי וקליני) ומפקחת בדימוס במסגרות החינוך המיוחד. כל חברי הצוות היו בעלי ניסיון רב באבחון, בטיפול ובניהול מסגרות של ילדים ונוער עם מוגבלויות. נוסף על כך, צורפו לצוות איש הערכה ומדידה, ובשלב המחקרים חוקרת בכירה ומומחית להערכה ומדידה וסטטיסטיקאי. הצוות התייעץ לאורך כל שלבי הפיתוח עם מומחים אקדמיים חיצוניים בתחומים השונים ועם מומחים מתחום החינוך המיוחד, שסייעו בשיפור הניסוח של ההיגדים ובשיוכם לנושאים השונים שהרכיבו את הממדים. כמו כן, תהליך הפיתוח לווה מתחילת דרכו על ידי ועדת היגוי שבה היו חברים אנשי אקדמיה מתחומים שונים (רפואה, חינוך, משפטים, מדידה והערכה) ונציגים ממשרד החינוך. תפקידה של הוועדה היה לדון בסוגיות המרכזיות בנוגע לפיתוח הכלי (לרבות אשרור של עקרונות הכלי ועקרונות פיתוח הכלי), ולשמש גורם המסייע בפרשנות של הנתונים האמפיריים שנאספו באמצעות הכלי.

הכלי בגרסתו הסופית מודד את התפקוד בחמישה ממדים: עצמאות תפקודית והתארגנות, תפקוד בתחום התקשורת, תפקוד רגשי, תפקוד חברתי ותפקוד קוגניטיבי-לימודי ושפתי (ממדים אלה אוחדו בעקבות ממצאי מחקר התיקוף). כל ממד מורכב מנושאים שונים (בין 4-7 נושאים בכל ממד; 32 נושאים בסך הכול), ובכל נושא יש בין 2 ל-13 היגדים – סך הכול 170 היגדים. כלי ההערכה הוא ממוחשב וממלא באמצעות רשת האינטרנט.

כמחצית מההיגדים מנוסחים באופן חיובי, והם משקפים תפקוד "תקין" (למשל, "מפסיק את התנהגותו כאשר הוא מבחין שהוא מפריע לילד אחר"). שאר ההיגדים מנוסחים באופן שלילי והם משקפים "קושי" בתפקוד (למשל, "אומר שלא ילך לו"). המעריך (מורה או איש צוות חינוך המכיר היטב את התלמיד) מתבקש להעריך בכל היגד את התנהגות התלמיד כפי שהיא נצפתה בפועל במשך 30 הימים שקדמו ליום ביצוע ההערכה. ההערכה מתבצעת על אחד משני סולמות ליקרט: סולם תדירות או סולם רמת תפקוד.

הקשיים הייחודיים בפיתוח הכלי: אי ודאות גדולה בנוגע להגדרה האופרטיבית של התכונות המתבקשות, לצד קושי בפיתוח ובתיקוף כלי שאין עבורו קריטריון לתיקוף ושלא פותח כמותו קודם לכן בארץ ובעולם; אי ודאות

בנוגע לשימוש שיעשה בכלי בפועל ובנוגע למודל התקצוב שישתמש בנתונים שיספק הכלי; מילוי על ידי אנשים שאינם מתחום ההערכה והמדידה הפסיכולוגית; הצורך במאות פריטים שיכסו את כל היבטי התפקוד הרלוונטיים; היחסיות של ההערכה במסגרות החינוך השונות (תלמידים משולבים, תלמידים בכיתות חינוך מיוחד בבתי ספר רגילים ותלמידים בבתי ספר של החינוך המיוחד); הערכת תפקוד בתנאים שאינם בהכרח התנאים שבהם התלמיד יתפקד בפועל (מסגרת אחרת, שינוי בתנאי הסיוע/תיווך/סביבת תפקוד); הצורך להשתמש בכלי לצרכים נוספים שלא נועד להם (כגון שימוש לצורכי אבחון, השמה, קביעת זכאות לשירותי חינוך מיוחד ועוד); פיתוח בסביבה שבה קיימת רגישות חברתית עצומה ושבעלי העניין בה חרדים שמא תוצרי הכלי יפגעו בתקציב המוענק לילדיהם.

