

הכינוס ה-15 של האגודה הישראלית לפסיכומטריקה

כ"ג בשבט תשע"ט 29.1.2019



מרכז שלוחה, דרך שלוחה ירושלים



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

התכנסות, רישום וארוחת בוקר	8:15 – 9:15	
דברי פתיחה וברכות יו"ר אפי: אייל גמליאל; יו"ר אפי הנכנס: צור קרליץ יו"ר הכינוס: עינת נוטע-קורן ואסטלה מלמד	9:15 – 11:00	
מושב א' כלי מיון ועדויות לתקפותם יו"ר: ענת בן-סימון		אודיטוריום
<ul style="list-style-type: none"> ■ תוקף הניבוי של כלי מיון ללימודי תואר ראשון במוסדות לא-אוניברסיטאיים אייל גמליאל, יונתן סער, תמר קנת-כהן, גיל גולדצויג 		
<ul style="list-style-type: none"> ■ עולים על המסיל"ה: מיון סטודנטים ייעודי ללימודי הוראה ג'ודי גולדנברג 		
<ul style="list-style-type: none"> ■ סיווג ממוחשב של חומרי קריאה לימודיים לדרגות כיתה צור קרליץ, ענת בן סימון, ענת בר-סימן-טוב, יונה דולב, אפי לוי, יואב כהן 		
הפסקת קפה	11:00 – 11:20	
מושב ב' 1 (מושב מקביל) אתגרים ודילמות במבחנים יו"ר: חגית גליקמן	11:20 – 12:40	אולם יהלום
<ul style="list-style-type: none"> ■ להכשיל או לא להכשיל: בחינות רישוי עריכת דין חנן גולדשמידט, אייל גמליאל, גיל גולדצויג 		
<ul style="list-style-type: none"> ■ השוואה בין רכיבי ציון הבגרות: ציון שנתי וציון בחינה יונתן סער, נורית ליפשוטט 		
<ul style="list-style-type: none"> ■ תוקף מטלת הכתיבה של הבחינה הפסיכומטרית תמר קנת-כהן, יונתן סער 		
<ul style="list-style-type: none"> ■ איך מודדים אוריינות מתמטית כחלק מ"היעדים להתפתחות בת-קיימא" של האו"ם? אתגרים בפסיכומטריקה ופוליטיקה של מדידה השוואתית עדו גל 		
מושב ב' 2 (מושב מקביל) סימפוזיון: הערכת הבעה בכתב במבחנים רחבי היקף בישראל יו"ר: יואל רפ, ספי פומפיאן		אודיטוריום
<ul style="list-style-type: none"> ■ מחונן הוליסטי או אנליטי? מהימנות הערכת כתיבה במבחן עברית לדוברי ערבית בכיתות ו' (עלד"ע) אלה שפירא, ליאת אריאלי-מאור, ענבל רון-קפלן 		
<ul style="list-style-type: none"> ■ אמידת מהימנות בין מעריכים – המקרה של מטלת הכתיבה הממזגת (הבעה בכתב) במבחן הבגרות הממוחשב בועז רזנבאום, רוני ברצלבסקי, יואל רפ 		
<ul style="list-style-type: none"> ■ האם וכיצד משפיעה הכשרת מעריכים על ההערכה של מטלות כתיבה במכפ"ל נעמי גפני, ספי פומפיאן, אבי אללוף, ג'ני שמולביץ, מיכל חטבי 		
<ul style="list-style-type: none"> ■ מדדי איכות מרכזיים של מערכי מטלת הכתיבה במכפ"ל – תיאור וניתוח רב שנתי 2013 – 2017 אבי אללוף, מרינה פרונטון, האשם ניג'ם 		

ארוחת צהריים	14:00 – 12:40	
מושב ג' 1 (מושב מקביל) שולחנות עגולים:	15:20 – 14:00	
<ul style="list-style-type: none"> ■ כיצד "לימודי התעודה בפסיכומטריקה" יקדמו את הפסיכומטריקה בישראל? אבי אללוף, צור קרליץ, ניצן פרידמן ■ כשבינה מלאכותית פוגשת פסיכומטריקה במגדל בבל ריטה סבר 		אולם יהלום
מושב ג' 2 (מושב מקביל) כישורים ודרכים לחידוד מדידתם		
יו"ר: ענת קדם		
<ul style="list-style-type: none"> ■ פיתוח סולם ציונים למבחן 2 בסוללת וודקוק-ג'ונסון 3 יואל רפ, אימאן עוואדיה, אלה שפירא, פרנקלין זרומב, אסטלה מלמד 		
<ul style="list-style-type: none"> ■ קבוצות הטרוגניות חכמות יותר: עדות ממחקר תאומים מאיר ברנרון, ליאור אברמזון, אילן יניב, אריאל כנפו-נעם 		
<ul style="list-style-type: none"> ■ קרבה פסיכולוגית משפרת ביצועים קוגניטיביים: עדויות מהבחינה הפסיכומטרית 		אודיטוריום
ברית הדר, ספי פומפיאן, צור קרליץ, נירה ליברמן		
<ul style="list-style-type: none"> ■ הערכה רב-ממדית של כדורגלני נוער, עם התייחסות לעמדת המשחק אורן אקהויז, גיל גולדצויג 		
אסיפת חברי אפי	16:00 – 15:20	

תוקף הניבוי של כלי מיון ללימודי תואר ראשון במוסדות לא-אוניברסיטאיים*

אייל גמליאל

המרכז האקדמי רופין

יונתן סער, תמר קנת-כהן
מרכז ארצי לבחינות ולהערכה (מאל"ו)

גיל גולדצויג

מכללה אקדמית תל אביב-יפו

מחקרי תיקוף של כלי המיון ללימודי תואר ראשון – ממוצע הבגרות והציון הפסיכומטרי, נעשים בישראל באופן שוטף לגבי סטודנטים באוניברסיטאות ולא נעשים לגבי סטודנטים במוסדות לימוד מוכרים שאינם אוניברסיטאות (להלן: מכללות). היעדרם של מחקרי תיקוף של כלי המיון במכללות הוא מפתיע, שכן בשנים האחרונות שיעור הלומדים לתואר ראשון במכללות אקדמיות הוא כפול בערך מזה של הלומדים באוניברסיטאות. בהיעדר נתונים על תוקף כלי המיון, ישנן מכללות שמשתמשות בכלי מיון מגוונים שתקפותם לא קיבלה ביסוס מחקרי, ובכלל זאת בחינות כושר שאינן הבחינה הפסיכומטרית, ראינות קבלה, ציוני מכינות כתחליף לממוצע בגרות, ציוני מכינות כתחליף לציון הפסיכומטרי, ועוד.

המטרה המרכזית של מחקר זה היא לבדוק את התוקף של כלי המיון הנפוצים – ממוצע הבגרות, הציון הפסיכומטרי וציון הסכם (שהוא שילוב של שניהם) – בניבוי ההצלחה (ממוצע ציוני התואר) בלימודי התואר הראשון בקרב סטודנטים במכללות. הנתונים למחקר זה התקבלו מחדר המחקר של הלשכה המרכזית לסטטיסטיקה. הנתונים היו של שנתוני לימוד שסיימו לימודים ב-32 מכללות בין השנים 2011 ל-2015. מספר התצפיות עליהם התבססו העיבודים היה 48,687. יחידת עיבוד הוגדרה כחוג לימודים במחזור מסוים במכללה מסוימת. הממצאים הראו שמעבר לכל יחידות העיבוד, מקדמי תוקף הניבוי של כלי המיון את ההצלחה בלימודי תואר ראשון במכללות (מתוקנים לקיצוץ תחום) התקרבו לערכים שמתקבלים באוניברסיטאות: 0.33 עבור הבגרות (לעומת 0.36 באוניברסיטאות), 0.35 עבור הציון הפסיכומטרי (לעומת 0.42), ו-0.43 עבור ציון הסכם (לעומת 0.45). בדיקה לפי תחומי לימוד העלתה שמקדמי התוקף בארבעה תחומים (מדעי הרוח ואומנויות, משפטים, הנדסה ואדריכלות, ומדעים) גבוהים יותר באוניברסיטאות לעומת המכללות, ודפוס זה מתהפך במדעי החברה ומקצועות העזר הרפואיים שם מקדמי התוקף במכללות גבוהים יותר. על מנת להעניק משנה תוקף לממצאים נבדקו מקדמי התוקף בשתי מכללות גדולות שנתונים מלאים שלהם הועמדו לרשות המחקר, ואלה הניבו ממצאים דומים לאלה שהתקבלו על בסיס הנתונים הזמינים בלמ"ס.

ממצאי מחקר זה מראים שתוקף כלי המיון בניבוי הצלחה בלימודי תואר ראשון במכללות גבוה כמעט כמו זה שקיים לגבי לימודי תואר ראשון באוניברסיטאות. בניגוד לנתוני הקבלה וההצלחה בלימודי תואר ראשון של סטודנטים באוניברסיטאות הזמינים במאל"ו, אין בנמצא גוף שמרכז נתונים דומים של סטודנטים במכללות. הגוף היחיד שמרכז נתונים אלה הוא הלמ"ס, אך הנתונים המצויים שם הם לרוב חלקיים וחסרים. מחקרי התיקוף חשובים על מנת לסייע למקבלי ההחלטות בנוגע לשימוש בכלי המיון ללימודי תואר ראשון, ובהיעדרם, ישנן מכללות שמשתמשות בכלי מיון מגוונים, למרות שתקפותם לא קיבלה ביסוס מחקרי משכנע. מאגר מידע ארצי של נתוני קבלה והצלחה בלימודי תואר ראשון של סטודנטים במכללות יאפשר לבחון את תוקף הניבוי של הכלים המגוונים לעומת ממוצע הבגרות והציון הפסיכומטרי ולקבל החלטות מושכלות בבחירתם.

* המחקר מומן על ידי קרן המחקר של המרכז הארצי לבחינות ולהערכה. אנו מודים לניר איצקוביץ על הסיוע באיסוף הנתונים ובעיבודם.

עולים על המסיל"ה: מיון סטודנטים ייעודי ללימודי הוראה

ג'ודי גולדנברג
האגף להכשרת עובדי הוראה במשרד החינוך, ומכון מופ"ת

אחת האקסיומות הידועות בתחום החינוך היא שרמת הלמידה של תלמידי בית הספר תלויה במידה רבה באיכות המורים. במהלך השנים האחרונות נערכת עבודה במשרד החינוך להעלאת רמת המורים במערכת החינוך בישראל. בין שאר הפרויקטים שמתקיימים לצורך מימוש מטרה זו, נעשית עבודה יסודית לשיפור ההתאמה של המועמדים ללימוד הוראה למקצוע המורה. לקראת שנת הלימודים תשע"ז, לאחר שלבים רבים של פיתוח, נכנסה לפעולה מערכת אבחון סטנדרטית להערכת ההתאמה ללימודי הוראה המשלבת בתוכה אספקטים בינאישיים, מוטיבציוניים, ואישיותיים. מערכת זו נקראת **יום מסיל"ה (מיון סטודנטים ייעודי ללימודי הוראה)**.

