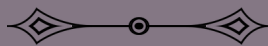


אפי | جاس | ISPA

אגודה ישראלית לפסיכומטריקה  
الجمعية الإسرائيلية للبيكومتريكا  
ISRAELI PSYCHOMETRIC ASSOCIATION  
מבחנים ■ מדידה ■ הערכה

הכינוס ה-17  
של  
האגודה הישראלית  
לפסיכומטריקה

י"ד-ט"ו בשבט תשפ"א 27-28.1.2021



כנס מקוון



2021

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

יום ד' 27.1.21

פתיחה יו"ר אפי: צור קרליץ עדכון מוועדת "מיהו פסיכומטריקא": ריטה סבר פתיחה יו"ר הכינוס: גיל גולדצויג ואימאן עואדיה	14:00 – 14:25
<b>מושב א'</b> סימפוזיון: מיון מקוון בפקודה: מעבר למיון מרחוק בצה"ל בתקופת הקורונה יו"ר: יאיר נעם	14:25 – 15:50
<ul style="list-style-type: none"> <li>■ המעבר להערכה מקוונת במיון התעסוקתי למועמדות לצה"ל (יום המא"ה) נמרוד בן-סניור</li> <li>■ מבד"ק מרחוק רביד אור להב, נטע ברזילי</li> <li>■ מיון מרחוק בחיל האוויר – מאילוץ להזדמנות דנה שושני שנקר</li> </ul>	
<b>הפסקה</b>	15:50 – 16:05
<b>מושב ב'</b> הרצאת אורח <b>The Use of Content-Referencing to Evaluate the Magnitude of Student Growth</b> Prof. Derek C. Briggs, University of Colorado, Boulder	16:05 – 17:05
*ההרצאה תועבר באנגלית	

יום ה' 28.1.21

פתיחה יו"ר הכינוס: גיל גולדצויג ואימאן עואדיה עדכון על לימודי התעודה בפסיכומטריקה: אבי אללוף	14:00 – 14:10
<b>מושב א'</b> סימפוזיון: הפקת מדדים משמעותיים מנתונים רועשים: פוטנציאל ואתגרים בשימוש בנתונים מסביבות למידה אותנטיות למדידה והערכה יו"ר: מירב אריאלי אטאלי	14:10 – 15:25
<ul style="list-style-type: none"> <li>■ זיהוי שיפור בשטף בעזרת ניתוח נתונים רועשים מקריאה קולית ממושכת שלא במסגרת מבחן ביאטה ביגמן-קלבנוב</li> <li>■ אומדן יכולות ומיומנויות לומדים בסביבות אינטראקטיביות המשלבות פריטים עם רמזים מירב אריאלי-אטאלי</li> <li>■ על האמינות של אנליטיקות למידה מקורסים מקוונים המוניים פתוחים (MOOCs) – זיהוי וניקוי של הטיות הנגרמות בגלל חשבונות מזויפים גיורא אלכסנדרון</li> </ul>	
<b>הפסקה</b>	15:25 – 15:35

**מושב ב'** (מושב מקביל) **השכלה גבוהה, נטיות דמוגרפיות וחברתית** 15:35 – 16:50  
 יו"ר: אימאן עואדיה

- **תוקף בחינת המתא"ם למיון לתארים מתקדמים בפסיכולוגיה**  
 מאיר ברנרון, תמר קנת-כהן, נועה גזית
- **אתגרים בצמצום הנשירה מההשכלה הגבוהה: בדיקת אפקטיביות של התערבות אינטרנטית**  
 מיכל פיליפס-ברנשטיין, איתמר גתי
- **התשואה לפסיכומטרי ולבגרות במונחי שכר עתידי**  
 כרמל אורן, דוד מעגן, יואל רפ

**מושב ג'** (מושב מקביל) **סוגיות בהערכה, העברה מרחוק ושימוש מאסיבי**  
 יו"ר: צור קרליץ

- **מידת הטיות בהישגי לומדים ממדינות מתפתחות בקורסים מקוונים המוניים פתוחים**  
 (Massive Open Online Courses) בעזרת תיאוריית התגובה לפריט  
 סער קרפ גרשון, גיורא אלכסנדרון
- **ימי קורונה קובעים עובדות בשטח – העברת בחינה חורצת גורל בהשגחה מרחוק:**  
 המקרה של בחינת הבקיאיות בערבית למועמדים ללמוד במכללות להוראה  
 חוה נוימן, תהילה סנד, יארא כריים, מאיר ליאני, מיכל שקד, ורוניקה ספיץ'
- **יישומים של ניתוח שפה טבעית להערכה אוטומטית של שאלות פתוחות במדעים**  
 מוריה אריאלי, טניה נזרצקי, גיורא אלכסנדרון

**הפסקה** 16:50 – 17:00

**מושב ד'** (מושב מקביל) **סימפוזיון: אוריינטציות של צעירים להשכלה גבוהה** 17:00 – 18:15  
 יו"ר: איתמר גתי

- **פיתוח שאלון אוריינטציות להשכלה הגבוהה**  
 תרצה וילנר
- **תוקף מתכנס של שאלון אוריינטציות להשכלה גבוהה**  
 יוליה ליפשיץ-ברזילרי
- **אוריינטציות להשכלה גבוהה בקרב סטודנטים ערבים**  
 שאדה קשקוש

**מושב ה'** (מושב מקביל) **פיתוח כלים ואמידת תכונות**  
 יו"ר: גיל גולדצויג

- **שימוש במדידה סובייקטיבית (דרך דיווח עצמי) לאומדן מהיר של יכולות מתמטיות של**  
 אנשים בוגרים: סוגיות ואתגרים  
 עדו גל
- **TO ERR IS HUMAN – פיתוח ותיקוף ראשוני לשאלון לאומדן מועדות לתאונות**  
 (Accident Proneness)  
 אירן דיאמנט
- **ניתוח עיסוקים של תפקיד הרופא המנתח – הגדרת התכונות והכישורים הנדרשים מרופאים**  
 מנתחים לצורך פיתוח מערכת מיון להתמחויות  
 נועה גזית, גלעד בן גל, רון אלישר

## סימפוזיון

# מיון מקוון בפקודה: מעבר למיון מרחוק בצה"ל בתקופת הקורונה

יאיר נעם

ענף פיתוח ומערכות מיון, צה"ל

צה"ל מהווה ארגון יחודי בחברה הישראלית כמעצב זהויות של אזרחים רבים. תחום המיון במערך מדעי ההתנהגות (ממד"ה) אחראי על תפעול, מחקר ופיתוח של תהליכי המיון המתקיימים בכניסה לצה"ל וליחידותיו המגוונות. תחום זה מבוסס על פסיכולוגים תעסוקתיים.

הסגר והגבלות התנועה של תקופת הקורונה איתגרו את היכולת של צה"ל להמשיך למיון מועמדים באופן שוטף. מקבץ הרצאות זה נועד להמחיש את פסיפס העשייה בתחום במיונים השונים בצה"ל. סימפוזיון זה נועד להציג את השינויים והממצאים שנבעו מכך ממספר יחידות בצה"ל: המיון הכלל צה"ל, המיון המטכ"ל לקצונה, מיון באגף המודיעין ובחיל האוויר.

## המעבר להערכה מקוונת במיון התעסוקתי למועמדות לצה"ל (יום המא"ה)

נמרוד בן-סניור

מדור מועמדים לשירות ביטחון, צה"ל

ההרצאה תעסוק במיון הכלל צה"ל ובאופן פרטני **את יום המא"ה** ומעברו למיון מקוון. מיון יום המא"ה (מיון, איתור וההשמה) כולל הערכה על גבי 13 סביבות של מועמדות לצה"ל במספר תחנות (סימולציות, תרגילי מצב ומבחנים ממוחשבים). יום המא"ה מתקיים מזה שנים במיקור חוץ. במהלך חודשי מרץ-ואפריל של שנת 2020 מדור מלש"ב, בשיתוף עם צוות מקצועי של חברת אדם מילא, פיתחו חלופה מקוונת. מאז, עשרות-אלפי מועמדות ביצעו את יום המא"ה המקוון ועל בסיס נתוניהן נערכו מחקרים לבחינת תפקודו בהשוואה לעבר. בהתאם לממצאים החיוביים שהתקבלו במחקר התקבלה החלטה לעבור באופן קבוע ליום מיון מקוון, תוך שיפור מתמיד.

## מבד"ק מרחוק

רביד אור להב, נטע ברזילי

מדור מבדקי קצונה, צה"ל

ההרצאה תעסוק **במדור מבדקי הקצונה**, שהינו התחנה הפסיכולוגית-אישיותית בתהליך המיון לקצונה בצה"ל. עם מעבר למיון מקוון, המועמדים מבצעים את המבחנים כל אחד מביתו או מיחידתו, דרך המחשב או מהטלפון, ועוברים ראיון פסיכולוג בצורה מקוונת. ההצגה תכלול בחינת הממצאים המחקריים שנערכו על התפלגות המדדים - תוך השוואה למיון הפרונטאלי, יכולת הניבוי של המיון המקוון, השפעה של הפלטפורמה (פלאפון בהשוואה למחשב) על הצלחה במיון ועוד. לסיכום נעסוק ביתרונות שבמיון מרחוק (רציפות תפקודית, חיסכון בזמן ומשאבים, התאמה לרוח התקופה ולדור המועמדים), לצד הקשיים והמגבלות שבכך.