מחקר התיקוף של הכלי להערכת תפקוד של תלמידים עם צרכים מיוחדים בכיתות א'-ו'

נוגה גוטספלד

מרכז לטכנולוגיה חינוכית (מט"ח)

כחלק מפיתוח הכלי ותיקופו נערך מחקר דו שלבי: השלב האחד במטרה לגבש ולברור את ההיגדים שישמשו בכלי הסופי וכדי לבדוק את תוקף הכלי; והשלב האחר בעיקר לצורך תיקוף הכלי וביסוס ההחלטות שהתקבלו על סמך נתוני המחקר בשלב הראשון.

מטרת המחקר בשלב הראשון הייתה לבחון באורח ראשוני את המאפיינים הפסיכומטריים של הכלי, לבחור את סט הפריטים שייכלל בנוסח הסופי, ולאסוף מידע לצורך תיקוף הכלי על היבטיו השונים, ובכלל זה על תוקף המבנה.

המדגם כלל 589 תלמידים בכיתות א'-ו בבתי ספר דוברי עברית (פיקוח ממלכתי וממלכתי-דתי בלבד) הלומדים בבתי ספר לחינוך מיוחד, בכיתות לחינוך מיוחד בבתי ספר רגילים, או המשולבים בכיתות רגילות בבתי ספר רגילים. התלמידים במחקר היו תלמידים שהוכרו כזכאים לתמיכה מתוקף חוק החינוך המיוחד. בנוסף, המדגם כלל קבוצה של תלמידים ללא לקות, אשר שימשו קבוצת ביקורת.

בשלב מחקרי זה, הכלי כלל 242 פריטים (היגדים) להערכת תפקוד התלמידים, שנחלקו בין הממדים השונים. בשלב הראשון בוצע ניתוח פריטים, ונמצא כי לרוב ההיגדים היו מדדי הבחנה טובים (עם הממד ועם הנושא שאליו הם שייכים) והתפלגות ציונים כמצופה. מדדי המהימנות (עקיבות פנימית) של כל הממדים ושל מרבית הנושאים עמדו באמות המידה המקובלות. על בסיס ניתוח הנתונים הפסיכומטריים של הפריטים נבחרו 170 ההיגדים לכלי הסופי על פי הקריטריונים האלה: (א) ביטול היגדים שאינם עומדים ברמת קשר סטטיסטי מספק עם הנושא שאליו הם שייכים, או שעומדים בקשר חזק יותר עם נושאים אחרים; (ב) ביטול היגדים הנמצאים בקשר חזק מדי עם היגדים אחרים (ביטול עודפות). לאחר צמצום מספר ההיגדים נערכו הבדיקות שנית, ונמצא כי הכלי מתאפיין בערכי מהימנות הולמים, הן בכל ממד והן בכל נושא.

בשלב השני, נבדק תוקף המבנה של הכלי, תוך כדי השוואה בין המבנה התאורטי (הטקסונומיה) כפי שהוגדר בשלב הפיתוח, ובין המבנה האמפירי שהתקבל על פי הנתונים. המבנה נבדק באמצעות ניתוח הקשרים האמפיריים בין הנושאים בכלי, בניתוח גורמים מגשש ובסילום רב-ממדי. ממצאי ניתוחים אלה הראו תמיכה חלקית בטקסונומיה שהוגדרה בשלב הפיתוח, ובעקבותיהם הוחלט לאחד בכלי הסופי שני ממדים: תפקוד קוגניטיבי-לימודי ותפקוד שפתי.