במבחני יום מסיל"ה המועמדים מוערכים באמצעות ארבעה סוגי כלים: מבחנים ושאלונים למדידת התאמה להוראה, מבדק קבוצתי הכולל שלושה תרגילים בינאישיים, ראיון אישי מעמיק מובנה למחצה, והערכת עמיתים. המועמדים מוערכים בכל הפעילויות ע"י צוות של שני מעריכים בלתי תלויים (פסיכולוג תעסוקתי ומדריך פדגוגי מסגל מוסדות ההוראה) המתחלפים בין התרגילים לראיונות.

העברת מבחני יום המסילה החלה בשנת 2015 בפילוט ראשוני, ובשנת 2016 עבר לשלב תפעולי כאשר ציוני יום המסיל"ה משפיעים על הקבלה ללימודי הוראה או על הקבלה לתכניות ייחודיות במשרד החינוך. נכון להיום נבחנו כ-2000 מועמדים ללימודי הוראה במבחני מסילה.

לאחרונה נערך מחקר לבחון את תוקף הניבוי של ציוני יום המסיל"ה למול קריטריון של הצלחה בלימודי הוראה, וזאת בהשוואה לציונים הקוגניטיביים המשמשים למיון ללימודי הוראה היום (ממוצע ציוני הבגרות והציון הפסיכומטרי). אוכלוסיית המחקר כללה את הסטודנטים שהשתתפו בפילוט הראשון לבניית יום המסיל"ה ואשר סיימו את שנת לימודיהם השנייה. הקריטריונים להצלחה בלימודים הוגדרו כ: ממוצע ציונים בשנה א' ובשנה ב' בלימודים, הערכות המדריכים במהלך העבודה המעשית, חוות דעת ייעודית שמולאה על ידי המדריכים הפדגוגיים, וכן נתונים לגבי שרידות או נשירה מהלימודים.

ממצאי המחקר מראים כי מבחני מסיל"ה מצליחים לנבא גם את ההצלחה בלימודים במידה בינונית-גבוהה ומובהקת סטטיסטית ($r=0.28$), וגם את הערכת ההוראה בפועל בעבודה המעשית ($r=0.29-0.47$) בדיסציפלינה הראשונה והשנייה). בנוסף, נמצא קשר בין ציוני מסיל"ה לבין נשירה מהלימודים ($r=-0.28$).

השוואת הממצאים הללו לגבי ציוני מסיל"ה לממצאים שנתקבלו לגבי כלי המיון הקיימים מראה כי כלי המיון הקיימים (בגרות ופסיכוכיני במשולב) מצליחים אף הם לנבא הצלחה בלימודים, אך הם אינם עומדים בקשר עם ההוראה בפועל בעבודה מעשית או עם נתוני הנשירה מהוראה.

מחקר תוקף זה הינו ראשוני ומבוסס על ציוני מסיל"ה שהופקו במסגרת הפילוט ואשר עברו מאז שינויים ותיקונים. במהלך השנים הבאות צוות פיתוח מסיל"ה ימשיך לעקוב אחר מדגם מחקר זה וכן אחר המחזורים התפעוליים של מסיל"ה. יתווספו קריטריונים בשלבים מתקדמים יותר, כולל סיום הלימודים, כניסה והתמחות בהוראה, וכן שרידות בהוראה. אנו מקווים כי הטמעת יום מסיל"ה כתהליך מקצועי וסטנדרטי למיון מועמדים ללימודי הוראה תביא לקידום ושיפור איכות הסטודנטים הלומדים הוראה. אנו מאמינים שבכוחו להשפיע גם על תפיסת לימודי הוראה בעיני המועמדים הפוטנציאליים למקצוע ועל המעמד של הלימודים הללו בעיני הציבור בכלל.

סיווג ממוחשב של חומרי קריאה לימודיים לדרגות כיתה

צור קרליץ, ענת בן סימון, ענת בר-סימן-טוב, יונה דולב, אפי לוי, יואב כהן
מרכז ארצי לבחינות ולהערכה (מאל"ז)

במאה השנים האחרונות חוקרים מנסים לאמוד את רמת הקושי של טקסטים באמצעות משוואות קריאות המבוססות על מאפייני הטקסט. לאחרונה, ההתפתחות של תחום עיבוד השפה הטבעית, (NLP) אפשרה פיתוח מאפייני טקסט מתוחכמים יותר מבעבר. רוב משוואות הקריאות פותחו עבור טקסטים בשפה האנגלית, ולמיטב ידיעתנו, לא פותחו משוואות קריאות עבור טקסטים בעברית. מטרת המחקר הנוכחי היא לפתח משוואות קריאות לסיווג חומרי קריאה לימודיים בעברית לדרגות כיתה, תוך שימוש במאפייני טקסט מבוססי NLP.

פיתוח משוואות הקריאות התבסס על טקסטים שנלקחו מ-26 ספרי לימוד וממקורות נוספים המיועדים לכיתות ב' עד ח', וכן על קטעי הבנת הנקרא מהבחינה הפסיכומטרית. בסך הכול כלל הקורפוס במחקר זה 775 יחידות טקסט (בנות כ-500 מילה כל אחת). הקורפוס נותח ניתוח מורפולוגי ותחבירי באמצעות מערכת ה-NiteRater (2018).

עבור טקסט נתון, מערכת ה-NiteRater מפיקה מעל 250 מאפיינים שיכולים לשמש לבניית משוואות קריאות. הקושי העיקרי בפיתוח המשוואות נובע מהשונות הנמוכה במאפייני טקסט בין דרגות כיתה סמוכות. על כן,

בשלב הראשון של המחקר נופו מאפיינים לא רלוונטיים מבחינה תיאורטית מתוך הסט המלא, ונותרו 120 מאפיינים בלבד. בשלב הבא נבחרו שלושה מקבצים של מאפיינים על בסיס שיקולים תיאורטיים (רלוונטיות למשימת הסיווג) ואמפיריים (גובה המתאם עם דרגת הכיתה). המקבצים כללו 57, 36 ו-23 מאפייני טקסט, ונבנו באופן היררכי כך שכל מקבץ כלל את כל המאפיינים של המקבץ הקודם וכן מאפיינים נוספים.

לכל מקבץ מאפיינים פותחו משוואות קריאות באמצעות שלוש שיטות: רגרסיה ליניארית, רגרסיה לוגיסטית ואלגוריתם SVM. כדי להעריך את מידת ההצלחה של מקבצי המאפיינים בניבוי דרגת הכיתה של טקסטים, נבנו שני מקבצים נוספים ששימשו בסיס להשוואה: המקבץ הראשון כלל את שני המאפיינים המקובלים ביותר במשוואות קריאות: אורך ממוצע של משפט ואורך ממוצע של מילה. ואילו המקבץ השני כלל את כל 120 המאפיינים שזוהו בשלב הראשון.

775 יחידות הטקסט שהופקו מהקורפוס לעיל שימשו לבדיקת טיב הניבוי בשיטת תיקוף צולב מסוג k-fold עם 10 חזרות. חישוב דיוק הניבוי התבצע מעבר ל-10 החזרות.

תוצאות התיקוף הצולב הראו יתרון ברור למקבץ המבוסס על 36 המאפיינים ולשיטת ה-SVM. לאחר שינויים קלים במקבץ זה הורכבה משוואת קריאות סופית על בסיס 34 מאפייני טקסט בלבד. משוואת הקריאות הסופית הצליחה לנבא במדויק את דרגת הכיתה שממנה נלקח הטקסט עבור 55% מהטקסטים, וב-31% טקסטים נוספים הניבוי סטה בדרגת כיתה אחת בלבד מעל או מתחת לדרגת הכיתה האמתית.

בעזרת ניתוח גורמים ושיקולים לשוניים קובצו המאפיינים במקבץ הסופי לארבעה גורמי שפה תיאורטיים: מורכבות תחבירית, עושר לשוני, דחיסות תוכנית וסגנון. הקיבוץ לגורמים מאפשר לאתר בצורה נוחה את היסודות המרכזיים המשפיעים על קושי הטקסט הנתון.

לבסוף, פותחה אפליקציה - מערכת מקוונת המאפשרת למפתחי חומרי לימוד לקבל אומדן אפריורי של רמת הקושי של טקסט לימודי ושל דרגת הכיתה לה הוא מתאים. תוצאות האומדן מוצגות באמצעות הסתברויות הסיווג של הטקסט לדרגות הכיתה. כמו כן מציג הניתוח את דרגות הכיתה שהטקסט מתאים להן עבור כל אחד מארבעת הגורמים התיאורטיים שנגזרו מ-34 המאפיינים.

להכשיל או לא להכשיל: בחינות רישוי עריכת דין

חנן גולדשמידט
המרכז לטכנולוגיה חינוכית (מטח)

אייל גמליאל
המרכז האקדמי רופין

גיל גולדצויג
מכללה אקדמית תל אביב-יפו

בשנים האחרונות נשמעות יותר ויותר טענות שבחינות ההסמכה של לשכת עורכי הדין – הבחינות לקבלת רישיון עריכת דין בישראל – גורמות לכישלון במכוון של הסטודנטים, וזאת על ידי העלאת דרגת הקושי של השאלות בבחינה. טענות אלה נשמעות עוד יותר מאז שינתה הלשכה את מבנה הבחינות, החל ממועד סתיו בשנת 2017. מטרתו של מחקר זה היא לנסות ולהשיב באופן מדעי על השאלה – האם שיעורי הנכשלים הגבוהים יחסית במועדים האחרונים של הבחינה נובעים מסיבות הקשורות לדרגת הקושי של הבחינה, או מסיבות הקשורות בשינוי של הרכב הנבחנים, ובפרט לגבי השנה האחרונה – שנת השינוי בבחינה. אל מול אלה שמסבירים את העלייה בשיעורי הנכשלים בקרב הנבחנים בשנים האחרונות בשל עלייה בדרגת הקושי של הבחינה, יש המסבירים זאת בשינוי שחל בהרכב קבוצת הנבחנים בשנים האחרונות: יותר ויותר מוסדות להשכלה גבוהה בהם רמת ההכשרה נופלת מזו שקיימת במוסדות הוותיקים, מצטרפים לרשימת אלה שמסמיכים בוגרים בתחום המשפטים.