## מיון מרחוק בחיל האוויר - מאילוץ להזדמנות

דנה שושני שנקר

מדור מו"פ למיון צוות אוויר, צה"ל

נציג את המיון מרחוק **בחיל האוויר**. ההרצאה תתמקד בשניים מהתהליכים בהם יושמו שינויים בתקופה זו: האחד הוא במסגרת המיון לצוות אוויר והשני במיון למקצועות התעופה (מדריכות סימולטור ועוד). בצוות אוויר - היה צורך לעבור לראיון מקוון, ממצאים בנושא יוצגו בהרצאה. במיון למקצועות התעופה - השלב הראשוני של מבחני הכשרים עבר לביצוע מרחוק. ממצאים מחקרניים ראשוניים של שינוי זה יוצגו גם כן.

## The Use of Content-Referencing to Evaluate the Magnitude of Student Growth

Professor Derek C. Briggs

School of Education, University of Colorado, Boulder

When students are given a test on more than one occasion on the same topic, it is natural to want to know how much they have grown, and to equate this growth to learning. But this is a lot harder than it appears because such evaluations require inferences about magnitudes, and inferences about magnitudes require well-defined and readily interpretable units of measurement. It has become conventional for researchers to interpret test score changes in terms of variability (e.g., standard deviation units) or some implicit function of time (e.g., weeks or months of “learning”). In this presentation I will briefly cover the historical and conceptual precedent for such moves, which can be traced back to the work of Gustav Fechner and Francis Galton between 1860 and 1890, Alfred Binet between 1904 and 1911, and Louis Thurstone between 1925 and 1932. I will then introduce a somewhat novel approach to interpreting growth magnitudes that rests upon a combination of cognitive theory, item mapping, and practical relevance. I call this a content-referenced approach to the measurement of growth.

## סימפוזיון

# הפקת מדדים משמעותיים מנתונים רועשים: פוטנציאל ואתגרים בשימוש בנתונים מסביבות למידה אותנטיות למדידה והערכה

מירב אריאלי אטאלי  
המרכז לטכנולוגיה חינוכית (מטח)

נתוני מדידה הם לעיתים קרובות נתונים "נקיים" במובן זה שאנו אוספים תגובות של נבחנים על משתנים מוגדרים היטב שבהם אנו מעוניינים. אנו מגדירים את "התכונה הנמדדת", מנסחים דרכים ספציפיות ומבוקרות למדידת התכונה ורק אותה, ואז אוספים את הנתונים תחת תנאים מבוקרים סטנדרטיים, ומבטיחים במידת האפשר שלא ייכנס רעש למדידה בדמות משתנים שאינם רלוונטיים.

אולם רוב הנתונים בעולם, כולל נתונים חינוכיים, כלומר, נתונים שנאספו במסגרת תהליך למידה, כמעט אף פעם אינם "נקיים" כל כך, ולעיתים קרובות קשה להפיק מתוכם מדדים העומדים בסטנדרטים פסיכומטריים. עם המעבר ללמידה דיגיטלית יותר בשנים האחרונות, ועוד יותר בחודשים האחרונים בגלל הקורונה, ישנם נתונים רבים זמינים שנאספו בדרך זו, ויש הזדמנות ללמוד ולהשתמש בהם. השאלה היא האם אנו יכולים לחלץ מדדי למידה רלוונטיים ומשמעותיים שהם "מספיק טובים" פסיכומטריים להערכה מעצבת או אולי כאלו שיוכלו אפילו להחליף הערכה מסכמת בהקשרים מסוימים.

בסימפוזיון זה אנו רוצים להציג ולדון בדרכים שונות להפקת מידע מדידה שימושי ומשמעותי מנתונים שבמקורם לא נועדו להערכה ומדידה.

נציג שלושה מחקרים המדגימים גישות שונות להתמודדות עם מקורות "רעש" שונים והמציעים פתרונות ישימים.

בהרצאה הראשונה תוצג דרך להפקת מדד שטף קריאה בעל-פה (קריאה קולית) יחד עם עדויות לשיפור בשטף הקריאה, מתוך נתונים שהושגו במסגרת פעילות קריאת ספרים עצמאית ממושכת עם טקסט אותנטי בתנאי כיתה רגילים בבית ספר יסודי, שכתוצאה מכך היו הנתונים רווי רעש התנהגותי וטכני.

ההרצאה השנייה תציג דרך להפקת מדד יכולת בהקשר של אפליקציה ללימוד שפה שמאפשרת לענות על פריטים תוך כדי שימוש ברמזים ובין השאר מעודדת התנהגויות תלויות משתנים בינאישיים שלא דווקא קשורים ליכולת, ותדון במספר מודלים להערכת יכולת הלומדים למרות רעש המדידה והשונות שעשויה להיראות לא רלוונטית.

ההרצאה השלישית תוסיף נקודת מבט כללית יותר על התנהגויות הלומדים בפלטפורמות הלמידה, תוך שימוש בנתונים מתוך קורסים מקוונים המוניים פתוחים (MOOCs). קורסים אלו פותחו לשירות תלמידים ומורים והם מאפשרים חופש ושונות רבה בשימוש. הרצאה זו תתמקד בהתנהגות רמייה – היקף התופעה, זיהויה והערכתה, השפעותיה על למידה מהנתונים, ובאופן ספציפי כיצד היכולת "לנקות" נתונים מזויפים אלה מסייעת להשגת מדדים משמעותיים יותר מהנתונים.

נחתום את הסימפוזיון בדיון על ההזדמנויות החדשות בהשראת עבודות אלו והאתגרים שהזדמנויות אלו מביאות איתן.



## סימפוזיון

# הפקת מדדים משמעותיים מנתונים רועשים: פוטנציאל ואתגרים בשימוש בנתונים מסביבות למידה אותנטיות למדידה והערכה זיהוי שיפור בשטף בעזרת ניתוח נתונים רועשים מקריאה קולית ממושכת שלא במסגרת מבחן

ביאטה בייגמן קלבנוב

Educational Testing Service, Princeton, NJ ETS

במערכת החינוך הפורמלית הערכה של למידה מתבצעת לרוב באמצעות מדידות חוזרות עם מבחנים שפותחו במיוחד לצורך זה. זהו המצב עם מדידות שטף קריאה קולית (Oral Reading Fluency) בבתי ספר יסודיים בארה"ב. במחקר זה אנו מראים כי ניתן לזהות שיפור בשטף על ידי ניתוח נתונים של קריאה עצמאית של ילדים בספר קריאה באפליקציית Relay Reader. זה מקרב אותנו לחזון של הערכה מתוך "קריאה לשם קריאה", לא לשם מבחן, עם יכולת לזהות שיפור בשטף הקריאה ללא התערבות ייעודית לצורך הערכה.

Relay Reader נועדה להקל על הגישה של הקורא המתפתח לספר באורך מלא על ידי קריאה לסירוגין, בה הילד והקריין (audiobook) מקריאים קטעים עוקבים מן הספר. הקטעים להקשבה ולהקראה מחושבים מחדש באופן דינמי בכל יום קריאה, בהתבסס על הגדרות המשתמש (מספר היעד של המילים בתור ההקשבה ובתור ההקראה). הקריאה הקולית של הילד מוקלטת ונשלחת לניתוח הכולל חישוב של שטף כמספר מילים שהוקראו נכון לדקה (Words Correct Per Minute), או (WCPM) באמצעות מערכת אוטומטית שפיתחנו.

המחקר שיוצג מבוסס על נתונים שנאספו מ-42 תלמידים בשתי כיתות (ד' ו-ה') בבית ספר במדינת ניו ג'רזי בארה"ב, שקראו את הספר "הארי פוטר ואבן החכמים" במשך כ-20 דקות ביום, שלוש פעמים בשבוע, במהלך סמסטר הסתיו של 2018. כל התלמידים סיימו את הספר; 95% עשו זאת תוך 15 שבועות. הילדים השתמשו באוזניות פשוטות עם מיקרופון מובנה וקראו בכיתתם או במסדרון הסמוך, על פי הוראות המורה.

שאלת המחקר היא האם ניתן להשתמש בנתונים אלה כדי לזהות שיפור בשטף הקריאה. כדי למזער את השפעת השינויים בזיהוי האוטומטי של הדיבור, התחלנו את הניתוח עם תמלולים ידניים של חלק מהנתונים, ואז חזרנו על הניתוח עם מדידות אוטומטיות.

השתמשנו באינדקס פרקים כדי למדל זמן; בספר 17 פרקים, באורך ממוצע של כ-4,500 מילים לפרק. לצורך התמלול התמקדנו ב-35 תלמידים שקראו עם הגדרות ברירת המחדל. בחרנו 68 קטעים - 4 בכל פרק, ובסה"כ 1,556 הקלטות; 2.6 הקלטות לפרק לקורא, בממוצע.