לבסוף, נבדקה היכולת של הכלי להבחין בין תלמידים בעלי לקויות שונות. מבדיקות אלה עלה כי הכלי מצליח לשקף הבדלים בין רמות התפקוד של תלמידים בעלי לקויות שונות, באופן התואם את המוכר לעוסקים בתחום.

המחקר בשלב השני התבצע מספר חודשים לאחר המחקר הראשון. בשלב זה נעשה שימוש בכלי הסופי (5 ממדים ו-170 ההיגדים שנבחרו). המחקר כלל מדגם גדול של תלמידים: 4536 תלמידים בעלי לקויות שונות בדרגות חומרה שונות, הנמצאים במסגרות שונות של החינוך המיוחד, ובכלל זה קבוצה של כ-200 תלמידים שאינם תלמידי חינוך מיוחד אשר שימשו כקבוצת ביקורת.

הממצא העיקרי והחשוב ביותר של מחקר רחב היקף זה הוא כי למרות ההבדלים המשמעותיים בגודל ובהרכב של המדגמים בשני המחקרים - ממצאיו דמו מאוד לאלו שהתקבלו במחקר התיקוף המצומצם. מדדי המהימנות של כל הממדים ושל רוב הנושאים היו גבוהים ועמדו באמות המידה המקובלות, ולרוב ההיגדים נרשמו מדדי הבחנה טובים. כמו כן, ניתוח הגורמים והסילום הרב-ממדי הניבו תוצאות דומות מאוד לתוצאות שהתקבלו במחקר המצומצם, המצביעות על יציבות גדולה במבנה הקשרים בין הנושאים הנמדדים בכלי. גם פרופיל התפקוד של תלמידים עם לקויות שונות היה דומה מאוד לפרופיל שהתקבל במחקר המצומצם.

הדמיון הרב בין הממצאים של שני המחקרים מבסס את תוקף הכלי ומחזק את המסקנות וההחלטות בנוגע להרכב כלי ההערכה שהתקבלו בעקבות המחקר המצומצם שנערך בשלב הראשון.

כלי להערכת תפקוד של תלמידים עם צרכים מיוחדים בחטיבת הביניים

אסטלה מלמד

מרכז לטכנולוגיה חינוכית (מט"ח)

לאחר תום פיתוח הכלי להערכת תפקוד בבית הספר היסודי, החל פיתוח הכלי המיועד לתלמידי חטיבת הביניים. הכלי לתלמידי חטיבת הביניים פותח על פי אותם עקרונות כמו הכלי לתלמידי היסודי, בהתאמות המתבקשות. בתום שלב המיני פיילוט נערך מחקר לגיבוש סופי של הרכב הפריטים שייכלל בנוסח הסופי, ולאיסוף מידע לצורך תיקוף הכלי, לרבות תיקוף הטקסונומיה.

הכלי בשלב מחקרי זה כלל 263 היגדים המתאימים להערכת תפקוד של תלמידים בגילי חטיבת הביניים. כלי זה פותח על בסיס הטקסונומיה של הכלי שיועד לשלב בית הספר היסודי ובתהליך פיתוח דומה. חלק קטן מההיגדים בכלי זה זהים להיגדים שפותחו עבור הכלי לתלמידי החינוך היסודי, חלק אחר מההיגדים מבוססים על היגדים שהופיעו בכלי לחינוך היסודי אך הותאמו לקבוצת הגיל, וחלק אחר מההיגדים פותחו במיוחד עבור כלי זה. המדגם כלל 836 תלמידים בחטיבות ביניים (כיתות ז-ט) בבתי ספר דוברי עברית (פיקוח ממלכתי וממלכתי-דתי בלבד).

ממצאים ראשונים מהמחקר מראים כי בדומה לממצאים עבור הכלי לחינוך היסודי, המאפיינים הפסיכומטריים של הכלי לחטה"ב - מדדי ההבחנה של הפריטים ואומדני המהימנות של הממדים והנושאים - עומדים בסטנדרטים המקובלים. גם במקרה זה הצמצום של מספר ההיגדים לטובת הנוסח הסופי של הכלי מונחה על ידי קריטריונים של התכנסות והבחנה של הקשרים הסטטיסטיים בין הפריטים והנושאים הנכללים בממדים.