הדרך בה נבקש להבחין מבחינה מתודולוגית בין שני ההסברים לעליית שיעורי הנכשלים – דרגת הקושי של הבחינה או רמת המוכנות של הניגשים אליה, תהיה באמצעות בחינה של אחוז העוברים את בחינות ההסמכה של הלשכה בקבוצה של עשרת המוסדות שמסמיכים בוגרים למקצוע המשפטים מזה למעלה מעשור. אם דרגת הקושי של הבחינה עולה בשנים האחרונות נצפה לראות ירידה בשיעור העוברים בקרב בוגרי עשרת המוסדות כולם. לעומת זאת, אם ההסבר לירידה בשיעורי העוברים נעוץ ברמת המוכנות לבחינה של הנבחנים שמגיעים יותר מהמוסדות החדשים, נצפה לדפוס שונה של שיעורי המעבר לאורך השנים.

ניתוח תוצאות מבחני הרישוי לאורך השנים מראה כי מרבית הירידה בשיעורי העוברים ניתנת לייחוס לירידה ברמת המוכנות הממוצעת שעמה מגיעים בוגרי מרבית, אך לא כל, המוסדות. הממצאים מראים דפוס שונה בקרב בוגרי שלושה מוסדות: בוגרי שלוש אוניברסיטאות משקפים יציבות בשיעורי המעבר, בוגרי שלושה מוסדות אחרים משקפים ירידה קלה בשיעורי המעבר, ואילו בוגרי ארבעה מוסדות נוספים (ששיעורם היחסי בקרב הנבחנים עלה מאד בשנים האחרונות) משקפים ירידה משמעותית מאוד. ממצאים אלה מראים שיתכן שדרגת הקושי של בחינת ההסמכה עלתה מעט (בהנחה שרמת הלומדים ורמת הכשרתם במוסדות הללו נותרה קבועה), אך ההשפעה של דרגת הקושי על שיעור הנכשלים בטלה בשישים לנוכח ההשפעה שמסתמנת לשינוי ברמת המוכנות לבחינה של כלל אוכלוסיית הניגשים אליה. שינוי הבחינה (הרפורמה) גרם לשינוי מזערי במוסדות השונים, כאשר בחלקן עלה במקצת אחוז המעבר ובחלקן ירד במקצת, זאת בתלות ברמת הסטודנטים במוסד, ככל הנראה. לסיכום, העלייה בשיעורי הנכשלים בבחינות ההסמכה של לשכת עורכי הדין מתפרשת על ידי בעלי עניין שונים כעדות לכך שדרגת הקושי של הבחינות הולכת ועולה. ניתוח מדוקדק של שיעורי הנכשלים בקרב בוגרי המוסדות השונים מעלה שיעוקר ההסבר לעלייה בשיעורי הנכשלים אינו טמון בעלייה בדרגת הקושי של הבחינה, אלא בירידה ברמת מוכנותם של הנבחנים בה.

השוואה בין רכיבי ציון הבגרות: ציון שנתי וציון בחינה

יונתן סער

מרכז ארצי לבחינות ולהערכה (מאל"ז)

נורית ליכשטט

הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

מחקר זה בוחן את הקשרים והפערים בין ציון הבגרות השנתי ("ציון המגן") לבין הציון בבחינת הבגרות עצמה, ברמת המקצוע הבודד ומעבר למקצועות, בפילוחים שונים של משתני רקע: מגזר-שפה, מגזרים דוברי ערבית, סוג פיקוח ומגדר. עוד נבדק הקשר בין ציונים אלה לבין הציון בבחינה הפסיכומטרית. המחקר נערך על 735,238 רשומות (מקצוע לתלמיד) השייכות ל-91,689 תלמידים שנבחנו בבחינות הבגרות במחזור 2013.

הממצאים מעידים על מתאם גבוה בין ממוצע הבגרות הכללי כשהוא מחושב לפי הציון השנתי לבין זה המחושב לפי ציון הבחינה, הן מעבר לכלל התלמידים (0.92) והן בפילוח לפי משתני הרקע. גם ברמת המקצוע המתאמים גבוהים, ונעים בטווח שבין 0.74 ל-0.91.

לצד המתאמים הגבוהים, קיימים פערים לא זניחים בין ציון הבגרות השנתי לבין ציון הבחינה, והללו משתנים בין קבוצות אוכלוסיה. הפער המתוקנן בין ממוצע הבגרות הכללי כשהוא מחושב לפי הציון השנתי לבין זה המחושב לפי ציון הבחינה הוא כחצי סטיית תקן. הפער במגזר-השפה דובר העברית (0.48) נמוך מאשר במגזר-השפה דובר הערבית (0.61). גם בין קבוצות הפיקוח והמגזר בתוך כל מגזר-שפה קיימים הבדלים בפער: במגזר-השפה דובר העברית, הפער בחינוך הממלכתי נמוך מאשר בחינוך הדתי; במגזר-השפה דובר הערבית, הפער במגזר הבדואי גבוה מאשר במגזר הערבי והדרוזי. בחינוך הממלכתי (בשני מגזרי-השפה) הפער בקרב בנות גבוה מאשר בקרב בנים, ואילו בחינוך הדתי דובר העברית - המצב הפוך.

פערים בין ציון הבגרות השנתי לבין ציון הבחינה בולטים גם ברמת בית-הספר. 15.7% מבתי-הספר נמצאו מקלים - קרי, הפער בהם בין ממוצע הבגרות המחושב על בסיס הציון השנתי לבין זה המחושב על בסיס ציון הבחינה הוא עשר נקודות לפחות. ישנם הבדלים משמעותיים בפער בין מגזרי-השפה: 10.6% מבתי הספר במגזר-השפה דובר העברית נמצאו מקלים לעומת 34.0% במגזר-השפה דובר הערבית. קיימת שונות גדולה במדד זה גם בין קבוצות פיקוח ומגזר; הטווח נע בין 1.7% בחינוך הממלכתי דובר העברית ל-63.3% במגזר הבדואי.

המתאם בין הציון הפסיכומטרי הכללי לבין ממוצע הבגרות המחושב על בסיס ציון הבחינה מעבר למגזרי-שפה (0.63) גבוה במידת מה מהמתאם המקביל בין הציון הפסיכומטרי לממוצע המבוסס על הציון השנתי (0.55). בקרב כל אחד ממגזרי-השפה הפער בין המתאמים קטן יותר. לציון השנתי במתמטיקה עדיפות על ציון הבחינה מבחינת גובה המתאם עם הציון הפסיכומטרי המקביל בתחום הכמותי: 0.64 לעומת 0.56. מנגד, בשפת אם (עברית או ערבית) ישנה עדיפות למתאם בין ציון הבחינה עם הציון הפסיכומטרי בתחום המילולי על פני המתאם המקביל עם הציון השנתי: 0.54 לעומת 0.41. במקצוע האנגלית אין הבדל משמעותי במתאם בין ציון הבחינה והציון השנתי לציון הפסיכומטרי המקביל (0.76 ו-0.74, בהתאמה).

המחקר מבוסס על מחזור בגרות אחד בלבד, ולכך יש השלכות באשר ליכולת להכליל ממצאיו. עוד יש לזכור כי ממצאי המחקר משקפים מציאות שבה מפקח משרד החינוך ברמה הבית ספרית על הפער בין הציון השנתי לציון הבחינה. החלטה אפשרית על ביטולן של בחינות חיצוניות תגרור ממילא את ביטולן של מנגנון הפיקוח. לא מן הנמנע שהדבר יגרור עלייה בציונים השנתיים, ואיתה גם עלייה בפערים בין קבוצות אוכלוסייה.

תוקף מטלת הכתיבה של הבחינה הפסיכומטרית

תמר קנת-כהן, יונתן סער
מרכז ארצי לבחינות ולהערכה (מאל"ז)

מטלת הכתיבה, שהוכנסה לבחינה הפסיכומטרית בספטמבר 2012, מבוססת על כתיבת חיבור (בזמן מוקצב של 30-35 דקות) בנושא נתון. הציון מבוסס על ממוצע ההערכות של שני מעריכים, ומשקלו בציון בתחום החשיבה המילולית הוא 25%. כצפוי, המהימנות של מטלה כזאת אינה גבוהה, בפרט בהשוואה לשאלות ברירה. כך, למשל, מהימנות מבחן חוזר (עם נוסחים מקבילים) של מטלת הכתיבה ושל פרקי שאלות הברירה בחשיבה מילולית היא 0.53 ו-0.80 בהתאמה. לאור הנתונים האלה, ובהיות המהימנות תנאי הכרחי לתוקף, מתעוררת שאלת התוקף של מטלת הכתיבה, באופן כללי ובפרט בהינתן שהיא מהווה תוספת לפרקי שאלות הברירה בחשיבה מילולית.

הסטנדרטים למבחנים חינוכיים ופסיכולוגיים (AERA, APA, NCME, 2014) מציעים מקורות אחדים לראיות שבהן אפשר להשתמש בתהליך התיקוף. בהצגה זאת נתמקד בראיות שמבוססות על קשרים עם משתנים אחרים, שכוללות: ראיות מתכנסות ומבחינות וקשרי מבחן - קריטריון.

ראיות מתכנסות ומבחינות נבדקו, ראשית, על ידי מתאמים בין ציוני מטלת הכתיבה לציונים בחלקים אחרים של הבחינה הפסיכומטרית. המתאמים שהתקבלו היו: 0.62 עם פרקי שאלות הברירה בתחום החשיבה המילולית, 0.51 עם תחום האנגלית ו-0.50 עם תחום החשיבה הכמותית.

שנית, נבדקו מתאמים בין ציוני הבחינה הפסיכומטרית, ושמטלת הכתיבה בפרט, לציוני יע"ל (בחינת מיון בעברית שמיועדת למי שנבחנים בבחינה הפסיכומטרית בשפה שאינה עברית, שכוללת חיבור ושאלות ברירה). המתאמים שהתקבלו מראים שבעוד שרוב מרכיבי הבחינה הפסיכומטרית מתואמים יותר עם פרקי שאלות הברירה מאשר עם החיבור ביע"ל (כך, למשל, המתאמים של פרקי שאלות הברירה בתחום החשיבה מילולית בבחינה הפסיכומטרית עם החיבור ועם שאלות הברירה ביע"ל הם 0.45 ו-0.55 בהתאמה), מטלת הכתיבה מתואמת יותר עם החיבור מאשר עם פרקי שאלות הברירה ביע"ל (0.34 ו-0.32 בהתאמה).

שלישית, נבדקו מתאמים בין ציוני הבחינה הפסיכומטרית, ומטלת הכתיבה בפרט, לציונים במקצועות חובה בתעודת הברגות. המתאמים שהתקבלו (שחושבו בקרב לומדים באוניברסיטאות) מספקים אינדיקציה לכך שתרומת מטלת הכתיבה מעבר לפרקי שאלות הברירה, למתאמים של הציון בתחום החשיבה המילולית עם ציוני מקצועות הברגות גבוהה יותר במקצועות ברגות שדורשים כתיבה (תוספת של 0.03 בלשון והבעה) מאשר במקצועות שאינם דורשים זאת (תוספת של 0.00 במתמטיקה).