לנתונים אלו התאמנו מודל לינארי מעורב (Mixed Linear Model): (א) WCPM הוא המשתנה התלוי; (ב) קטעים וקוראים הם גורמים משתנים (random factors); (ג) כיתה ושני משתנים טקטואליים כגורמים קבועים (fixed factors); (ד) מספר פרק כגורם קבוע עם שיפוע משתנה. התקבל אפקט מובהק למקדם הלינארי החיובי של מספר הפרק: בכל פרק, ממוצע WCPM צפוי לעלות ב-1.28 מילים. לאחר מכן התאמנו מחדש את המודל ל-6,814 הקלטות (של 1,207 קטעים שונים) מכל 42 הילדים עם הערכות WCPM אוטומטיות. התוצאות היו דומות, עם עלייה צפויה של 1.12 מילים לפרק.

התוצאות מראות כי מעקב לא פולשני אחר שטף קריאה בזמן קריאת ספר עם Relay Reader הינו אפשרי בשל מספרן הרב של התצפיות שנאספו עבור כל קורא, וזאת למרות חוסר שליטה בטקסט, רעש סביבתי והתנהגותי משמעותי, ושינויים של המערכת האוטומטית לזיהוי דיבור. לפיכך, אנו מתקבלים לחזון של קריאה לשם קריאה, לא לשם מבחן, עם יכולת למדוד ולהעריך למידה. השלבים הבאים כוללים ניסויים עם ספרים אחרים והעברה (transfer) מספר לספר, וכן חקירת קשרים בין שטף כפי שנמדד כאן עם אומדנים אחרים של שטף ועם מיומנויות קריאה אחרות, כגון הבנה.

## סימפוזיון

# הפקת מדדים משמעותיים מנתונים רועשים: פוטנציאל ואתגרים בשימוש בנתונים מסביבות למידה אותנטיות למדידה והערכה אומדן יכולות ומיומנויות לומדים בסביבות אינטראקטיביות המשלבות פריטים עם רמזים

מירב אריאלי אטאלי  
המרכז לטכנולוגיה חינוכית (מטח)

אחת ממטרות הערכה מעצבת היא לעזור למורים להתאים את ההוראה ללומדים כמו גם לעזור ללומדים להיות מודעים ליכולות ולידע שלהם ובכך להתאים או לבחור את הצעד הבא הדרוש ללמידה. סביבת למידה דיגיטלית אינטראקטיבית אוספת נתונים כברירת מחדל, כחלק אינטגרלי מתהליך העבודה של הלומדים וללא צורך בפעולות אקטיביות נוספות. האם אנו יכולים לנצל את אסוף הנתונים האוטומטי הזה ולהשתמש בו כדי להעריך את יכולתם של הלומדים לצורך הערכה מעצבת מבלי שהם יצטרכו להפסיק את הלמידה כדי לעבור מבחן? עד כמה הנתונים שנאספים כברירת מחדל יכולים להעיד על יכולת? האם וכיצד הנתונים ההתנהגותיים (למשל, שימוש ברמזים) מפריעים או משפיעים על אומדני היכולת?

הפקת ציון או מדד יכולת מסביבת למידה הינה מאתגרת מכיוון שברוב המקרים סביבת הלמידה כוללת מאפיינים שהותאמו ללמידה, ולא בהכרח הותאמו למדידה. מאפיינים אלה עשויים לכלול משוב או מערכת תגמול, או אפשרויות בחירה, כגון ניסיונות תשובה מרובים על שאלות, שימוש בעזרה או ברמזים לפני מענה לשאלה, ועוד כהנה התנהגויות שיכולות להוסיף "רעש" למדידה. במחקר זה אנו מתמקדים בסביבת למידה שמאפשרת שימוש ברמזים ומציעים מודל להפקת אומדן יכולת למרות תוספת השונות שעשויה להיראות לא רלוונטית.

לצורך פיתוח ובחינת מודלים אפשריים, השתמשנו בנתונים של Duolingo, פלטפורמת לימוד שפה מקוונת עם יותר מ-200 מיליון משתמשים רשומים. קורס דואולינגו מאורגן בשיעורים, שבכל אחד מהם יש סט שאלות (= פריטים) שלומדים עונים עליהן ומקבלים משוב נכונות מידי. בחלק מסוגי הפריטים יש רמזים, המופיעים כאופציה של רחף מעל המילה/ים (hover-over). בסוג פריט זה, ריחוף מעל מילה במשפט פותח חלון עם התרגום של מילה זו. הלומדים יכולים לרחף מעל מילה אחת או יותר, ואפילו לקבל את התרגום של כל המילים במשפט. בקשת רמז יכולה להיות תלויה ביכולת או בידע של הלומדים, אך יכולה להיות מושפעת גם מגורמים אחרים מלבד יכולת. מבחינת הלומד, תשובה נכונה לפריט מזכה בנקודה, ללא קשר אם נעשה שימוש באופציית הרחף כדי לחשוף מילים מתורגמות.

המטרה שלנו היתה להשתמש במידע לגבי שימוש ברמזים כמידע נוסף על היכולות או הכישורים הנמדדים, גם אם לא בצורה ישירה. בדקנו כמה מודלים שבהם השימוש ברמז ברמת הפריט (כן או לא) נכלל בדרכים שונות בחישוב הציון לפריט (כלומר, מודל משותף של נכונות ושימוש ברמז), וגם כשלוש חשבון תכונה סמויה נוספת של "נטייה להשתמש ברמז". אנו השתמשנו במודלים ממשפחת IRT. את פיתוח המודלים השונים ביצענו על מאגר נתונים אחד, ולאחריו בדקנו ישום ובחירת המודל המתאים במאגר נתונים מקביל. מאגר הנתונים הראשון היה קורס לימוד ספרדית לדוברי אנגלית (כלומר ספרדית-מאנגלית) והשני היה קורס לימוד אנגלית לדוברי פורטוגזית (כלומר אנגלית-מפורטוגזית).

בהצגה זו נסביר את הרציונל למודלים השונים ונראה איזה מהמודלים נמצא מתאים יותר לנתונים, וכיצד המודל הנבחר מוסיף להבנתנו את תהליך הלמידה והמדידה. נדון בסוגיית תקפותם של אומדני יכולות אלו ובתועלת הפוטנציאלית שלהם בתהליך הלמידה וההערכה.

## סימפוזיון

# הפקת מדדים משמעותיים מנתונים רועשים: פוטנציאל ואתגרים בשימוש בנתונים מסביבות למידה אותנטיות למדידה והערכה על האמינות של אנליטיקות למידה מקורסים מקוונים המוניים פתוחים (MOOCs) - זיהוי וניקוי של הטיות הנגרמות בגלל חשבונות מזוייפים

גיורא אלכסנדרון  
מכון ויצמן למדע

MOOCs (מוקס, וביחיד מוק) משנים את הנוף החינוכי ומציעים גישה לתכנים לימודיים באיכות גבוהה לכל מי שהוא בעל חיבור לאינטרנט. בעת הלמידה במוק, הלומדים משאירים "עקבות דיגיטליים" מפורטים הנשמרים בקבצי יומן (log files). עם יותר מ-100 מיליון לומדים גלובליים, עקבות אלו מייצרים נתוני עתק המספקים הזדמנות חסרת תקדים לחקור תהליכי למידה בסביבות דיגיטליות תוך שימוש בשיטות חישוביות מתקדמות, ולפתח מנגנונים שיסייעו לענות על הצרכים הייחודיים של אוכלוסיות לומדים שונות.

עם זאת, הנתונים שנוצרים על ידי קהל הלומדים ההטרוגני והאנונימי מכילים 'רעש' רב. בין היתר, רעש זה הוא תוצאה של משתמשים המנצלים לרעה את האנונימיות והאופי הפתוח של המוקים כדי לשפר את ציוניהם, תוך שימוש בטכניקות בלתי לגיטימיות כגון הקמת חשבונות מרובים המופעלים על ידי משתמשים מזוייפים. חשבונות אלה משמשים ל-'קצירת' פתרונות נכונים למטלות המקנות ניקוד, ושימוש בפתרונות הני"ל לקבלת קרדיט (לרוב, תעודת סיום) בחשבון הראשי של המשתמש. בחמש השנים האחרונות חקרנו נושא זה בהרחבה, תוך התמקדות בשיטות לאיתור התנהגויות כאלו, הערכת שכיחותן והשפעתן על מדדים שונים, ודרכים למונען.

אלגוריתמי הגילוי שלנו זיהו כי במוקים המוקדמים שהוצעו על ידי MIT (המכון הטכנולוגי של מסצ'וסטס) על פלטפורמת edX, נושא זה אכן היה משמעותי מאוד (בהערכה גסה, כ-10% ממקבלי התעודות השתמשו בחשבונות כפולים). כפי שהדגמנו, התנהגות לומדים אלה פעלה כ-'רעש' שהשפיע על ניתוח למידה ומודלים סטטיסטיים בדרכים שונות. לדוגמה, מכיוון שחשבונות שהיו מעורבים בפעילות כזו הפגינו ביצועים גבוהים מבלי ליצור אינטראקציה משמעותית עם תכני הלמידה בקורס (למשל, הם נטו לדלג על סרטונים ודפי הסבר), אלגוריתמים של למידת מכונה זיהו תכנים שונים כחסרי ערך לימודי, וזיהו מסלולי למידה המדלגים מעליהם כאפקטיביים מבחינה פדגוגית וככאלו שרצוי להמליץ עליהם ללומדים עתידיים. בנוסף, הביצועים הגבוהים של הלומדים הללו ודפוסי התגובה החריגים שלהם היטו מודלים של IRT והורידו את אומדן היכולת של שאר הלומדים במוק (בממוצע, הפחיתו את ציון התקן של הלומדים האחרים בכ-0.3).