בניתוחי המבנה - ניתוח גורמים מגשש וסילום רב-ממדי - התקבלו ממצאים דומים מאוד לאלו שהתקבלו במחקרים על הכלי לחינוך היסודי. על אף הבדלים מסוימים, המבנה הכללי של תחומי התפקוד הנמדדים בשני

הכלים נשמר. ניכר כי למרות ההבדלים בין שתי החטיבות החינוכיות (יסודי וחט"ב) ובתפקוד הנדרש מהתלמידים, ולמרות ההבדלים בין שני הכלים שפותחו (בעיקר בהיגדים המרכיבים אותם), הרי המבנה התאורטי/טקסונומי בבסיס הכלים יציב ועקבי בשתי דרגות הגיל.

הממצאים שנאספו במחקר על הכלי לחטה"ב ובמחקר על הכלי לבית הספר היסודי מתקפים אלו את אלו. הממצאים מציינים תמונה מעודדת בנוגע למהימנות ולתוקף של כלי ההערכה בכללותו, ומספקים עדות לכך שממדי התפקוד הבית-ספרית כפי שהוגדרו בתהליך פיתוח הטקסונומיה של הכלי הראשון (לכיתות א-ו') הם אוניברסליים דיים (לפחות מעבר לחטיבות גיל שנחקרו עד כה) ועקביים דיים כדי שאפשר יהיה להעריך בעזרתם את התפקוד של תלמידים בחינוך המיוחד, ולסייע ביישום ההמלצות של ועדת דורנר כנדרש.

טעות מדידה בשיפוטי חיבורים: בדיקה אמפירית של "כשל המערך השלישי"

יואב כהן

מרכז ארצי לבחינות ולהערכה

בהערכה של תוצרי כתיבה (חיבורים) בבחינות עתירות סיכון (high stakes tests) נהוג להשתמש בשני מעריכים בלתי תלויים השופטים את איכות החיבור, והציון הסופי נקבע על-ידי מיצוע שתי ההערכות. יש מערכות מבחנים שבהן מיושם הכלל הבא: במקרים של אי-הסכמה ניכרת בין שני המעריכים (וכל מערכת מגדירה לעצמה מהי אי-הסכמה ניכרת). ניתנת גם הערכה שלישית, והציון הסופי נקבע על-ידי מיצוע הערכה זו עם ההערכה הקרובה לה ביותר מבין שתי ההערכות המקוריות (ר' למשל Donlon 1984, p.90, GMAC, 2001, Abramson, 2004). הרציונל בבסיס נהל זה – נהל "החלפת ההערכה הקיצונית" – הוא שקרבה בין הערכות מבטאת מהימנות גבוהה של הציון, ועל כן אם שני מעריכים אינם מסכימים זה עם זה קרוב לוודאי שאחד מהם טועה במידה ניכרת ועל-כן יש להחליפו במערך אחר (לעיתים לצורך הערכה שלישית נעזרים במערך שנחשב למנוסה או מיומן יותר).

במחקר קודם שהתבסס על סדרת סימולציות מונטה קרלו, התבררה תופעה מפתיעה, והיא שנהל "החלפת ההערכה הקיצונית" – שמטרתו לנטר ולשמור על איכותן של ההערכות – במקום לגרום לצמצום בטעות המדידה, גורם דווקא להגדלתה (בהקשר הנוכחי טעות המדידה מוגדרת במונחי תורת המבחנים הקלאסית, כפער בין הציון הנצפה לבין הציון האמיתי). קראנו לתופעה זו בשם "כשל המערך השלישי".