הקשרים עם שלושת סוגי המשתנים שתוארו לעיל מספקים ראיות מתכנסות ומבחינות לפירוש של ציון מטלת הכתיבה כמעיד על יכולת כתיבה (ולא על קונסטרוקט אחר).

קשרי מבחן - קריטריון נבדקו על ידי מתאמים בין ציוני הבחינה הפסיכומטרית לממוצע ציוני שנה א' באוניברסיטה. המתאם (כל המתאמים שלהלן תוקנו לקיצוץ תחום) בין מטלת הכתיבה לקריטריון (0.26) משקף עוצמה סבירה של קשר, בהינתן שמדובר בפרט פתוח של 30-35 דקות. הממצא החשוב יותר הוא שתוקף הניבוי של הציון הפסיכומטרי (0.42) לא נפגע בעקבות הכנסת מטלת הכתיבה. אומדן לתוקף הניבוי של ציון פסיכומטרי היפותטי, שאינו כולל את מטלת הכתיבה, אך מבוסס על זמן היבחנות כולל זהה, מניב ערך דומה. ממצאים אלה מעידים שמטלת הכתיבה מוודדת יכולת רלוונטית וייחודית להצלחה בלימודים אקדמיים, למרות מגבלות המהימנות שלה, ולפיכך מספקים הצדקה לשימוש בה בתהליך הברירה.

איך מודדים אוריינות מתמטית כחלק מ"היעדים להתפתחות ברת-קיימא" של האו"ם? אתגרים בפסיכומטריקה ופוליטיקה של מדידה השוואתית

עדו גל

אוניברסיטת חיפה, החוג לשירותי אנוש

הרצאה זו עוסקת בסוגיות מתודולוגיות ומקצועיות הקשורות לצורך לפתח גישה למדידה ודיווח של אינדיקטור בינלאומי מוסכם לגבי רמת אוריינות מתמטית של בוגרים, כחלק ממערכת היעדים להתפתחות ברת-קיימא של האו"ם (SDG) הכוללת כ-170 מדדים. תכנית המדידה הבינלאומית היחידה בתחום, PIAAC (סקר מיומנויות בוגרים של OECD), אשר הסבב השני שלה מתוכנן ל-2023, אומצה רק ע"י מספר קטן יחסית של 33 מדינות (כולל ישראל) מרביתן מדינות מפותחות. לרוב מדינות העולם, מעל 180 במספר, יש צורך במערכת חדשה ופשוטה יותר המותאמת למאפיינים ההטרוגניים של מדינות העולם.

ההרצאה מתמקדת באינדיקטור 4.6.1 אשר אמור לאפשר לכל מדינה לדווח על "אחוז הגברים והנשים אשר הגיעו לרמה נדרשת הן באוריינות מתמטית הן באוריינות לשונית". מדד זה אמור להתבסס על מדידה קוגניטיבית ישירה (באמצעות מבחן) של רמת האוריינות. ההרצאה תסקור כמה מהאתגרים והדילמות אתם מתמודדת קבוצת המומחים של אונסק"ו בתחום זה, בעיקר:

1. סוגיות הקשורות להמשגה ולקביעת רמת המינימום לדיווח המדד בתחום אוריינות מתמטית. נושא זה מעלה שאלות לגבי המשמעות הפוליטית והכלכלית של המידע המדווח.
2. הצורך לפתח במקביל כמה מתכונות למדידה שמותאמות למגבלות יכולת טכנית ולוגיסטית במדינות רבות, כולל שימוש במבחן אדפטיבי מבוסס-מחשב, לצד כלי קצר ומודפס (ללא מחשב) שניתן ליישם במהירות ובקלות יחסית בתנאי שטח מאוד שונים.
3. שיפור תוקף המדידה על ידי הפחתת ההשפעה של יכולות לשוניות ורמת הבנת הנקרא של נבחנים על המדידה של אוריינות מתמטית.
4. הצורך בניתוח השוואתי של מאגרי פריטים שפותחו בנפרד על ידי כמה יצרני סטטיסטיקה, והשוואת רמות הקושי של הפריטים, בלי שניתן לבצע מחקרי linking כמקובל.

הנושא של אוריינות מתמטית הן של בוגרים הן של תלמידים זוכה לתשומת לב בכל מדינות העולם, ויש אליו התייחסות בעולם ובישראל בכמה תכניות מדידה. המדדים שיפותחו בתכנית הבינלאומית שמותנעת על ידי אונסק"ו עשויים להידרש בעתיד לדיווח על ידי מדינת ישראל, וחלק מהסוגיות העולות בהקשר של המיזם הנוכחי צצות ועולות בפרוייקטים רבים בעולם ובישראל. ההרצאה תעדכן ביחס למגמות עולמיות הנמצאות בפיתוח בימים אלו, ותתאר את המורכבות של פיתוח גישות למדידה ברת-השוואה של משתנה חברתי חשוב, בתוך זירה פוליטית רוויה אשר משפיעה על תהליך הפיתוח ויוצרת אילוצים מתודולוגיים ואתגרים אחרים.

סימפוזיון

הערכת הבעה בכתב במבחנים רחבי היקף בישראל

יואל רפ

הרשות הארצית למדידה והערכה (ראמ"ה), מרכז ארצי לבחינות ולהערכה (מאל"ו)

ספי פומפיאן

מרכז ארצי לבחינות ולהערכה (מאל"ו)

כתיבה היא אחת מארבע מיומנויות השפה העיקריות, ויש לה חשיבות עליונה לאדם המודרני. הכתיבה משמשת את הכותב במגוון משימות: היא אמצעי תקשורת, ביטוי עצמי ולמידה, והיא מסייעת בהקשרים יומיומיים, אקדמיים ואומנותיים.

יכולת הכתיבה נחשבת ליכולת מורכבת, ולמידת הכתיבה הן אחד ההיבטים הקשים והמורכבים ביותר בהוראת שפות ובלמידתן. הערכה של יכולת התלמידים בכתיבה ובהבעה בכתב היא חלק חיוני ואינטגרלי מתהליך הלמידה. אומדן של יכולת זו מסייע למערכת החינוך בכללותה, והוא משמש גם לצורכי קבלה, מיון והשמה למוסדות אקדמיים ותעסוקתיים. לא פעם הערכת מיומנות הכתיבה וההבעה בכתב משולבת בכלי המדידה והמבחנים המשמשים להערכת מיומנויות מרכזיות שפתיות אחרות, כגון הבנת הנקרא, הבנת הנשמע וידע לשוני.

בישראל מערכת החינוך מעריכה את יכולת הכתיבה של התלמידים בכמה וכמה תחנות במהלך לימודיהם. בשנים הראשונות, בשלב רכישת מיומנויות היסוד של השפה (קריאה וכתיבה), המורים משתמשים במבדקים סטנדרטיים כדי להעריך את יכולת הכתיבה של תלמידי ישראל וכדי לברר אם הגיעו לרמת הכתיבה המצופה מהם. מבחני המיצ"ב, הבדקים את הישגיהם של תלמידי ישראל ונערכים בכיתות ה' ו-ח', מעריכים את מיומנות ההבעה בכתב במבחני השפות (הן שפות האם - עברית וערבית - והן אנגלית). יכולת הכתיבה בעברית של תלמידים דוברי ערבית נבדקת במבחני עלד"ע (עברית לדוברי ערבית), המועברים במתכונת מדגמית בקרב תלמידי כיתות ו' בבתי ספר ערביים. לבסוף, כתיבה נבדקת במבחני הבגרות בשפות (הן שפות האם - עברית וערבית - והן שפות זרות ושניות). נוסף על מבחנים רחבי היקף אלו במערכת החינוך, יכולת הכתיבה נבדקת גם באמצעות מערכות המיון להשכלה גבוהה, הן כחלק מהמבחן הפסיכומטרי והן במסגרת מבחני ידע בעברית למועמדים שאינם דוברי עברית כשפת אם. מערכות אלו מעריכות את יכולת הכתיבה באמצעות מטלות כתיבה שבהן יש לכתוב קטע טקסט בנושא מסוים (עם גריין מובנה או בלעדיו) ובסוגת כתיבה מסוימת (טיעונית, תיאורית וכיוצא באלה). מעריכים המומחים לכתיבה, לשפה או להוראתה בודקים את תוצרי כל מטלה כזאת לאחר שקיבלו הכשרה ייעודית לבדיקתה.

הסימפוזיון הנוכחי יבחן את מצב ההערכה רחבת היקף של יכולת הכתיבה בישראל דרך ארבע עבודות מחקר שכל אחת מהן עוסקת באחת מתחנות ההערכה. עבודה אחת עוסקת בפרק הכתיבה הממזגת שבבחינת הבגרות המתוקשבת במקצוע "עברית - הבנה, הבעה ולשון", עבודה אחרת עוסקת בפרק הכתיבה של מבחן עלד"ע (עברית לדוברי ערבית), ושתי הנותרות עוסקות במבחן הכתיבה הנכלל בבחינה הפסיכומטרית לקבלה לאוניברסיטאות בישראל.

כל העבודות בוחנות היבטים חשובים ומגוונים של הערכת הכתיבה. הן בודקות אם מחוון הוליסטי עדיף על מחוון אנליטי, מהי השפעתם של מגוון גורמים על איכות בדיקתן של מטלות כתיבה בידי מעריכים אנושיים ואיזו הכשרה יש להעניק להם כדי לשפרה. כמו כן הן מעלות שאלות פסיכומטריות גרידא, כגון כיצד נכון לבדוק את מהימנות הבדיקה האנושית.

סימפוזיון הערכת הבעה בכתב במבחנים רחבי היקף בישראל מחווון הוליסטי או אנליטי? מהימנות הערכת כתיבה במבחן עברית לדוברי ערבית בכיתות ו' (עלד"ע)

אלה שפירא, ליאת אריאלי-מאור, ענבל רון-קפלן
הרשות הארצית למדידה והערכה (ראמ"ה)

אחת הדרכים המקובלות להערכת יכולות שפה ואוריינות היא מבחן הכולל, בין השאר, מטלות לבדיקת יכולת הבעה בכתב שבהן התלמיד נדרש להפיק טקסט. מטלות כאלה דורשות מהתלמיד אינטגרציה של כמה מיומנויות שפתיות: המשגה לשונית, התאמת שפה להקשר, בהירות, הכרה בנמען ועוד. מאחר שיכולת הבעה בכתב היא מורכבת מבחינה קוגניטיבית, קשה למדוד אותה ולהעריכה באופן תקף ומהימן.