אמנם רעש כזה עשוי להשפיע על אמינותם של מודלים סטטיסטיים שונים, אך למעשה מחקר זה הגביר בסופו של דבר את אמוננו ביכולת להפיק תובנות משמעותיות מנתוני למידה בסביבות מקוונות ובפרט במוקים. הסיבה פשוטה - בניגוד לסביבות למידה 'מסורתיות' (לא מקוונות) בהן קיומו של הרעש ידוע, אך מדידה ואפיון מדויק שלו לרוב אינם אפשריים, הנתונים העשירים הנאספים בסביבות למידה דיגיטליות מודרניות מאפשרים לנו במקרים רבים לזהות, למדוד ולבטל את ההשפעה של סוגים שונים של רעש. במקרה שהוצג - רעש הנוצר מהתנהגויות לא לגיטימיות, הנפוצות מאוד גם בסביבות למידה לא-מקוונות. בעזרת השיטות שפיתחנו לזיהוי והערכת רעש כזה, מפתחי מוקים ב-MIT יכלו לשנות את מודלי ההערכה שלהם כדי למנוע התנהגויות כאלה, והצלחנו לכמת את היעילות של מודלים כאלה ביצירת איזון בין אמינות ההערכה לפדגוגיה ממוקדת-לומד.

בהרצאה זו נציג תוצאות ממחקרים שונים על דרכים לזהות ולמדוד את ההטיה הנגרמת על ידי שימוש בחשבונות כפולים, דרכים למנוע אותה, וכיצד למנף אותה לגילוי מקורות רעש אחרים.

## תוקף בחינת המתא"ם למיון לתארים מתקדמים בפסיכולוגיה

מאיר ברנרון, תמר קנת-כהן, נועה גזית  
מאל"ו

בחינת המתא"ם משמשת כלי מרכזי במיון לתארים מתקדמים בפסיכולוגיה ברוב מגמות הלימוד. מטרתה של הבחינה היא לבדוק את השליטה בתכנים ובמיומנויות שנרכשים במהלך לימודי התואר הראשון בפסיכולוגיה באופן מהימן וסטנדרטי. למרות השימוש הנרחב הנעשה בציוני המתא"ם בשני העשורים האחרונים, העדויות לתוקף הניבוי של הבחינה הינן מועטות ונקודתיות. מטרת המחקר הנוכחי היא לבחון את תוקף המתא"ם, ובפרט את תוקף הניבוי. המחקר מבוסס על תלמידי תואר שני בפסיכולוגיה משישה מוסדות לימוד. נמצא כי ציוני המתא"ם מתואמים חיובית עם ציוני התזה, ממוצע ציוני הקורסים המתודולוגיים וממוצע התואר השני. עוד נמצא כי בחינת המתא"ם תורמת למיון מעבר לממוצע התואר הראשון. כמו כן נמצאו עדויות אודות התוקף המתכנס והמבחין של המתא"ם. בהצגה נדון בהמלצות הנוגעות למבנה הבחינה ובהשלכות הממצאים על טיב המיון לתארים מתקדמים בפסיכולוגיה.

## אתגרים בצמצום הנשירה מההשכלה הגבוהה: בדיקת אפקטיביות של התערבות אינטרנטית

מיכל פיליפס-ברנשטיין, איתמר גתי  
האוניברסיטה העברית בירושלים

כ-30% מהסטודנטים לתואר ראשון לא מסיימים את לימודיהם תוך 6 שנים. ההתמודדות עם צמצום הנשירה היא אחד האתגרים של האוניברסיטאות והמכללות. התמקדנו בהערכת המוכנות הפסיכו-חברתית של סטודנטים שנה א' להשתלבות מוצלחת בלימודים. לשם כך, פותח שאלון "מוכנות לקראת לימודים אקדמיים" (Psychosocial Readiness for College) לאיתור סטודנטים בסיכון לנשירה על בסיס שישה מאפיינים פסיכו-חברתיים (מסוגלות עצמית אקדמית, מחויבות ללימודים, נוחות חברתית, מעורבות בקמפוס, משמעת עצמית וחוסן אישי), שנמצאו תורמים להתמדה בלימודים מעבר ליכולות אקדמיות ומצב סוציו-אקונומי. ה-PRC הוא שאלון אינטרנטי קצר (8-6 דקות) הכולל 33 היגדים (5 היגדים לכל אחד מששת המימדים, פריט חימום ושני פריטי תוקף), שיש לדרג אותם בסולם מ-1 (במידה מועטה ביותר) עד 7 (במידה רבה ביותר). לשאלון גם משוב אישי מילולי וגרפי המבוסס על אלגוריתם לאיתור סטודנטים בסיכון לנשירה. מטרת המחקר הייתה לבחון את התרומה לצמצום הנשירה של עצם המענה על השאלון מצד אחד והאפקט של המשוב האישי, מצד שני.

מחקרים קודמים (N=12,000) מצאו שלשאלון תכונות פסיכומטריות טובות מאד במונחי מהימנות פנימית (Cronbach's α=0.93-0.78), המובחנות של ששת הסולמות נתמכה ב-EFA ו-CFA. נמצאה שקילות מדידה בקרב קבוצות שונות (מכללות-אוניברסיטה, גברים-נשים, מעל/מתחת לגיל 23, גבוה/נמוך, וסטטוס דורי בהשכלה גבוהה). נמצא תוקף מתכנס (עם ה-BIG 5 ושאלון ה-HEO, תוקף מבחין (מול הבגרות והפסיכומטרי) ותוקף ניבוי הן של נשירה והן של ממוצע ציוני שנה א'.

באוקטובר 2019 פנינו לכל הסטודנטים שנה א' באוניברסיטה להשתתף בסקר שמטרתו ללמוד על "השתלבות בלימודים אקדמיים". הסטודנטים חולקו מקרית לשלוש קבוצות. קבוצה 1 קיבלה את שאלון ה-PRC, ומשוב אישי עליו. קבוצה 2 קיבלה את שאלון ה-PRC ואחריו את שאלון ה-HEO - אוריינטציות להשכלה גבוהה, המתייחס למשמעות של הלימודים האקדמיים, וקיבלה משוב על שאלון ה-HEO. קבוצה 3 קיבלה את שאלון ה-HEO ושאלון של אוריינטציות לעבודה, עם משוב על שאלון ה-HEO.

אחוז הנשירה (על-פי דיווח עצמי ובקשת הפסקת הלימודים עד 18 יוני) בקרב הסטודנטים שלא נענו להזמנה להשתתף בסקר (42% מתלמידי שנה א') היה 4.36%, גבוה משמעותית מזה של המשתתפים בסקר 2.42%,  $p < 0.001$ . בנוסף:

- למשוב האישי המידי של שאלון ה-PRC יש תרומה נוספת להפחתת הנשירה מעבר לעצם המענה לשאלון (של כ-29%).
- השאלון + המשוב האישי על ה-PRC מפחיתים את הנשירה לעומת אלה שקיבלו משוב אבל על שאלון אחר (HEO) מ-3.87% ל-1.72%,  $p < 0.01$ .
- ממוצע המוכנות בהתבסס על ממוצע ששת מימדי ה-PRC היה נמוך משמעותית בקרב הנושרים מאשר המתמידים ( $d=0.37$ ).
- הנושרים בסמסטר המקוון דיווחו כבר באוקטובר 2019, על משמעת עצמית נמוכה בהרבה  $d=0.91$ . ממצא זה משקף שסטודנטים עם משמעת עצמית גבוהה יותר התמודדו טוב יותר עם אתגרי הלמידה המקוונת.

התוצאות תומכות באפקטיביות של התערבות ממוקדת באמצעות שאלון המוכנות והמשוב המידע עליו לצמצום הנשירה. חשוב להדגיש שהמשתנים הפסיכו-חברתיים מתייחסים למאפיינים שהם אישיים ולא קבוצתיים איתור סטודנטים בסיכון לנשירה באמצעות משתנים פסיכו-חברתיים מאפשר איתור טרם תחילת הלימודים, והפעלת התערבות מוקדמת בהתאם לצרכיו האישיים של הסטודנט, עוד בטרם נוצרו פערים לימודיים.