מאחר ונהל ההחלפה הוא מושרש ביותר ועל פניו נראה כמוצדק, ומאחר וכשל המערך השלישי הוכח עד כה רק בסימולציות הנשענות על הנחות סטטיסטיות מסוימות, יש חשיבות לבדיקת קיומו האמפירי של הכשל בתנאי אמת. בסימולציות היה מידע מושלם אודות "הציון האמיתי" של כל חיבור, אך במציאות אין בדרך כלל מידע כזה כי אם רק מידע על הציון הנצפה.

בעזרת ריבוי הערכות לכל חיבור ניתן לאמוד בצורה טובה את הציון האמיתי ועל כן גם לבדוק האם כשל המערך השלישי אכן מתקיים גם בתנאי אמת. הבדיקה נערכה ע"י ניתוחן של הערכות שניתנו ל-500 חיבורים וכל חיבור הוערך 13-15 פעמים. מתברר כי המסקנות שהוסקו מתוך הסימולציות הן אכן תקפות עבור נתונים אמפיריים, כלומר כשל המערך השלישי מתקיים לא רק בסימולציה אלא גם בסיטואציה אותנטית של הערכה. נהל "החלפת ההערכה הקיצונית" במקרים שבהם יש פער גדול בין שני המעריכים אכן מגדיל את טעות המדידה במקום להקטינה. טעות המדידה שלאחר נהל ההחלפה של ההערכה הקיצונית, גדולה בהרבה מטעות המדידה המקורית כאשר הנהל מיושם במקרים של פערים גדולים מאד בין שתי ההערכות המקוריות. בדומה לתוצאות הסימולציה, נהל החלפת ההערכה הקיצונית מצמצם את טעות המדידה דווקא במקרים שבהם הפער בין ההערכות הוא קטן.

לתוצאות המחקר השלכות תיאורטיות ומעשיות גם יחד. מבחינה תיאורטית מתקבל אישוש לטענה שמתוך ההפרש בין שתי טעויות מדידה (זהו ההפרש בין שתי הערכות) לא ניתן ללמוד על סכום טעויות המדידה. מבחינה מעשית ההשלכות הן ברורות: תוכנית מבחנים אשר מיישמת את נהל ההחלפה הקיצונית, קרוב לוודאי – ולפחות בתנאים שנבדקו במחקר זה – משקיעה משאבים בהגדלת טעות המדידה במקום בהקטנתה.

הטיה אופטימית בשיפוט מתאם פירסון

אייל גמליאל, רואי גור

המרכז האקדמי רופין

נניח שמקדם המתאם של פירסון בין ציוני הפסיכומטרי וציוני ההצלחה בלימודים של קבוצת סטודנטים גדולה הוא 0.5. מה יהיה להערכתם ממוצע ההפרשים המוחלטים בין שני ציוני התקן (בפסיכומטרי ובהצלחה בלימודים) של כל הסטודנטים? ממצאי מחקר זה מנבאים שההערכה שלכם תהיה הערכה חסרה: בהיעדר מודלים נורמטיביים זמינים, אנשים ברמות מומחיות שונות בסטטיסטיקה משתמשים בחשיבה היוריסטית דומה, אשר גורמת להם להעריך באופן חסר מאוד את הפער הממוצע בין כל שני ציוני תקן. כך, בדוגמה שלעיל אנשים יעריכו את הפער הממוצע בין כל שני ציוני תקן בחצי סטיית תקן, בעוד שלמעשה הפער עומד על 0.8 סטיית תקן. הטיה זו היא משמעותית, שכן בעוד שפער של חצי סטיית תקן יאופיין כפער בינוני, הרי שפער של 0.8 סטיית תקן הוא פער גדול (על-פי הקטגוריות שטבע כהן, Cohen, 1988).

המחקר מציג את הקשר הצפוי בין מתאם פירסון ובין ממוצע ההפרשים המוחלטים בין ציוני תקן תחת הנחות מסוימות, אומד את הקשר האמפירי בין משתנים אלה, ומציג את השיפוט הסובייקטיבי של אנשים לגבי הקשר בין שני משתנים אלה.