בשדה החינוכי מקובל להשתמש באחת משתי שיטות להערכת הבעה בכתב: שיטה הוליסטית ושיטה אנליטית. השיטה ההוליסטית רואה בהבעה בכתב יכולת שלמה, שאין להעריכה באמצעות פירוק למרכיבים. לכן מחווון הערכה המבוסס על השיטה ההוליסטית מנחה את המערך לתת ציון אחד לביצוע במטלת הבעה בכתב. השיטה האנליטית רואה בהבעה בכתב יכולת המורכבת מכמה מרכיבים, שיש להעריכם בנפרד ובאופן בלתי תלוי ככל האפשר. לכן מחווון המבוסס עליה דורש מהמערך להתמקד בהיבט אחד של הכתיבה בכל פעם (כגון איכות התוכן, לכידות וקישוריות, תחביר ומורפולוגיה) ולתת לו ציון שאינו תלוי בציוני המרכיבים האחרים. הציון הסופי הוא הממוצע המשוקלל של ציונים אלו.

במחקר הנוכחי ביקשנו לבדוק אם הערכת מטלת הבעה בכתב לפי השיטה האנליטית מהימנה יותר מהערכת אותה מטלה לפי השיטה ההוליסטית, וכן לבדוק אם השיטה הנבחרת משפיעה על הציון הסופי במטלה וכיצד. כמו כן ביקשנו להשוות בין חוויית המעריכים בהליך הערכה אנליטי לבין חווייתם בהליך הוליסטי. לשם כך השתמשנו במטלת הבעה בכתב במבחן בעברית כשפה שנייה לתלמידים דוברי ערבית בכיתה ו' – מבחן עלד"ע (לפרטים על המבחן ראו אתר ראמ"ה). במסגרת המחקר בדקנו כ-400 מטלות משתי סוגות כתיבה (תיאורית וטיעונית) באמצעות מחווון הוליסטי ובאמצעות מחווון אנליטי.

מצאנו כי הערכה בגישה האנליטית מהימנה יותר מהערכה בגישה ההוליסטית. בגישה האנליטית עמדה המהימנות בין שופטים (על פי מתאם פירסון) על 0.75, ואילו בגישה ההוליסטית היא עמדה על 0.64. בחינה מדוקדקת של כל אחד מהמרכיבים העלתה כי מקור הפער הוא בעיקר במהימנות של הערכת רכיבי הלשון. נוסף על כך, מצאנו כי השיטה האנליטית תורמת לשביעות רצונם של המעריכים יותר מהשיטה ההוליסטית. המעריכים דיווחו כי המחווון האנליטי מקל עליהם את בדיקת החיבורים ואת קבלת החלטות הנוגעות להם.

אנו מציעים שילוב של שני סוגי המחווונים המתאפיינים במהימנות גבוהה והן ביעילות בבדיקה.

סימפוזיון הערכת הבעה בכתב במבחנים רחבי היקף בישראל אמידת מהימנות בין מעריכים - המקרה של מטלת הכתיבה הממזגת (הבעה בכתב) במבחן הבגרות הממוחשב

בועז רוזנבאום, רוני ברצלבסקי, יואל רפ
הרשות הארצית למדידה והערכה (ראמ"ה)

מבחן הבגרות המתוקשב במקצוע "עברית - הבנה, הבעה ולשון" כולל, בין השאר, מטלת כתיבה מורחבת שבה הנבחן מתבקש לסכם בכתב שלושה מאמרים המוצגים כדפי אינטרנט. החיבורים מוערכים באמצעות מחוון אנליטי המבחין בין שלושה עשר תבחינים, או ממדים, של הכתיבה. הציון בכל תבחין נע על פני ארבע רמות (0-3).

במבחני הבגרות המודפסים במקצוע העברית שני מעריכים בלתי תלויים בודקים כל מחברת בחינה בבדיקה אורכית, כלומר מעריכים את תשובות הנבחן לכל השאלות בטרם יעברו לנבחן הבא. בשאלות הפתוחות במבחני הבגרות הממוחשבים, לעומת זאת, מעריכים בודקים את השאלות הפתוחות בבדיקה רוחבית, כלומר מעריכים את תשובות כל הנבחנים לשאלה מסוימת בטרם יעברו לשאלה הבאה, וכל תשובה נבדקת פעמיים, בידי שני מעריכים אקראיים. היתרון העיקרי של בדיקה רוחבית הוא ההתמחות שהיא יוצרת אצל הבודקים בנוגע לכל שאלה ושאלה. כמו כן, מכיוון שהבודקים עוברים כל העת מנבחן לנבחן, אין הם מפתחים אפקט הילה ביחס לנבחן כלשהו.

בסופו של דבר, הן במבחן המודפס והן במבחן הממוחשב שני מקבצי הערכות מופקים לכל נבחן, אך בעוד שבמבחן המודפס כל מקבץ משקף את הערכותיו של מעריך אחד, במבחן הממוחשב כל מקבץ משקף את הערכותיהם של כמה מעריכים.

ראמ"ה בדקה את המהימנות בין המעריכים במבחן הממוחשב. בדיקות המהימנות נערכו על הערכות התשובות לשאלות הפתוחות, ובכללן הערכות מטלת הכתיבה בכל תבחין ותבחין. הבדיקות המובאות כאן נערכו על הציונים במטלת הכתיבה לפני ביצוע הערכה בכירה. במקרה זה האתגר באמידת המהימנות הוא כפול: ראשית, מעריכים רבים העריכו נבחנים רבים, ושנית, גם בקרב צמדי מעריכים שהעריכו את אותן קבוצות נבחנים, סדר ביצוע הערכות אינו אחיד. אנו נציג כמה דרכים לחישוב מהימנות המעריכים - מתאם פשוט, ממוצעי המתאמים של צמדי המעריכים ומתאם ICC (Intra-Class Correlation) - ונדון ביתרונותיהן ובחסרונותיהן.

סימפוזיון הערכת הבעה בכתב במבחנים רחבי היקף בישראל האם וכיצד משפיעה הכשרת מעריכים על ההערכה של מטלות כתיבה במכפ"ל

נעמי גפני, ספי פומפיאן, אבי אללוף, ג'ני שמולביץ, מיכל חטבי
מרכז ארצי לחינות ולהערכה (מאל"ו)

החל מאוקטובר 2012 הבחינה הפסיכומטרית כוללת מטלת חיבור. הציון במטלה נקבע על בסיס ממוצע ההערכות של שני מעריכים. בכל מועד, לאחר שהמבחן מועבר, אנשי הצוות המילולי בוחרים חיבורים ועל בסיסם בונים סדנה להכשרת המעריכים. אורכה של הסדנה כחמש שעות, ואופייה - פרונטלי או מתקשב - נקבע בהתאם לניסיון המעריך. הכשרת המעריכים דורשת השקעה רבה של משאבי כוח אדם.

במחקר הנוכחי נבדקה השפעת ניסיון המעריך ואופי הכשרתו על איכות הערכותיו, וזאת בעזרת ניסוי ש-29 מעריכים העריכו בו 150 חיבורים שנכתבו בתגובה לאותה המטלה. כמו בהערכת המטלה בפועל, בניסוי העריך כל מעריך את החיבורים בשני ממדים, תוכן ולשון, והציון שהתקבל היה ממוצע הציונים בשני הממדים. המעריכים נבדלו זה מזה באופי הכשרתם ובהיקף ניסיונם.

המשתנים הבלתי תלויים היו סוג ההכשרה וותק המעריכים. במחקר נכללו שלושה סוגי הכשרה: סדנה ארוכה (10 מעריכים), סדנה קצרה (13 מעריכים) והכשרה ללא סדנה, ובה המעריך מקבל הדרכה כתובה ומשוב קצר על דיוק ההערכה שנתן לעשרה חיבורים (6 מעריכים). כל המעריכים בניסוי עברו בעבר סדנה ארוכה אחת או יותר. למחצית מהמעריכים בכל קבוצה היה ניסיון רב בהערכה (למעלה מחמש סדנאות), ולמחצית האחרת היה ניסיון מצומצם יותר (עד חמש סדנאות בעבר).

המשתנים התלויים במחקר, המעידים על טיב ההערכה, היו:

1. המתאם של ציוני החיבורים עם ציונם ה"אמיתי" - ממוצע הציונים שכל משתתפי הניסוי העניקו להם
2. הפער בין הציון שהעניק המעריך לחיבור ובין ציונו ה"אמיתי"
3. המתאם של ציון החיבור עם הציון שקיבל כתבו בפרק החשיבה המילולית בבחינה הפסיכומטרית
4. המתאם בין ציוני החיבור בשני הממדים
5. משך הערכה ממוצע לחיבור (הספק)

לא נמצא שהוותק או סוג ההכשרה של המעריך קשורים למתאם שבין ציוני המעריכים לבין הציון האמיתי או לפער ביניהם. כמו כן לא נמצא קשר בין הוותק או סוג ההכשרה לבין המתאם של הציון בחיבור עם הציון בפרקים המילוליים הסגורים. המתאמים בין שני ממדי ההערכה (תוכן ולשון) היו דומים זה לזה בכל הקבוצות. המתאם בין שני הממדים היה הנמוך ביותר בקרב המעריכים שעברו הכשרה ללא סדנה. לבסוף, לא נמצאה השפעה של סוג ההכשרה על ההספק של המעריכים.

מסקנת המחקר היא כי סדנה ארוכה אינה מביאה לידי הערכה מדויקת יותר או מהירה יותר. תוצאה זו מעניינת משום שהיא עומדת בניגוד לתפיסה כי הכשרת מעריכים אינטנסיבית משפרת את ההערכה. נראה כי מעריכים שרכשו ניסיון בהערכה זקוקים להתנסות קצרה, דוגמת ההכשרה ללא סדנה שהוצעה בניסוי, וסדנה אינטנסיבית אינה משפרת את הערכתם. כמו כן לא נמצא קשר בין איכות ההערכה לבין הוותק של המעריך.

ראוי לציין כי כל המעריכים במאל"ו, כולל אלה שהשתתפו בניסוי, עברו סדנה ארוכה לפני שהתחילו לעסוק בהערכה. לאור התוצאות הנוכחיות יש לשקול לשנות את סוג ההכשרה הניתן למעריכים ותיקים ולהסתפק בסדנאות קצרות ובהדרכה כתובה עם משוב.