## התשואה לפסיכומטרי ולבגרות במונחי שכר עתידי

כרמל אורן  
מאל"ו

דוד מעגן  
למ"ס

יואל רפ  
מאל"ו

מחקר זה מנסה לאמוד את התשואות של מרכיבי מערכת המיון להשכלה הגבוהה בארץ, במונחי שכר שנתי, בגיל 35. במוקד המחקר: הציין הפסיכומטרי, ציוני שלושת תחומיו - כמותי, מילולי ואנגלית, ממוצע הבגרות כולל בונסים למקצועות מוגברים. בנוסף, נבדק אפקט הבחירה במקצועות בגרות מוגברים, וכן האפקטים של מסגרת קבלת תואר ראשון, תחום הלימודים לתואר, וענף תעסוקה כלכלי - על השכר, ועל התשואה של משתני המיון. המתאמים הגבוהים עם השכר הם המשתנים: לימודים בתחומי הייטק, פסיכומטרי כמותי, מגדר, פסיכומטרי כללי, עיסוק בהייטק, מתמטיקה 5 ופיזיקה 5. בתחתית הרשימה, המשתנים המתואמים שלילית עם השכר הם מקצועות הבגרות הלא מוגברים (אנגלית 3 ו-4, מתמטיקה 3). מקצועות בגרות הומניים, חוגי הלימוד חברה ורוח, תואר ממכללה אקדמית לחינוך, ותעסוקה בענף חינוך. כשנבדקת התשואה של משתני המיון תוך פיקוח על משתני רקע קבועים (מגדר, מגזר שפה, השכלת הורים, שנת שכר, וחודשי עבודה), עלייה בסטיית תקן אחת בפסיכומטרי מלווה בעלייה של כ-19% בשכר, עלייה בסטיית תקן אחת בתחום הכמותי (כשבנוסף למשתני הרקע, שני תחומי הפסיכומטרי האחרים מוחזקים קבוע) - מלווה בעלייה של כ-22% בשכר, ואילו עלייה בסטיית תקן בממוצע בגרות-בונסים מלווה בשיפור של כ-16% בשכר. עדיפות הפסיכומטרי (או מרכיביו) על הבגרות מבחינת התשואה, או עוצמת הקשר עם שכר עתידי, עקבית עם ממצאי מחקרי תוקף של מערכת המיון, שבהם ההישג המנובא הוא ציונים באקדמיה, ולפסיכומטרי עדיפות בניבוי. במובן זה ניתן לראות בממצאי המחקר הנוכחי סוג של עדויות נוספות לתוקף מערכת המיון להשכלה הגבוהה, שאולי ניתן לכנותו "תוקף כלכלי".

הקשר בין משתני המיון לשכר מתווך ע"י משתני מסגרת קבלת תואר ראשון וענף תעסוקה. פיקוח עליהם מוריד את אומדני התשואה של מרכיבי המיון, אם כי אינו מאפס אותם, לכ-7% לתחום הכמותי בפסיכומטרי וכ-5% לממוצע בגרות בונסים.

פערי מגדר ומגזר לא עמדו במוקד המחקר. עם זאת מעניין לציין כי בעוד שבנתונים הלא מפקחים, הפערים בשכר בין המינים הם 45% לטובת גברים, ו-36% לטובת דוברי עברית. פערים אלה מצטמצמים משמעותית כאשר מפקחים על משתני סוג המוסד, חוג הלימודים ומגזר התעסוקה, לכ-24% לטובת גברים וכ-14% לטובת דוברי עברית. מאחר שהיקף המשרה אינו מפקח במחקר זה, סביר שהפער בשכר (לשעה) בין המינים קטן אף יותר.

## מדידת הטיות בהישגי לומדים ממדינות מתפתחות בקורסים מקוונים המוניים פתוחים בעזרת תיאוריית התגובה לפריט

סער קרפ גרשון, גיורא אלכסנדרון  
מכון ויצמן למדע

קורסים מקוונים המוניים פתוחים (Massive Open Online Courses–MOOCs), או מוקים (וביחיד: מוק) מאפשרים לכל המעוניין בכך גישה חופשית דרך רשת האינטרנט לתכנים לימודיים ברמה גבוהה. מוקים מהווים זירת פעילות חינוכית חדשה וחדשנית, בעלת היקפי פעילות גדולים וגדלים, ומודלים עסקיים המתאגרים את המבנה המסורתי של ההשכלה הגבוהה. מוסדות השכלה גבוהה, כמו גם מוסדות חינוך וגופי הכשרה אחרים משתמשים בקורסים אלה בדרכים שונות לצורך הקניית ידע, מתן קרדיט אקדמי, ועוד, לעיתים תוך שימוש במודלים היברידיים (blended learning) המשלבים הוראה מסורתית עם רכיבים מקוונים מבוססי-מוק. פעילות הלומדים במוקים מתועדת בצורה רציפה ואוטומטית בקבצי יומן (log files). כמויות הלומדים הגדולות והנתונים המפורטים מייצרים נתוני עתק של למידה, ומזמנים אפשרויות למחקר על התנהגות לומדים בסביבות למידה דיגיטליות.

למרות ההבטחה הגלומה במוקים ל"דמוקרטיזציה של הידע", מחקרים מצביעים על פערים והטיות ביחס להישגיהם של לומדים מקבוצות שונות, ובעיקר, הטיה לרעת לומדים שאינם דוברי אנגלית כשפת אם ודוברים ממדינות מתפתחות. אולם, המחקרים הקיימים סובלים ממספר מגבלות, וביניהן התבססות על כמות נמוכה של לומדים. בנוסף, המחקרים הקיימים התמקדו בתיאור התופעה, אולם לא בחנו שאלה מהותיות – האם ניתן לשפר את סיכויי ההצלחה של לומדים מקבוצות אלו בעזרת עיצוב נכון של הקורסים?

כדי לענות על פערים אלו, ניתחנו מוקים שנוצרו במכון הטכנולוגי של מסצ'וסטס (MIT). מוקים אלו תוכננו ע"פ עקרונות טכנו-פדגוגיים עדכניים המעודדים למידה פעילה בעזרת הערכה מעצבת, וכוללים רכיבי למידה שונים כגון סרטוני וידאו קצרים, טקסטים מעובדים ומונגשים, פורומים, ומגוון אמצעי הערכה ומישוב עצמי הפזורים לאורך הקורס. לרשותנו עמדו 158 הרצות של 120 קורסים שהופצו על פלטפורמת edX בין השנים 2013–2018, עם סך של 2.8 מיליון נרשמים ו-78,562 מסיימים. המשתנים הבלתי תלויים בהם התמקדנו הם שפת אם (אנגלית/לא-אנגלית; להלן דוברים ילידיים/לא-ילידיים), ורמת הפיתוח של המדינה (מפותחת/מתפתחת, לפי מדדים מקובלים). המשתנים התלויים שנמדדו הם אחוז המסיימים – מדד ההצלחה הנפוץ ביותר במוקים, ויכולת הלומדים, שהוערכה תוך שימוש בתיאוריית התגובה לפריט (IRT). למרות יתרונותיה הרבים, ישנם IRT למוקים מייצר אתגרים שונים למשל, התמודדות עם ריבוי נסיונות, פידבק, וציונים חלקיים, ועד כה זכה לתשומת לב מועטה.

ספציפית, מחקר זה עונה על השאלות הבאות: האם יש הבדלים בין אחוזי המסיימים ביחס למשתנה הפיתוח (גבוה/נמוך), בקרב דוברים לא-ילידיים? האם השונות במידת הטייה זו קשורה למאפיינים של הקורס עצמו או מוסברת רק על ידי התחלפות אוכלוסיית הלומדים בקורס? האם הטייה באחוז המסיימים תתבטא גם כהטייה בהערכת היכולת עבור הלומדים המסיימים? בכך המחקר תורם לשדה בשלושה אופנים תומכים ומשלימים: (1) התוצאות אותן קיבלנו מאתגרות מחקרים קודמים על חשיבות השפה במוקים, ומחזקות את חשיבות משתנה הפיתוח להבנת מאפייני הלמידה של לומדים במוק. (2) לראשונה נבדקה השאלה האם הטייה זו קשורה גם למבנה ותוכני הקורס ונמצא כי מעבר להתחלפות אוכלוסיית הלומדים במוק, גם המבנה ותכני המוק משפיעים על הישגי הלומדים. (3) על בסיס מדגם קטן ומאופיין היטב של לומדים במוק הראינו שבניגוד להטייה בסיכויים לסיים את הקורס, בקרב המסיימים אין הטייה ביכולת, כפי שנמדדת על ידי תיאוריית התגובה לפריט.

## ימי קורונה קובעים עובדות בשטח - העברת בחינה חורצת גורל בהשגחה מרחוק: המקרה של בחינת הבקיאות בערבית למועמדים ללמוד במכללות להוראה

חנה נוימן, תהילה סנד, יארא כריים, מאיר ליאני, מיכל שקד, ורוניקה ספיץ'  
המרכז לטכנולוגיה חינוכית (מטח)

אחד התנאים ההכרחיים לקיום בחינות חורצות גורל הוא הבטחת תנאים סטנדרטיים ושווים לכלל הנבחנים ושמירה על טוהר הבחינות. ללא קיום תנאים סטנדרטיים המאפשרים להפחית את טעות המדידה וללא הקפדה על טוהר הבחינות, מהימנות ציוני הבחינה מוטלת בספק. ימי הקורונה זימנו לנו אתגרים חדשים, וביניהם - הצורך לקיים בחינות מרחוק, כל נבחן בביתו הוא. הרצאה זו תעסוק בתיאור מקרה של בחינות הבקיאות בערבית למועמדים ללמוד במכללות להוראה, אשר פותחו במטח, והועברו מרחוק בשל הסגר בספטמבר 2020. נציג בה את הדילמות שהתמודדו איתן ואת הפתרונות שמצאו בתנאים הקיימים. נעלה לדיון כללי את שאלת ההיבחות מרחוק, טכנית ומהותית, ובמיוחד עבור בחינות חורצות גורל. האתגר הראשון בקיום בחינה מרחוק היה בחירת הסביבה הטכנולוגית. אנו בחרנו בסביבת "אופק" מבית מטח, המכונה גם "myTestBox" (MTB), שהיא סביבה שפותחה עבור מבחני כיתה ומעולם לא נוסתה עבור מבחנים קובעי גורלות. הדילמה העיקרית שעלתה בשלב הזה הייתה כיצד להבטיח את טוהר הבחינות, כלומר להבטיח הזדהות תקינה של הנבחנים ולמנוע העתקות באופנים שונים, ובפרט באמצעות גלישה באינטרנט. מכיוון שהבחינה הייתה מקוונת, דילמה נוספת שעלתה הייתה שנדרש קשר אינטרנט רציף לכל משך זמן ההיבחות. בפועל התעוררו בעיות טכניות שנבעו מניתוקים בתקשורת האינטרנט אצל חלק מהנבחנים, בעיות שהקשו על איסוף מלא של הנתונים ועל מתן ציונים. כדי להתמודד את אתגרי טוהר הבחינות ובניסיון לאתר העתקה או רמאות מסוג אחר, בדקנו בנפרד את השאלות הסגורות ואת החיבור.