בהנחה שמשנתנה מתפלג נורמאלית, היחס בין ממוצע ההפרשים המוחלטים בין ציוני תקן (Mean Absolute Differences between paired standard scores - Z-scores, להלן MADZ) לבין מקדם המתאם של פירסון (r_p) הוא (בקירוב): $MADZ = 1.13 - 0.68 \times r_p$

קשר זה אומת בשלושה בסיסי נתונים אמפיריים: (1) שאלוני דיווח עצמי; (2) ציוני קורסים של סטודנטים בקורסי חובה; (3) ציוני קבלה והצלחה בלימודים של סטודנטים.

בדיקה של שיפוט של סטודנטים לתואר ראשון במדעי החברה, בוגרי קורס סטטיסטיקה תיאורית, וכן של סטודנטים לתואר שני במדעי החברה, כמו גם אנשים העוסקים במחקר, העלתה שבקרב כל אחת מהקבוצות, כ-70% מהמשתתפים שפטו את הקשר בין MADZ למקדם המתאם לנוסחה: $MADZ = 1 - r_p$

שיפוטם סובייקטיביים אלה מוטים באופן משמעותי לעומת הקשר הנכון תיאורטית וזה שנמצא אמפירית. השיפוט הסובייקטיבי מעריך באופן חסר מאוד את MADZ עבור כל ערכי מקדם המתאם של פירסון. לדוגמה, עבור מקדם מתאם של 0.7, אנשים מעריכים את MADZ בערך של 0.3 (פחות משליש סטיית תקן), בעוד שערכו האמיתי הוא כמעט שני שלישי (0.65) סטיית תקן.

כך, עבור ערכי מתאם פירסון חיוביים השיפוט של אנשים הוא אופטימי מדי: הם סבורים כי מקדם מתאם כלשהו מצביע על מרחקים קטנים הרבה יותר בממוצע בין כל שני ציוני תקן של אותה תצפית. הטיה זו בשיפוט נובעת, ככל הנראה, מחשיבה היוריסטית שאליה אנשים פונים בהיעדר מודלים נורמטיביים זמינים. להטיה זו תורמים גם הוראת הסטטיסטיקה והיעדר דיון בסוגיה בספרות המקצועית. להטיה זו יש יתרונות גם לצרכני הספרות המדעית וגם לחוקרים המפרסמים – כולם גם יחד "מרוויחים" פסיכולוגית מכך שהם סבורים שמתאמים חיוביים מצביעים על קשר גבוה יותר מהקשר הקיים בפועל. רווח תועלתני נוסף יש לחוקרים, אשר ממצאי מחקריהם נשפטים כמצביעים על קשרים סובייקטיביים גבוהים מאלה הקיימים בפועל.

יישום גישות שונות לחישוב רווח סמך ולאמידת מתאם מתוקן לקיצוץ תחום

תמר קנת-כהן

מרכז ארצי לבחינות ולהערכה

אילה כהן, אטי דובא, טטיאנה אומנסקי

הטכניון

במחקרי תוקף להערכת יעילות כלי מיון מקובל להשתמש במקדם המתאם בין כלי המיון לקריטריון. מחקרים כאלה מתבצעים לרוב במדגמים שעברו ברירה מוקדמת, הגורמת לצמצום בטווח ציוני החזאים והקריטריון ולהקטנת מקדמי המתאם ביניהם. במחקר זה התמקדנו בקשר שבין החזאי שלפיו נעשתה הברירה והקריטריון. מקובל להתייחס למצב כזה כאל Thorndike's Case 2 (Thorndike, 1949), ועבורו מוצעת נוסחה מתאימה לתיקון המתאם עבור קיצוץ התחום.