סימפוזיון הערכת הבעה בכתב במבחנים רחבי היקף בישראל מדדי איכות מרכזיים של מערכי מטלת הכתיבה במכפ"ל – תיאור וניתוח רב שנתי 2013 – 2017

**אבי אללוף, מרינה פרונטון, האשם ניג'ם
מרכז ארצי לבחינות ולהערכה (מאל"ו)**

באוקטובר 2012 נוספה מטלת כתיבה לבחינה הפסיכומטרית. שינוי זה היה דרמטי: הוספת חלק פתוח לבחינה סטנדרטית שלאורך 28 שנים, למן היווצרה, הכילה שאלות סגורות בלבד. בשנים האחרונות הרבה מאל"ו לעקוב אחר מטלת הכתיבה ולחקור אותה. כך, למשל, חושבו מתאמים בין ממדי ההערכה ומתאמים בין המעריכים, נאמדה מהימנות המטלה, נחקרו מדדי איכות של המעריכים, נערך מחקר המערער על המוסכמה ולפיה פער גדול בין שתי הערכות מחייב הערכה שלישית, ולעומתו נערך מחקר השואל אם לא די במעריך אחד הבודק את אותה המטלה פעמיים. מובן שגם תרומת המטלה לתוקף הבחינה נבדקה. מחקר שנערך לאחרונה ומוצג בסימפוזיון זה עוסק בהשפעת אופי הכשרת המעריכים על טיב ההערכה, ומחקרים אחרים נערכים בימים אלו. עבודה זו מתמקדת במערכי מטלת הכתיבה. מעריך שאיכותו גבוהה תורם למהימנות המטלה ולתוקף המבחן כולו, ואילו מעריך שאיכותו נמוכה פוגע בהם. כשם שבעת בניית מבחן סגור יש לבדוק את איכותה של כל שאלה, כך יש לבדוק את איכותו של כל מעריך. לכן מאל"ו עוקב אחר ביצועי המעריכים ושולח אליהם משוב לאחר כל מועד בחינה. מעריכים שאיכותם נמוכה אינם מועסקים עוד. כמו כן, בכל שנה 15 עד 20 אחוז מהמעריכים נושרים מעצמם וחדלים לעבוד במאל"ו.

מטרת המחקר הנוכחי היא לגבש תמונה רחבה וחמש-שנתית של מערכי מטלת הכתיבה, וזאת לפי מדדים חשובים של ביצועיהם (איכות, הספק ויציבות) ולפי משתנים ביוגרפיים הנוגעים להם (גיל, מגדר וניסיון בהערכה). המחקר מנסה לברר אם איכות ההערכה היא תכונה יציבה של מעריך, אם חל בה שינוי לאורך השנים ואם יש קשר בינה לבין ההספק שלו, וכן אילו מעריכים נוטים לנשור. נוסף על כך, מחקר כזה יכול לספק מידע על מעריכים ספציפיים.

במחקר נבדקו המעריכים בעברית והמעריכים בערבית. המשתנים הבלתי תלויים במחקר היו המשתנים הביוגרפיים (גיל, מגדר, וניסיון בהערכה), והמשתנים התלויים בו היו מדדי איכות ההערכה, ההספק והיציבות. איכות ההערכה נאמדה על סמך הדמיון שבין הערכות המעריך לבין הערכות המעריכים האחרים. המדדים הרלוונטיים כאן הם מתאם, מדדי הסכמה ופער. מדד נוסף לאיכות הוא המתאם שבין הערכת המעריך לבין קריטריון חיצוני, כמו הציון בחלק הסגור בפרק החשיבה המילולית בבחינה הפסיכומטרית. הספק ההערכה נאמד על סמך מספר החיבורים שבדק המעריך בפרק זמן נתון – מספר שיש לו השלכות על עלות ההערכה, היות שהמעריכים מתוגמלים לפי מספר שעות עבודה. היציבות חושבה לפי סטיית התקן של מדד האיכות בחמש שנים.

שולחן עגול I

כיצד "לימודי התעודה בפסיכומטריקה" יקדמו את הפסיכומטריקה בישראל?

אבי אללוף, צור קרליץ, ניצן פרידמן
מרכז ארצי לבחינות ולהערכה (מאל"ו)

באקדמיה בישראל כמעט אין כיום אנשי סגל בתחום של המדידה וההערכה ומכיוון שאין בתחומים אלה תוכניות לימודים לתואר שני ושלישי, אין הכשרה אקדמית של דור מומחים חדש. אנשי המקצוע מוכשרים חלקית במוסדות שבהם הם עובדים. כדי לתקן זאת, התגבשה ועדה ללימודים מתקדמים בפסיכומטריקה, שהמליצה על מתן מלגות לימודים לדוקטורנטים ולפוסט דוקטורנטים, ועל פתיחת מסלול לימודי תעודה בפסיכומטריקה – מסלול שאכן נפתח בתחילת 2018. המימון הנדרש למיזם מקורו בקרן פילנתרופית, והוצאתו לפועל מתבצעת על ידי מאל"ו.

המצב בעולם – החוסר באנשי מקצוע בפסיכומטריקה בישראל מאפיין מדינות רבות נוספות בעולם, המתמודדות איתן בדרכים שונות – הקמת מרכזים אקדמיים, פתיחת מסלולי לימוד מתקדמים, השקת סמינרים וסדנאות לפיתוח מקצועי, קידום תוכניות מנטורים ויצירת שיתופי פעולה מקצועיים. אפילו בארה"ב, "ארץ המבחנים" שבה יש תוכניות מעולות באוניברסיטאות רבות, קיים חוסר בתלמידים לתארים מתקדמים ובאנשי מקצוע בתחום. ארגונים דוגמת ה-NCME מתחבטים בסוגית הכשרת הדור הבא של הפסיכומטריקאים.

לימודי התעודה בפסיכומטריקה – הלימודים מיועדים לאנשים העוסקים במדידה והערכה במסגרת עבודתם ומעוניינים בפיתוח מקצועי, ולאנשים המעוניינים להשתלב בתחום בעתיד. הלימודים מתאימים לאנשים העובדים בתחומי מדידה והערכה במערכת החינוך, במשרדי ממשלה ובארגונים ציבוריים, ולסטודנטים לתארים מתקדמים המעוניינים לקבל הכשרה פסיכומטרית וליישמה בעבודת המחקר שלהם. מטרת התוכנית היא לתת רקע תאורטי בפסיכומטריקה; להקנות חשיבה ביקורתית ועצמאית בתחום הפסיכומטריקה – כולל התייחסות להשפעות המבחנים על החברה; ללמד מגוון שיטות כמותיות וסטטיסטיות שימושיות בתחום הפסיכומטריקה; ולפתח את המיומנויות הנדרשות לפיתוח כלי מדידה והערכה.

סטטוס המיזם נכון לאוקטובר 2018 – מלגות – הוענקו ארבע מלגות: אחת לדוקטורנטית בארה"ב, ושלוש לדוקטורנטיות הלומדות את לימודי התעודה, ועבודות המחקר שלהן מתבצעות בליווי פסיכומטריקאי. יצא קול קורא לשלוש מלגות פוסט דוקטורט בחו"ל. התוכנית ללימודים מתקדמים בפסיכומטריקה כוללת שישה קורסים. אחד מהם – "סטטיסטיקה ושיטות מחקר מתקדמות" נלמד במגוון מוסדות; "מבחנים וחברה" נלמד במכללה האקדמית תל-אביב יפו. יתר הקורסים פותחו במאל"ו: פיתוח כלי מדידה והערכה, יסודות הפסיכומטריקה, תורת התגובה לפריט וכריית נתונים, וסוגיות מעשיות במדידה ובהערכה. "פיתוח כלי מדידה והערכה" הוא קורס מקוון שהחל מ-2019 יהיה פתוח ברשת לציבור כולו, ללא עלות.

המיזם והאגודה הישראלית לפסיכומטריקה – כמיזם שמטרותיו מתלכדות עם מטרות אפי, המיזם קיבל את ברכת אפי. התוכנית הראשונית של לימודי התעודה הועברה לצוות יועצים מאפי, שהצעותיהם יושמו בגרסה הסופית של לימודי התעודה. קולות קוראים למלגות מתפרסמים באתר אפי, ועדכונים ומידע על לימודי התעודה נשלחים לתפוצת הארגון.

השולחן העגול – אנו מעוניינים להפיק את התועלת המרבית מהמיזם, וסבורים שדיון פתוח בכך במסגרת הכנס ישרת מטרה זו. בתוכנית "שולחן עגול" בן שלושה חלקים: (א) הצגת הפרויקט והקורסים שעברו עד כה, כולל לקחים מההעברה (ב) פאנל המונה חמישה נציגים חברי אפי (ד"ר חגית גליקמן, ראמ"ה; גיא אושר להב, אדם מילוא; לירון צוברי, אוניברסיטת תל אביב; נועם מרקוביץ, האוניברסיטה העברית; נעמה קצין, אוניברסיטת בן גוריון), ששייבו לשאלות שיופנו אליהם על המיזם ועל תרומתו (ג) דיון מונחה בהשתתפות הקהל. בדיון נעלה שאלות לגבי אפשרויות להגעה לקהלים רחבים יותר של לומדים במיזם, צורכי הפיתוח המקצועי של חברי אפי, גורמים המונעים מאנשים להתפתח מקצועית, ועוד.

שולחן עגול II

כשבינה מלאכותית פוגשת פסיכומטריקה במגדל בבל

ריטה סבר

האוניברסיטה העברית בירושלים

מדידה פסיכולוגית ו/או חינוכית נתקלת בקשיים כשהיא נעשית בקרב אוכלוסיות בעלות שונות תרבותית גבוהה. העוסקים בה חווים תסכול וקוצר-יד כל שגוברת המודעות לכך שגם כאשר הם משתמשים בכלי מדידה שנמצאו בעלי תוקף גבוה – הכלים הללו תקפים בעיקר עבור נבחנים מהתרבות ההגמונית ומהזרם המרכזי; ואילו עבור נבחנים מתרבויות אחרות ומי ששפת אמם איננה השפה המקומית הדומיננטית – תקפם מוטל בספק אפילו כאשר ניתן להציע לנבחנים הללו תרגומים לשפתם. תרגום גרידא איננו פותר את הבעיה, ובכל מקרה קיים קושי רב (ונחוצים משאבים רבים) להפיק מבחן בשורה של שפות שונות ולתקף אותו בכל אחת מהשפות השונות. הקושי הולך ומתעצם ככל שגלי הגירה מגדילים את מספר השפות והתרבויות השונות שצריכות לקבל מענה באותה חברה; וככל שמתחזקות הדרישות לשוויון הזדמנויות בקרב בניהן של קבוצות תרבות מודרות.

במקביל, הולך ומתרחב השימוש בטכנולוגיות של בינה מלאכותית וביכולת של מכונות לבצע משימות שבעבר יכלו רק בני אדם לבצע. והשאלה היא אם אפשר (וכדאי) להיעזר בטכנולוגיות הללו כדי להתמודד עם הבעייתיות הפסיכומטרית שתוארה לעיל.