### 1. בדיקת החיבורים לזיהוי העתקות

מעריכים מנוסים העריכו את החיבורים. בנוסף להדרכה שהם קיבלו, כמו בכל מועד, הם התבקשו לציין אם יש חשד להעתקה, ואם כן, לצרף את מה שלדעתם הוא המקור להעתקה - עמודי פייסבוק ואתרי אינטרנט. כל חיבור נבדק על ידי שני מעריכים. החיבורים החשודים בהעתקה הועברו למעריך בכיר, שאישר את קביעת המעריכים, וכל החיבורים הללו נבדקו אל מול הקישורים. מתוך 67 חיבורים חשודים בהעתקה, שלושה נמצאו תקינים, וכל השאר סווגו כחשודים בהעתקה.

### 2. בדיקת השאלות הסגורות והתפלגות הציונים לזיהוי אמינות הציונים

לצורך בדיקת אמינות הציונים נעשו בדיקות סטטיסטיות שונות. בבדיקת הציונים של השאלות הסגורות נמצא פיזור שנראה סביר, ללא אפקט תקרה, וכן נעשו השוואות שונות מול התוצאות של בחינות במועדים קודמים, שהתקיימו במרכזי בחינה בהשגחה מסורתית. למשל, השוואת התפלגות הציונים ומדדים מרכזיים (ממוצע, סטיית תקן, חציון, רבעונים), בדיקת מהימנות פנימית, ניתוחים ברמת הפריט (קושי ומידת הבחנה), דפוסי תשובות, ועוד.

### מסקנות

ביצוע בחינה חורצת גורל בבית הנבחנים הוא, כאמור, אתגר שרוב הארגונים שמעבירים בחינות מתמודדים איתו בימים אלה. בהתנסות שתיארנו כאן מיוצגים סוגי הכשלים הצפויים בדרך ההיבחות הזאת. עם זאת, מהשוואת המדדים השונים של מהימנות ואמינות נתוני הבחינה בשיטה הזאת, לעומת בחינה במרכזי בחינה עם השגחה מסורתית, ניכר שהושגה רמה סבירה של אמינות ציונים. בהנחה שפתרון מלא להשגחה אמינה על בחינות ביתיות עדיין אינו בנמצא, נציע כמה אפשרויות שיוכלו לשפר עוד את המהימנות של בחינות חורצות גורל בבית הנבחנים



## יישומים של ניתוח שפה טבעית להערכה אוטומטית של שאלות פתוחות במדעים

טניה נזרצקי, מוריה אריאלי, גיורא אלכסנדרון  
מכון ויצמן למדע

כתיבה היא אחד מאמצעי התקשורת העיקריים בקהילה המדעית ומרכיב מרכזי באוריינות מדעית. מחקרים קודמים הראו כי כתיבת הסברים מדעיים משפרת את ביצועי הלמידה של התלמידים ומקדמת את הבניית הידע, אולם מחקרים רבים גם הצביעו על העובדה שמספר לא מבוטל של סטודנטים מתקשים לבנות הסברים נאותים. בהתאם לכך, פיתוח יכולת הלומדים לבנות הסברים מדעיים הינו יעד מרכזי של הוראת המדעים.

הערכה מעצבת המספקת משוב היא גורם מפתח בשיפור הכתיבה המדעית, מאחר שהמשוב הוא האמצעי העיקרי שבאמצעותו התלמידים יכולים להעריך ולשפר את כתיבתם. מטרתו של המשוב המעצב היא להנחות את התלמידים על ידי מתן מידע על הפערים בין ביצועיהם הנוכחיים לביצועים הרצויים. כדי לתמוך בשיפור איכות כתיבתם של התלמידים, על המשוב המעצב להיות ספציפי לכתיבתו של כל אחד מהם ולספק המלצות ישימות לשיפור והסברים מדוע יש צורך בשיפורים אלה. הוראה מעצבת של כתיבה מדעית דורשת זמן רב מהמורים להכנת חומרים, קריאה, עריכה ומתן משוב. המורים, המתמודדים עם כיתות גדולות ועומס משימות, מתקשים למצוא את הזמן הדרוש לתהליך זה, מה שגורם לעיכוב ניכר בתזמון ובאיכות המשוב שהתלמידים מקבלים. התוצאה היא שתהליכי הוראה מעצבת איכותיים הכוללים משוב מעצב ומותאם אישית מתרחשים לעיתים נדירות בכיתות מדעיות.

בשנים האחרונות, במטרה לתת מענה לאתגרים בהערכת כתיבה מדעית, פותחו מערכות הערכה אוטומטיות המבוססות על כלים לעיבוד שפה טבעית (NLP). עם זאת, מערכות אלו מיועדות בעיקר למטרות הערכה מסכמת ולא מעצבת. בנוסף, רוב המחקר בתחום מתמקד בניתוח כתיבה באנגלית. עברית הינה שפה עשירה מורפולוגית, ועיבודה מציב אתגרים ניכרים עבור שיטות NLP סטנדרטיות. עד היום לא פורסמו עבודות העוסקות בהערכה של כתיבה מדעית ושאלות פתוחות בהוראת המדעים בעברית.

התרומה שלנו: במחקר זה אנו מדגימים כיצד מערכת NLP המשתמשת בטכניקות למידה עמוקה יכולה להעריך את איכות ההסברים המדעיים על בסיס אוסף של תשובות בדוקות על ידי מומחים אנושיים. תרומתו של המחקר היא כפולה. ראשית, בניגוד לעבודות קודמות בנושא הערכה אוטומטית שהתמקדו בעיקר בהערכה מסכמת, אנו מתמקדים בהערכה מעצבת. במחנך הבדיקה שהמודלים שלנו לומדים נכללות קטגוריות הלוכדות היבטים חיוניים שתגובות התלמידים צריכות לכלול, שמהן ניתן להפיק משוב מעצב וממוקד. שנית, זהו מחקר חלוץ בנושא של שימוש בניתוח שפה טבעית בעברית להערכה אוטומטית של שאלות פתוחות במדעים.

במחקר אנו מציגים בנייה של מודלים להערכה מעצבת אוטומטית של שאלות פתוחות בביולוגיה, ומודדים את איכות ההערכה שהם מספקים על ידי השוואה להערכה הניתנת על ידי מומחים. המודלים משתמשים בכלי עיבוד שפה טבעית בעברית ובלמידת מכונה כדי 'ללמוד' מתוך אוסף תשובות בדוקות (על ידי מומחה) כיצד להעריך תשובות חדשות. התוצאות מראות שהמודלים האוטומטיים מספקים הערכה אוטומטית הדומה להערכת המומחים עבור תשובות לא מוכרות.

עבודה זו היא הצעד הראשון לקראת בניית כלים להערכה אוטומטית של שאלות פתוחות בהוראת המדעים בעברית, תוך שימוש בשיטות של למידת מכונה ועיבוד שפה טבעית. שיטות אלו יוכלו לסייע למורים בתהליכי הוראה מעצבת של כתיבה מדעית. בנוסף לקידום תחום מחקר זה בעברית, אנו מקווים שמחקר זה יסלול את הדרך לעבודות דומות בשפות עשירות מורפולוגית אחרות כגון ערבית. מאחר ושפות אלו מקבלות תשומת לב מועטה בקהילת המחקר, וישנו מיעוט של כלים התומכים בלומדים בשפות הללו, למחקר כזה ישנה חשיבות גלובלית.

## סימפוזיון

### אוריינטציות של צעירים להשכלה גבוהה

איתמר גתי

האוניברסיטה העברית בירושלים

תואר אקדמי חשוב עבור השתלבות מיטבית בעולם העבודה. לימודים אקדמיים משפיעים על הניעות החברתית, רמות השכר וההזדמנויות החברתיות. הבנת המשמעות של השכלה גבוהה בחייהם של צעירים יכולה לסייע ליועצי קריירה ויועצים אקדמיים ביצירת הלימה בין המוטיבציות של הסטודנטים לבין החוויה האקדמית שלהם. מטרת מחקר זה לפתח ולבחון מודל רב-מימדי וכלי למדידתו – שאלון אוריינטציות כלפי השכלה גבוהה (Higher Education Orientations) המורכב מחמישה מימדים – רכישת מקצוע (חשיבות הכשרה אקדמית להשתלבות במקצוע), לחץ סביבתי (ריצוי אנשים הקרובים לך), רכישת ידע והשכלה (אמצעי להשגת ידע והרחבת אופקים), יוקרה (אמצעי להעלאת הסטטוס החברתי), ומסגרת חברתית (הזדמנות להיכרות עם אנשים חדשים והגדלת המעגל החברתי).