ללא הערכת שגיאת אומד המתאם המתוקן, לא סביר לקבל החלטות מעשיות. לצרכים מעשיים יש צורך ברווח סמך עבור המתאם (American Psychological Association, 2010). מאחר שלא ניתן לפתח באופן אנליטי נוסחה לסטיית התקן של האומד למתאם המתוקן, לא ניתן היה בעבר הרחוק לבנות רווח סמך מדויק. כיום, תודות להתפתחות בכוח המחשוב והשימוש הרב בשיטות של סימולציה, ניתן לשפר משמעותית את שיטות ההסקה הסטטיסטית. בפרט, ניתן על ידי שיטות כמו ה-Bootstrap לבנות רווח סמך מדויק למדי למתאם המתוקן.

מטרת המחקר הייתה לחשב ולהשוות חמש שיטות לחישוב רווחי סמך למתאם המתוקן. בארבע השיטות הראשונות (Chan & Chan, 2004) חושב האומד הנקודתי לפי הנוסחה המקובלת (Thorndike, 1949), והרווחים עצמם התבססו על חישובים תאורטיים אסימפטוטיים או על מדגמי Bootstrap. בשיטה החמישית חושב האומד הנקודתי באמצעות השלמת חלק הנתונים החסר (imputation), ובהתאם נבנה הרווח סביבו. אומד נקודתי בשיטת ההשלמה הוצע בעבר (Wiberg & Sundström, 2009), והוספת רווח הסמך היא חידוש של המחקר הנוכחי.

רווחי הסמך חושבו בשתי אוכלוסיות:

1. עבור נתוני הקבלה של הטכניון לשנת 2009. החזאי היה סכם הקבלה (המורכב מציוני הבגרות ומציוני הבחינה הפסיכומטרית) והקריטריון היה ממוצע ציוני שנה א'. לכל פקולטה חושבו האומדים הנקודתיים וחמישה רווחי הסמך.
2. עבור אוכלוסייה סימולטיבית. מאחר שעבור נתוני הטכניון לא ידוע הערך האמתי של מקדם המתאם, לא ניתן ללמוד מהם על טיב השיטות השונות. לכן, נעשו ארבע סימולציות, שבהן יצרנו מדגמים מהתפלגות ידועה. כל סימולציה התבססה על מדגם של 400 תצפיות מהתפלגות נורמלית דו-ממדית (עם "חזאי" ו"קריטריון"). ארבע הסימולציות נבדלו זו מזו בערכו של הפרמטר (מקדם המתאם) ו/או ביחס הברירה. בכל סימולציה חושבו האומדים וחמישה רווחי הסמך.

הממצאים מנתוני הטכניון ממחישים את הצורך להיזהר בהסקת מסקנות על בסיס האומד הנקודתי בלבד, בפרט במדגמים הקטנים וכאשר הערך של המתאם הלא מתוקן נמוך. במקרים אלה רווח הסמך מכסה למעשה כמעט את כל התחום התיאורטי [-1, 1] ולכן אין הרבה משמעות לתוצאה הנקודתית. נתוני הסימולציות מעידים שבכל השיטות ובכל הדוגמאות להוציא מקרה אחד, הערך הנכון של הפרמטר כלול ברווח. הרווח המתקבל בשיטת השלמת נתונים הוא לרוב הצר ביותר ובמובן זה עדיף. בימים אלה מתבצע מחקר רחב היקף כדי לאושש את הממצא בדבר עדיפותה של שיטה זו וכן כדי לבדוק אם היא נותנת כיסוי נכון, בהתאם לרמת הסמך הנדרשת. אחת ממטרות המחקר הייתה תיעוד תוכנה המתאימה לאמידה המרווחית. קוד שנכתב בתוכנת SAS במסגרת המחקר מאפשר לחשב רווחי סמך לפי כל אחת מחמש השיטות.

כשיחלו

כשיחלו

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

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

מאובות ופצלות
בנושא האודף הישראית אפסיכומטריקה
אפסר אפסנות א:

שרה שפירא, מרכז ארצי לבחינות ולהערכה sara@nite.org.il

אתר האגודה הישראלית לפסיכומטריקה www.ispa.org.il