בינה מלאכותית (Artificial Intelligence (AI), מבוססת על הרעיון שמערכות ממוחשבות יכולות ללמוד מנתונים, לזהות דפוסים, לארגן מידע, לזהות יחסים, להפיק ניבויים, להתאים עצמן לתשומות חדשות, לקבל החלטות (כמעט) ללא התערבות אנושית ולבצע משימות כאילו-אנושיות. דוגמאות שאנו מכירים כיום – כגון מחשבים המשחקים שחמט או מכונות אוטונומיות – נשענות בין היתר על טכנולוגיות של למידה עמוקה deep learning ושל עיבוד שפה טבעית natural language processing. שימוש בטכנולוגיות הללו מאפשר למחשבים ללמוד לבצע משימות ספציפיות ע"י עיבוד כמויות גדולות מאד של נתונים וזיהוי דפוסים בתוך הנתונים.

בתחום החינוך, יישום מתודולוגיות של מיצוי מידע שימושי מתוך מאגרי מידע גדולים מכונה ניתוח למידה – learning analytics (LA) וכריית מידע בחינוך educational data mining (EDM)². הוא משמש להפקת תובנות המבוססות על כריית מידע – data mining (EDM) – מתוך מאגרי מידע גדולים על סטודנטים וניתוח כמות גדולה של נתונים מורכבים. מה שעשוי להיות רלבנטי במיוחד לפסיכומטריקה בחברה מרובת תרבויות הוא השימוש בטכנולוגיות האלה מאפשר לסרוק מאגרי מידע (דיגיטאליים) גדולים המצויים במערכות חינוך בשביל לגלות דפוסים המופיעים רק אצל מספר קטן של סטודנטים³.

בשולחן העגול הזה המשתתפים מוזמנים להציע תשובות יצירתיות לשאלות כגון: איך, אם בכלל, ובאילו תנאים, תוכל הפסיכומטריקה להיעזר בבינה מלאכותית בכל הנוגע לשימוש במדידות פסיכולוגיות וחינוכיות באוכלוסיות בעלות שונות תרבותית גבוהה? מה היתרונות ומה המגבלות של שימוש בבינה מלאכותית בפסיכומטריקה? האם אפי אמורה להציע לחברה סדנאות בנושאים בתחום הבינה המלאכותית? האם רכישת מיומנויות של שימוש ב-AI בכלל, וב-learning analytics בפרט, רלבנטית להכשרת פסיכומטריקאים כבר כיום?

- 1 Sever, R. (2019). Caught in the mystery – What happens to psychologists in cross-cultural personality assessment and/or treatment? Paper to be presented at the WAPP 2019 (World Conference on Personality) in Hanoi, Vietnam, 2-6 April 2019.
- 2 Siemens, G. & Baker S.J.d (2012). Learning analytics and educational data mining: towards communication and collaboration. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge. Vancouver, British Columbia, Canada — April 29 - May 02, 2012 . Pages 252-254
- 3 Clark Jeria Frías (2017) Learning Analytics: When Artificial Intelligence meets Education. Published on March 10, 2017. <https://www.linkedin.com/pulse/learning-analytics-when-artificial-intelligence-meets-jeria-frías>

פיתוח סולם ציונים למבחן 2 בסוללת וודקוק-ג'ונסון 3

יואל רפ, אימאן עוואדיה, אלה שפירא, פרנקלין זרומב
הרשות הארצית למדידה והערכה (ראמ"ה)

אסטלה מלמד

מרכז ארצי לבחינות ולהערכה (מאל"ו)

סוללת מבחני וודקוק-ג'ונסון 3 לבדיקת כשרים קוגניטיביים של ילדים ומבוגרים תורגמה והותאמה לעברית ולערבית בישראל. הפרויקט, אשר תואר בהרחבה בכנסי אפי קודמים, עומד בפני סיום והגיע לשלב חישוב הציונים והנורמות.

אחד המבחנים אשר הציבו אתגר מיוחד בפני צוות התרגום וההתאמה בישראל בשלב חישוב הנורמות היה מבחן 2, מבחן הבודק למידה חזותית-שמיעתית. האתגר נוגע לאופן קביעת הציון "הגולמי" הסופי שעל בסיסו יחושב הציון היחסי (ציון ה"נורמה"). הייחוד נובע מהשילוב שבין אופי המבחן (הכולל בתוכו למידה של סמלים שונים המשמשים לפענוח "משפטים" המורכבים מהם) וקיומן של נקודות עצירה במבחן.

כמו בכל המבחנים בסוללת וודקוק-ג'ונסון, גם במבחן זה נקבעו מספר נקודות עצירה שמטרתן למנוע מנבחנים מתקשים להתמודד עם משימות שהסיכוי שלהם להצליח בהן נמוך מאוד. נקודות העצירה נקבעו לאורך המבחן במיקומים אשר בסבירות גבוהה אפשר להניח שנבחנים, שלא הצליחו מספיק עד לאותה נקודה, לא יצליחו בפריטים הבאים אחר כך. באופן מעשי, כאשר עוצרים את המבחן של נבחן מסוים בנקודת עצירה כזו, מניחים כי הוא היה טועה בכל הפריטים הבאים אחרי אותה נקודה ומחשיבים את כל הפריטים שלאחר נקודת העצירה כאילו הנבחן טעה בהם. ברם, הנחה זו פחות תקפה למבחן 2 כיוון שבמבחן זה נעשית למידה של סמלים (שמהם מורכבים "משפטים") ואותם סמלים ממשיכים להופיע בכל פריטי המבחן, כלומר, גם בפריטים המוצגים לנבחן לאחר נקודות העצירה השונות. לכן, במבחן זה סביר להניח כי נבחן שנעצר בנקודת עצירה כלשהי היה מצליח לזהות לפחות חלק מהסמלים שנכללים במשפטים שבהמשך המבחן. לכן, לא נכון לזקוף באופן גורף לכל נבחן שעצר, את מספר הטעויות המקסימלי (דהיינו את מספר הסמלים) בכל משפט באותם משפטים שאליהם הנבחן לא הגיע.

זקיפה של ערך מקסימלי של טעויות במשפט למי שלא ביצע את המשפט, מגדילה באופן מלאכותי את הפער בין נבחנים שנעצרו בנקודת עצירה כלשהי לאלו שלא ומטה לרעה את הציון של הנבחנים שנעצרו.

בעבודה זו תוצג שיטה שפותחה במטרה לאמוד את מספר הטעויות שהיה מבצע נבחן שעצר במשפטים המופיעים במבחן לאחר נקודת העצירה בה עצר, אילו משפטים אלו הועברו לו. פיתוח האומדנים התבסס על ביצועיהם של 676 הנבחנים בשפה העברית שהשתתפו במחקר החלוץ שנערך בשנת 2016 ושב, לא נעשה שימוש בכללי עצירה, וכל המשפטים במבחן הועברו לכל הנבחנים (למעט מספר מקרים שבהם נבחנים או בוחנים עצרו מיוזמתם את המבחן בשל התעייפות הנבחן, תסכול, או סיבות אחרות).

בעבור כל נקודת עצירה חושבה נוסחת הרגרסיה הלינארית שקשרה בין מספר הטעויות המצטברות עד לנקודת העצירה, לבין מספר הטעויות המצטבר עד סוף המבחן כולו. נוסחה זו שימשה לקביעת האומדנים של מספר הטעויות המצטבר של כל נבחן, בכל נקודת עצירה.

בעבודה הנוכחית תוצג בהרחבה שיטת פיתוח האומדנים, הממצאים השונים, ואופן התיקוף של ההליך בכללותו. לסיים, ידונו השלכות הציינון בשיטה זו, יתרונותיה וחסרונותיה.

קבוצות הטרוגניות חכמות יותר: עדות ממחקר תאומים

מאיר ברנרן

מרכז ארצי לבחינות ולהערכה (מאל"ז)

ליאור אברמזון, אילן יניב, אריאל כנפו-נעם
האוניברסיטה העברית בירושלים

מחקרים רבים בנושא קבלת החלטות מוצאים כי צירוף אומדניהם של כמה שופטים מניב תוצאה מדויקת יותר מאומדנו של שופט יחיד. תופעה זו, המכונה "חוכמת ההמונים", התגלתה במגוון רחב של תחומים, כגון פסיכולוגיה, כלכלה ורפואה. כך, למשל, ממוצע התחזיות של קבוצת כלכלנים מומחים בדבר שיעור האינפלציה נמצא מדויק יותר ממרבית התחזיות של כלכלנים יחידים. גם הערכות המופקות לפי מודלים של מדידה והערכה, כגון תורת המבחנים הקלסית, מושפעות מתופעה זו. כך, למשל, צירוף של שתי הערכות של מטלת הכתיבה בבחינה הפסיכומטרית מדויק יותר מהערכה בודדת.

במחקר זה אנו טוענים כי התועלת שחוכמת ההמונים מביאה לקבוצות המורכבות משופטים שמאפייניהם מגוונים רבה מזו שהיא מביאה לקבוצות המורכבות משופטים שמאפייניהם אחידים. המחקר מתמקד בשני גורמים המאפיינים שונות קבוצתיות. הגורם הראשון הוא התורשה: שופטים שונים זה מזה מבחינה גנטית, ושני בגנים מכתוב שוני בדפוסי חשיבה והתנהגות. הגורם השני הוא גורם הסביבה: שופטים שונים זה מזה במאפיינים כגון חינוך, סביבת מגורים, דת, דעות פוליטיות וכיוצא באלה.

מדוע האומדנים המצורפים של קבוצות מגוונות צפויים להיות מדויקים יותר מאלה של קבוצות אחידות? אנו טוענים כי מגוון הדעות, נקודות המבט ודפוסי המחשבה בקבוצות מגוונות עולה על אלו של קבוצות אחידות. מסיבה זו הדמיון בין אומדניהם של שופטים מקבוצה מגוונת קטן יותר, ולכן צירופם מדויק יותר.

בניסוי שנערך השתתפו 106 זוגות תאומים (70 זוגות תאומים מונוזיגוטיים ו-36 זוגות תאומים טרוזיגוטיים מאותו מגדר). הנבדקים התבקשו לאמוד את מספר הסוכריות בכמה צנצנות. בניסוי זה, שיש בו קריטריון הערכה אובייקטיבי, אפשר לחשב את מידת הדיוק (המרחק מהאמת) של כל צירוף שופטים. צירופי השופטים שנבדקו מנו שני שופטים כל אחד. נערכה השוואה בין שני סוגים של צירופים: צירוף של שני שופטים תאומים (שופטים שהם זוג תאומים) וצירוף של שני שופטים שאינם תאומים (כלומר שופטים הזרים זה לזה).