### פיתוח שאלון אוריינטציות להשכלה הגבוהה

תרצה וילנר

האוניברסיטה העברית בירושלים

במחקר 1 בחנו את המבנה של שאלון אוריינטציות כלפי השכלה גבוהה, בשני מדגמים של צעירים. בנתוני המדגם הראשון (N=726) ביצענו ניתוח גורמים מגשש (EFA) על 25 הפריטים של שאלון HEO. נמצא ש-58% מהשונות מוסברת על ידי חמשת המימדים (טווח 9% ועד 14%), וחציון הטעינות של הפריטים על הגורם הייתה 0.74 (טווח 0.36 ועד 0.84). לאחר התאמות בפריטי השאלון הרצנו את השאלון על מדגם שני של צעירים (N=703). ניתוח גורמים מאשש (CFA) תמך במבנה 5 המימדים (CFI = .88, RMSEA = .064, SMSR = .07).

### תוקף מתכנס של שאלון אוריינטציות להשכלה גבוהה

יוליה ליפשיץ-ברזילר

האוניברסיטה העברית בירושלים

במחקר 2 תיקפנו את שאלון HEO באמצעות בחינת הקשר עם משתנים נוספים (N=695). נמצא כי אוריינטציות רכישת מקצוע ורכישת ידע והשכלה מתואמים עם אסטרטגיות התמודדות פרודוקטיביות ( $r = .23$ ,  $r = .22$ , בהתאמה); לעומתם, אוריינטציות לחץ סביבתי, יוקרה ומסגרת חברתית מתואמים עם אסטרטגיות חיפוש עזרה ( $r = .21$ ,  $r = .30$ ,  $r = .22$ , בהתאמה), אסטרטגיות לא פרודוקטיביות ( $r = .48$ ,  $r = .36$ ,  $r = .22$ , בהתאמה) ועם יותר קשיים בבחירת מקצוע ( $r = .41$ ,  $r = .37$ ,  $r = .24$ , בהתאמה). מתוך ממצאים אלו עולים שני נתיבים בתהליך קבלת החלטות קריירה. בנתיב הראשון צעירים אשר מונעים מהרצון לרכוש מקצוע או לרכוש ידע והשכלה, שמופעלים על ידי מוטיבציות פנימיות, משתמשים בדרכי התמודדות פרודוקטיביות, ומתקדמים יותר בתהליך קבלת ההחלטה. בנתיב השני, צעירים מונעים מהרצון ליוקרה, מסגרת חברתית או רציית גורמים חיצוניים ← שמשקפים מוטיבציות חיצוניות ← חווים יותר קשיים ← משתמשים בדרכי התמודדות לא פרודוקטיביות ← מתקדמים פחות בתהליך ההחלטה. לדעתנו יותר נגיש להבנה מבחינת האם זה יהיה סימן ← או + זה תלוי בטיב הקשר בין השלבים.

## אוריינטציות להשכלה גבוהה בקרב סטודנטים ערבים

שאדה קשקוש  
האוניברסיטה העברית בירושלים

במחקר 3 פיתחנו ובדקנו גרסה ערבית של שאלון HEO במדגם של 1055 סטודנטים מהחברה הערבית ובחנו את התוקף הבו-זמני של השאלון. תוצאות שקילות המדידה של שאלון HEO מצביעות על כך שלאחר הורדת פריט אחד לחמשת המימדים יש התאמה מספקת מעבר למגדר ( $CFI = .88$ ,  $RMSEA = .037$ ,  $SRMR = .06$ ) והתאמה טובה מעבר לדת - נוצרים ומוסלמים ( $CFI = .86$ ,  $RMSEA = .033$ ,  $SRMR = .07$ ). המאפיינים הפסיכומטריים של הגרסה הערבית של שאלון HEO מראים שגרסה זו ראויה לשימוש בקרב צעירים ערבים, ומדגישים את חשיבות הרגישות התרבותית בעת מתן ייעוץ אקדמי לסטודנטים.

## שימוש במדידה סובייקטיבית (דרך דיווח עצמי) לאומדן מהיר של יכולות מתמטיות של אנשים בוגרים: סוגיות ואתגרים

עדו גל

החוג לשירותי אנוש, אוניברסיטת חיפה

ההרצאה תעסוק בסוגיות ודילמות הקשורות לצורך לאמוד אוריינות מתמטית של בוגרים, תוך שימוש ב'מדד מהיר' (rapid measure) שהוא קצר, מהיר, פשוט וזול לתיפעול. ככלל, הצורך במדדים מהירים עולה במצבים רבים, כגון בהקשר של מדידת אוריינות בריאות או אוריינות פיננסית בהקשרים קליניים, ו/או במחקרים אקדמיים הקשורים לתהליכי קבלת החלטות ומערכים בין השאר עיבוד מידע כמותי או הסתברותי. החיפוש אחרי מדדים מהירים צף שוב בעשור האחרון בהקשר של הצורך בפיתוח מדדים לבדיקת מצבן של כל מדינות העולם ביחד ל"יעדים להתפתחות בר-קיימא" Sustainable Development Goals של האו"ם לשנים 2015-2030, ובמיוחד ב SDG4 הקשור למדידת רמת אוריינות לשונית ואוריינות מתמטית של אנשים בוגרים בכל מדינה. המחבר עוסק בתחום זה כחלק מעבודה מתמשכת עם אונסק"ו (במיוחד UNESCO Institute for Lifelong Learning) וגופים בינלאומיים אחרים.

מטרת ההרצאה לשתף בהתלבטויות ובאתגרים המקצועיים אתם אנו מתמודדים לאור הצורך לפתח מדדים מהירים לתכנית SDG, עם דגש על תחום האוריינות המתמטית. כרקע, סקר מיומנויות בוגרים (PIAAC) של OECD מספק מערכת מדידה מתקדמת ואמינה לאומדן מיומנויות יסוד של בוגרים ברמה הלאומית, כולל יכולות מתמטיות. (גם ישראל השתתפה בסקר זה ב-2016 ונערכת כעת לסבב שני). אולם, סקר PIAAC דורש רמת תיחום גבוהה ומשאבים רבים, ורק כ-40 מדינות מעורבות בו. לפיכך, מרבית מדינות העולם, אשר אינן מסוגלות או מעוניינות להשתתף בתכנית PIAAC, ניצבות בפני החלטות מורכבות בבואן לחפש מדדים חדשים שהם זולים ופשוטים יותר ומסוגלים לספק, ולו חלקית, את דרישות האינדיקטורים של תכנית SDG, ובמקביל יכולים לסייע למקבלי החלטות בכל מדינה להבין את התפלגות רמות האוריינות באוכלוסייה שלהם.

ההרצאה תסקור מקצת הממצאים הידועים לגבי התוקף והמהימנות של מדדים מהירים לגבי רכיבים של אוריינות מתמטית, תדגים כלים בתחום, ותתמקד בהתלבטויות לגבי אפשרות השימוש במדד סובייקטיבי לאומדן רמת אוריינות מתמטית, כתחליף או כחלק ממדידה קוגניטיבית רגילה של מיומנויות יסוד במדינות שאינן מסוגלות או מעוניינות ליישם גישות מדידה המיועדות להשוואות בינלאומיות, לאור המורכבות והעלות שלהן. הדילמות בתחום זה נוגעות מזווית ייחודית ביחסי הגומלין המורכבים בין תקפות, אמינות, שימושיות, ותרומה של כלי מדידה לקבלת החלטות ברמה הלאומית בעולם.

## TO ERR IS HUMAN - פיתוח ותיקוף ראשוני לשאלון לאומדן מועדות לתאונות (Accident Proneness)

אירן דיאמנט  
האקדמית תל אביב יפו

טעויות אנוש, שגיאות וכשלים מוגדרים כמגיפה השקטה של העידן המודרני. כמעט בכל הסקטורים התעסוקתיים קיימת עדות לפגיעה בגוף ונפש, ברווחיות, בתפוקה עקב מופעים שונים של טעויות אנוש, תאונות וכשלים אנושיים. למרות התקדמות ניכרת מבחינה טכנולוגית, מהווה הגורם האנושי, מאז ומתמיד, 'חוליה חלשה'. הנחת יסוד במושג 'מועדות לתאונות' (Accident Proneness) היא כי אינדיבידואלים מסוימים נוטים למעורבות בתאונות ובכשלים יותר מאשר אחרים. מושג זה, ששורשים מוקדמים לפיתוחו היו בפרסומים סטטיסטיים של גרינווד ווודס (Greenwood & Woods, 1919) לגבי התפלגות לא נורמלית של התרחשות תאונות במגזר התעשייתי באותה תקופה, עבר מסע הפכפך בהגדרתו, בדרכי המדידה והתקפות שלו לאורך השנים. עם זאת, בעשורים האחרונים, התבסס מעמדו של המושג במידה ניכרת, הן מבחינה סטטיסטית והן מבחינה תאורטית: התפלגות תאונות וכשלים באוכלוסייה אינה 'נורמלית'; פרופילים מסוימים של איפיונים אישיותיים-קוגניטיביים קשורים, באופן שיטתי, למעורבות בתאונות. על רקע זה של ביסוס למושג ה'מועדות לתאונות', בולט בהעדרו כיום כלי ייעודי להערכת הפרט.

בהתבסס על המודל של ריזון (Reason, 1990) לסיווג כשלים, המבחין בין מועדות לטעויות (lapses) לשגיאות (errors) ולעבירות (violations), מוצג בהרצאה זו ניסיון ליצירת כלי אומדן ל'מועדות לתאונות', תוך הבנייה מושגית ותיקוף ראשוני. מספר מחקרים בוצעו בהקשר זה: מחקר ראשוני נועד לבדוק את מבנה השאלון ולבדוק איכויות פסיכומטריות בסיסיות שלו באוכלוסייה כללית (773 נבדקים); מחקר המשך נועד לבדוק את השאלון בקרב מועמדים לתעסוקה במסגרת מיון תעסוקתי (187 נבדקים) ולבדוק קשרים עם מדדי אישיות (CPI) ויושרה; מחקר אחרון בדק קשר בין ה-AP להתנהגות בריאות 'בטיחותית' בהקשר למגיפת הקורונה (430 נבדקים).

בהתבסס על מספר שלבי פיתוח ובדיקת איכויות פסיכומטריות ראשוניות של הכלי, נמצא שאלון עם מבנה ארבעה גורמים: טעויות, שגיאות תכנון, עבירות פרוצדורליות ועבירות א-נורמטיביות. שאלון זה תומך בטיפולוגיה הבסיסית של ריזון, אך מוסיף הבחנה בין עבירות מסוג 'פרוצדורלי' לבין עבירות 'א-נורמטיביות', הבחנה שעלתה בספרות לגבי התנהגות תעסוקתית חריגה. השאלון נמצא עם נתוני מהימנות מספקים ותמיכה בהשערות לגבי קשרים עם אפיונים אישיותיים (שליטה עצמית, קונפורמיות) ויושרה וכן עם התנהגות בריאות 'בטיחותית' בקורונה.

בהרצאה זו נדון בהמשגה כללית של מושג המועדות לתאונות ותרומתו האפשרית לתחומים שונים במרחב הפסיכולוגיה התעסוקתית, בהם מיון תעסוקתי, ייעוץ וטיפול תעסוקתי.

## ניתוח עיסוקים של תפקיד הרופא המנתח - הגדרת התכונות והכישורים הנדרשים מרופאים מנתחים לצורך פיתוח מערכת מיון להתמחויות כירורגיות

נועה גזית, רון אלישר, גלעד בן-גל  
האוניברסיטה העברית בירושלים

ההכשרה של מתמחים בתחומים כירורגים ברפואה אורכת זמן רב וכרוכה בהשקעה של משאבים רבים. על אף המשאבים הרבים שמושקעים, חלק מהמתמחים אינם מגיעים לרמה התפקודית הנדרשת בסוף ההתמחות. לכן, תהליך מיון יעיל שיאפשר לתכניות ההתמחות לאתר את המועמדים המתאימים ביותר הוא חשוב ונחוץ. עם זאת, כלי המיון המשמשים כיום למיון להתמחויות הם בעלי תוקף נמוך בניבוי תפקוד בהתמחות.

המחקר שיוצג הוא מחקר מקדים שנערך כחלק ממחקר רחב יותר שמטרתו שיפור תהליך המיון להתמחויות כירורגיות ברפואה. במחקר נערך ניתוח עיסוקים (Job Analysis) שנועד להגדיר את התכונות והכישורים הרצויים של רופאים מנתחים במאה ה-21, ואת התכונות והכישורים הרלוונטיים למיון להתמחויות. הצורך במחקר זה עלה בשל מיעוט הספרות בנושא וכן בשל חוסר העדכניות שלה.

ניתוח העיסוקים התבסס על מספר שיטות. בשלב הראשון, נערכה סקירת ספרות מקיפה וכן בוצעו תצפיות על עבודתם של רופאים בחדר ניתוח. בנוסף, נערכו ראיונות מובנים עם 104 רופאים מנתחים בכירים בישראל (מכל התחומים הכירורגיים). הראיונות התמקדו בתכונות וכישורים שחשובים לדעת המשיבים לרופאים בתחומים כירורגיים ולמיון עבור תחומים אלו. הראיונות הוקלטו ותומללו, ולאחר מכן נותחו באמצעות אנליזת תוכן (Content Analysis).

בשלב השני, נבנה שאלון שהופץ לרופאים מומחים ומתמחים בתחומים כירורגים (1,108 משיבים). בשאלון הוצגה למשיבים רשימה של תכונות וכישורים והם התבקשו לדרג את מידת החשיבות של כל אחד מהם עבור עבודתם של רופאים מנתחים ועבור מיון להתמחויות כירורגיות. רשימת התכונות שנכללה בשאלון נבנתה על בסיס סקירת הספרות, התצפיות והראיונות שנערכו.

ניתוח הנתונים של הראיונות והשאלונים הראה שהרופאים שהשתתפו במחקר הדגישו את החשיבות של תכונות וכישורים משלוש קבוצות: כישורים טכניים, כישורים קוגניטיביים ותכונות אישיות. ממצאים אלו עולים בקנה אחד עם ההגדרות שהוזכרו בסקירת הספרות שערכנו. עם זאת, הגדרת התכונות והכישורים שהתקבלה במחקר הנוכחי כללה גם תכונות וכישורים שלא הוזכרו או שלא קיבלו משקל רב בהגדרות העבר. ככל הנראה שינוי זה נובע מכך שבשני העשורים האחרונים חלו שינויים רבים בתחום הכירורגיה שמשפיעים גם על הכישורים הנדרשים להצלחה בתחום זה.

## אנשי קשר ומציגים בכינוס אפי 2021

דוא"ל	שם		
tzur@nite.org.il	צור קרליץ	יו"ר אפי	יום ד', 28.12.21
ritasever9@gmail.com	ריטה סבר	עדכון מועדת "מיהו פסיכומטריקאי"	
iawadie.rama@education.gov.il	אימאן עואדיה	יו"ר הכינוס	
giligold@mta.ac.il	גיל גולדצויג		
yair_noam@yahoo.com	יו"ר - יאיר נעם	מושב א' מיון מקוון בפקודה: מעבר למיון מרחוק בצה"ל בתקופת הקורונה	
nimrodbensior@idf.il	נמרוד בן-סניור		
orravid@gmail.com	רביד אור להב		
danashenkar@hotmail.com	דנה שושני שנקר		
derek.briggs@colorado.edu	Prof. Derek C. Briggs	מושב ב' The Use of Content-Referencing to Evaluate the Magnitude of Student Growth	
avi@nite.org.il	אבי אללוף	עדכון לימודי התעודה בפסיכומטריקה	יום ה', 28.12.21
meirav.attali@gmail.com	יו"ר - מירב אריאלי אטאלי	מושב א' הפקת מדדים משמעותיים מנתונים רועשים: פוטנציאל ואתגרים בשימוש בנתונים מסביבות למידה אותנטיות למידה והערכה	
bbeigmanklebanov@ets.org	ביאטה ביגמן-קלבנוב		
giora.alexandron@weizmann.ac.il	גיורא אלכסנדרון		
iawadie.rama@education.gov.il	יו"ר - אימאן עואדיה	מושב ב' השכלה גבוהה, נטיות דמוגרפיות וחברתית	
meir@nite.org.il	מאיר ברנרן		
michal.phillips-be@mail.huji.ac.il	מיכל פיליפס-ברנשטיין		
carmel@nite.org.il	כרמל אורן		
tzur@nite.org.il	יו"ר - צור קרליץ	מושב ג' סוגיות בהערכה, העברה מרחוק ושימוש מאסיבי	
havan@cet.ac.il	חווה נוימן		
saar.gershon@weizmann.ac.il	סער קרפ גרשון		
moriah.ariely@weizmann.ac.il	מוריה אריאלי		
itamar.gati@huji.ac.il	יו"ר - איתמר גתי	מושב ד' אוריינטציות של צעירים להשכלה גבוהה	
tirza.willner@mail.huji.ac.il	תרצה וילנר		
yuliya.lipshits@mail.huji.ac.il	יוליה ליפשיץ-ברזילרי		
shada.kashkoush@mail.huji.ac.il	שאדה קשקוש		
giligold@mta.ac.il	יו"ר - גיל גולדצויג	מושב ה' סוגיות בהערכה, העברה מרחוק ושימוש מאסיבי	
iddo@research.haifa.ac.il	עדו גל		
irendi@mta.ac.il	אירן דיאמנט		
noa@nite.org.il	נועה גזית		
ispaisrael@gmail.com	מייל לפניית למועצת אפי	ארגון	
metal@nite.org.il	מיטל אילון		

תגובת והצעת  
בנושא האגודה הישראלית לפסיכומטריקה

נא להפנות אלי:

ד"ר צור קרליץ, יו"ר אפי [ispaisrael@gmail.com](mailto:ispaisrael@gmail.com)



אתר אפי [www.ispa.org.il](http://www.ispa.org.il)



עיצוב ועריכה גרפית | יעל פרץ | מאל"ו