הממצא העיקרי היה כי ללא קשר לסוג התאומים, צירוף של שני שופטים תאומים הניב אומדן מדויק פחות מצירוף של שני שופטים שאינם תאומים. ניתוחי המשך השוו בין שני סוגי התאומים. מודלים קלסיים של חקר תורשה וסביבה מניחים כי הסביבות שתאומים משני הסוגים גדלים בהן דומות זו לזו. עם זאת, רק תאומים מונוזיגוטיים נושאים אותו מטען גנטי בדיוק, ואילו תאומים טרוזיגוטיים חולקים חצי מהמטען הגנטי שלהם זה עם זה. לכן ההשוואה בין תאומים מונוזיגוטיים ותאומים טרוזיגוטיים עשויה ללמד על השפעת השונות הגנטית על כל תופעה. נמצא כי בקרב התאומים המונוזיגוטיים, ההשוואה בין אומדני השופטים התאומים לאומדני השופטים שאינם תאומים הייתה מובהקת ואילו בקרב התאומים הטרוזיגוטיים ההשוואה לא הייתה מובהקת.

ממצאי המחקר עולה כי שונות קבוצתית בכלל, ואולי אף שונות גנטית בפרט, חשובה להחלטות קבוצתיות. ככל שמאפייני השופטים בקבוצה שונים יותר זה מזה, כך האומדנים שהקבוצה מפיקה מדויקים יותר. בעולם שמחוץ למעבדה אפשר להגדיל את מידת השונות באמצעות בחירת שופטים מרקעים מגוונים מבחינת מוצא, מגדר, גיל, סביבה חינוכית, דת, דעות פוליטיות וכיוצא באלה.

קרבה פסיכולוגית משפחת ביצועים קוגניטיביים: עדויות מהבחינה הפסיכומטרית

ברית הדר

אוניברסיטת תל אביב, המחלקה לפסיכולוגיה

ספי פומפיאן, צור קרליץ
מרכז ארצי לבחינות ולהערכה (מאל"ו)

נירה ליברמן

אוניברסיטת תל אביב, המחלקה לפסיכולוגיה

במחקר הנוכחי בחנו כיצד מרחק פסיכולוגי משפיע על ביצועים קוגניטיביים ובאופן ספציפי, בקונטקסט של הבחינה הפסיכומטרית. מרחק פסיכולוגי הוא רצף (קצה אחד "קרוב" וקצה שני "רחוק") אשר מציין עד כמה אובייקט או אירוע רחוקים מהחוויה של אדם באותו הזמן ואותו המקום. מחקרים קודמים מצאו שמרחק פסיכולוגי משפיע על אופן העיבוד הקוגניטיבי כך שאובייקטים רחוקים מיוצגים מנטאלית באופן אבסטרקטי ואובייקטים קרובים באופן קונקרטי, אך לא ידוע כיצד מרחק פסיכולוגי משפיע על ביצועים קוגניטיביים. במחקר זה ערכנו שלושה ניסויים בהם תפעלנו מרחק פסיכולוגי (למשל, השאלה נשאלה על התרבות המעקודית העתיקה או על ישראל של ימינו), ומדדנו ביצוע בשאלות הסקה מילולית אשר לקוחות מהבחינה הפסיכומטרית. שיערנו שפתרון השאלות יהיה מדויק יותר תחת התנאי של קרבה פסיכולוגית לעומת מרחק פסיכולוגי. כמו כן, לחצי מהשאלות הוספנו גם מידע שאינו רלבנטי לפתרון כדי לבחון האם קרבה פסיכולוגית תעזור להתעלם ממידע זה. בניסוי הראשון, 129 נבדקים פתרו שאלות הסקה מילולית בתנאי מעבדה. בניסוי השני, 871 נבחנים פתרו שאלות הסקה מילולית כחלק אינטגרלי מהבחינה הפסיכומטרית. בניסוי השלישי, 387 נבדקים פתרו שאלות הסקה מילולית באינטרנט. בשלושת הניסויים מצאנו שקרבה פסיכולוגית שיפרה את הביצוע בשאלות ההסקה המילולית (בניסוי השני רק עבור נבחנים עם ציון פסיכומטרי נמוך מהמוצע). כמו כן, מצאנו שהוספת מידע לא רלבנטי פגעה בביצוע באופן כללי, ללא אינטראקציה עם מרחק.

הערכה רב-ממדית של כדורגלני נוער עם התייחסות לעמדת המשחק

אורן אקהויז, גיל גולדצויג
מכללה אקדמית תל אביב-יפו

תהליכי הערכת הכישרונות הנוכחיים בענף הכדורגל ניתנים לאפיון בחוסר קונצנזוס בולט ומסתמכים בעיקר על בסיס הערכות סובייקטיביות של מאמנים וסקאוטרים. אלו, מושפעים מלחצים שמופעלים על הגורמים המקצועיים להגיע להישגים קצרי טווח.

מטרת המחקר הנוכחי הייתה לספק תמונת מצב תיאורית של כדורגלני נוער, המשחקים בעמדות משחק שונות. לצורך המחקר, נבנה מערך הערכה רב-ממדי, המשלב מדידה של נתונים פיזיים; שאלון ביוגרפי [מקורי], המודד תחושת סיפוק מהעמדה אליה משויכים, הערכה עצמית, מודלים לחיקוי ומאפיינים מנטאליים; שאלון הערכה עצמית של יכולות פיזיות וטכניות [מקורי]; ושאלון מיומנויות מנטאליות (OMSAT-4) [תורגם לעברית בנוהל סטנדרטי]. המחקר נערך בקיץ 2016 והתבסס על השתתפות התנדבותית של 104 כדורגלנים בגילאים 13-16, משלושה מועדוני כדורגל.

הממצאים הצביעו על קבלה וסיפוק מעמדות המשחק (79.9% הציגו שיעורי סיפוק גבוהים) וכן על הבדלים פיזיים בין שחקנים מעמדות שונות (שוערים גבוהים וכבדים יותר באופן מובהק, ובלמים נוטים להיות גבוהים וכבדים יותר בהשוואה לקשרים). עוד נמצאו, הבדלים מובהקים במרבית הדירוגים העצמיים של היכולות הטכניות בין שחקנים מעמדות שונות: כדורגלנים בעמדות הגנתיות נטו לדרג עצמם גבוה יותר ב-"אלמנטים הגנתיים"; כדורגלנים בעמדות קישור נטו לדרג עצמם גבוה יותר ב-"אלמנטים מקשרים"; כדורגלנים בעמדות התקפה נטו לדרג עצמם גבוה יותר ב-"אלמנטים התקפיים". באשר למיומנויות מנטאליות, לא נמצאו הבדלים מובהקים רלוונטיים. יחד עם זאת, כשהתבקשו המשתתפים לשייך לעצמם מאפיינים מנטאליים (מתוך מבחר אופציות), נמצאו מספר מגמות; לפיהן: בלמים נטו לדרג "מנהיגות", קשרים מרכזיים נטו לדרג "ראיית משחק" ו-"יצירתיות", וחלוצים נטו לדרג "תנועתיות".

ממצאי מחקר זה מצביעים על נטייה להצבת שחקנים גבוהים וכבדים יותר בעמדות השוער והבלם, בניגוד להצבת שחקנים נמוכים וקלים יותר בעמדות המגן החיצוני, הקשר המרכזי והקשר החיצוני. נטייה זו מייצגת מגמה מוכחת אשר קיבלה תימוכין במחקרים שונים שנערכו במקומות שונים בעולם. עוד ניתן להסיק, כי התפישה של השחקן את יכולותיו הטכניות ומאפייניו המנטאליים קשורה לעמדת המשחק שלו. התייחסות השחקנים לאלמנטים מסוימים כאל יתרונות או חסרונות יחסיים, עולה בקנה אחד עם מה שמצופה מהם במסגרת השתייכותם לעמדת משחק ספציפית.

לדוגמא: שחקן הגנה, אשר שויך לעמדתו על סמך יתרון גובה, מדרג עצמו גבוה ביכולות התיקול והנגיחה ובה בעת רואה ביכולות המסירה והסיומת כמגרעות יחסיות. בנוסף, הוא נוטה לשייך לעצמו מאפיינים מנטאליים כמו "אגרסיביות" ו-"מנהיגות" ופחות "יצירתיות" ו-"חזון".

על כן, ניתן לשער כי שיוכם החד משמעי של הכדורגלנים לעמדות משחק, בגילאים מוקדמים, עשוי ליצור הטיות שעלולות למנוע מהם להתפתח באופן כוללני (הוליסטי). זאת, מתוך הבנה שתחושת המסוגלות העצמית משפיעה על הביצוע הסופי.

תהליך הערכה מובנה, כדוגמת זה שהמחקר מציג, יאפשר שילוב של מדדים סובייקטיביים, כמו הערכות מאמנים ועמיתים, עם מדדים אובייקטיביים כמו מדדים טכניים של יכולת ביצוע. פיתוח של סוללת הערכה סטנדרטית יכול ליעל את תהליכי האיתור וההכשרה של כדורגלנים צעירים ולאפשר התפתחות טובה יותר של השחקנים.

אנשי קשר ומציגים בכינוס אפי 2019

שם	דוא"ל	
אייל גמליאל	eyalg@ruppin.ac.il	יו"ר אפי
צור קרליץ	tzur@nite.org.il	יו"ר אפי הנכנס
עינת נוטע-קורן	einatno@openu.ac.il	יו"ר הכינוס
אסטלה מלמד	estela.melamed@gmail.com	
יו"ר - ענת בן-סימון	anat@nite.org.il	מושב א' כלי מיון ועדויות לתקפותם
אייל גמליאל	eyalg@ruppin.ac.il	
ג'ודי גולדנברג	judyg@zahav.net.il	
צור קרליץ	tzur@nite.org.il	
יו"ר - חגית גליקמן	hglickman.rama@education.gov.il	מושב ב' 1 אתגרים ודילמות במבחנים
חנן גולדשמידט	cgoldsch@idc.ac.il	
יונתן סער	yonatan@nite.org.il	
תמר קנת-כהן	tami@nite.org.il	
עדו גל	iddo@research.haifa.ac.il	
יו"ר - ספי פומפיאן	yoel@nite.org.il	מושב ב' 2 הערכת הבעה בכתב במבחנים רחבי היקף בישראל
	sephi@nite.org.il	
אבי אללוף	avi@nite.org.il	מושב ג' 1 שולחנות עגולים
ריטה סבר	ritasever9@gmail.com	
יו"ר - ענת קדם	anatk@cet.ac.il	מושב ג' 2 כישורים ודרכים לחידוד מדידתם
יואל רפ	yoel@nite.org.il	
מאיר בררון	meir@nite.org.il	
ברית הדר	britt.hadar@gmail.com	
אורן אקהויז	orenekhoiz@gmail.com	
מארגנת הכינוס - שרה שפירא	ispaisrael@gmail.com	ארגון
דורית דהן	dorit@nite.org.il	
מירי פוקס	mirif@nite.org.il	
מור פלולי	events@SHALVA.org	

כ'א/א

כ'א/א

תלובות והצעות
בנושא האגודה הישראלית לפסיכומטריקה
נא להפנות אל:

שרה שפירא, האגודה הישראלית לפסיכומטריקה ispaisrael@gmail.com

אתר אפי www.ispa.org.il