

אפי | جاس | ISPA

אגודה ישראלית לפסיכומטריקה

الجمعية الإسرائيلية للبيكومتريكا

ISRAELI PSYCHOMETRIC ASSOCIATION

מבחנים ■ מדידה ■ הערכה

הכינוס השנתי ה-22 של האגודה הישראלית לפסיכומטריקה

3 בפברואר 2026

בית ציוני
אמריקה,
תל אביב

בתמיכת:

מאל"ו

מרכז ארצי לבחינות ולהערכה (נ"ר)

المركز القطري للاختبارات والتقييم | NATIONAL INSTITUTE FOR TESTING & EVALUATION

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

<p>התכנסות רישום וארוחת בוקר מושב פתיחה (מליאה) דברי פתיחה וברכות – יו"ר אפי: עינת נוטע-קורן יו"ר הכינוס: מירב אריאלי-אטאלי ועינת נוטע-קורן</p> <p>■ הרצאת אורח: לחשוב מהר לחשוב לאט בבינה מלאכותית גל צ'צ'יק, אוניברסיטת בר-אילן, NVIDIA</p>	<p>8:30–9:00 9:00–10:10</p>
--	---------------------------------

<p>הפסקת קפה</p>	<p>10:10–10:30</p>
-------------------------	--------------------

<p>כשהערכה פוגשת בינה מלאכותית (מליאה)</p> <p>■ הערכת חשיבה מסתעפת באמצעות מודלי שפה גדולים: בחינה פסיכומטרית של תוקף ומהימנות במבחן השימושים האלטרנטיביים ערן הדס, ארנון הרשקוביץ</p> <p>■ יישום תיאוריה פסיכומטרית לזיהוי פריטים המבדילים בין בני אדם לצ'אטבוטים ליכל זיינפלד, אלונה סטרוגצקי, זיוה בר-דב, רון בלונדר, שלי רפ, גיורא אלכסנדרון</p> <p>■ מה יידרש מבינה מלאכותית כדי להיות שותפה בבניית טיעון הערכתי? לקחים מפתרון בעיות שיתופי יואב ברגנר</p> <p>■ אותו תוכן, שיפוט שונה: כיצד זהותו של נותן המשוב מעצבת את עמדות הסטודנטים כלפי משוב שנוצר על-ידי בינה מלאכותית לעומת משוב אנושי טניה נרצקי, פאולה מג'יה-דומנסאין, ויניטרה סוואמי, ג'יבריל פרג', טניה קיזר</p>	<p>מושב א'</p> <p>יו"ר: צור קרליץ 10:30–11:50</p>
--	--

<p>ממיומנויות להישגים (מושב מקביל)</p> <p>■ נחישות (Grit) כגורם ממתן בקשר בין יכולות קוגניטיביות והישגים בלימודי STEM בתיכון הילה סגל, מאור כלפון-חכמיגרי</p> <p>■ ידע משנה חקירה – שינויים קוגניטיביים ביכולת שאלת שאלות והשפעתם על הערכה אקדמית תובל רז, יועד קנת</p> <p>■ ציוני מבחן פיזה כמנבאים הישגים לימודיים עתידיים בקרב תלמידים בישראל – מחקר אורך נחום בלס, זמירה מברך, דוד מעגן, יואל רפ</p>	<p>מושב ב'</p> <p>יו"ר: יוסי מחלוף 12:00–13:00</p>
--	---

<p>התפתחות מיומנות קריאה (מושב מקביל)</p> <p>סימפוזיון: ניבוי התפתחות אוריינית בשפות שמיות: מחקר אורך מהגן ועד כיתה ד'</p> <p>■ מנבאים מוקדמים לקריאה בעברית: מודל רב-מרכיבי מהגן עד כיתות א' ו-ד' רותם ינון, דנה טל, שלי שאול, יניב קנט-מימון, תמי קציר</p> <p>■ מסלולי התפתחות של קריאה ואיות בערבית מהגן ועד כיתה ד' אפנאן ח'ורי-מטאניס, אבי קרני</p> <p>■ מודל התפתחותי רב-מרכיבי של הבנת הנקרא בעברית מכיתות א' עד ד' לירון חזן, שחר דותן, רותם ינון, תמי קציר</p>	<p>מושב ב'</p> <p>יו"ר: תמי קציר 12:00–13:00</p>
--	---



<p>ארוחת צהריים</p>	<p>13:00 – 13:45</p>
<p>הענקת פרס אפי והרצאה מוזמנת (מליאה)</p> <ul style="list-style-type: none"> הענקת פרס אפי לעבודת דוקטורט יו"ר ועדת הפרס: הלנה קמרון שיפור תהליך המיון להתמחויות כירורגיות ברפואה: הערכה של כישורים טכניים, יכולות קוגניטיביות ומאפייני אישיות באמצעות כלים מבוססי מציאות מדומה וגיימיפיקציה נועה גזית, כלת פרס אפי הרצאה מוזמנת: שימושי בינה מלאכותית במערכות תומכות למידה ביאטה בייגמן-קלבנוב 	<p>מושב ג'</p> <p>יו"ר: מירב אריאלי-אטאלי 13:45 – 14:45</p>
<p>מדידת תפקודים קוגניטיביים וביצועיים (מושב מקביל)</p> <ul style="list-style-type: none"> מחקר תיקוף חוזר של מת"ל (מערכת תפקודי למידה) לאבחון לקות למידה והפרעת קשב במבוגרים יערה לדור ויצמן, ענבל חת, ענת בן-סימון פחות אבל מספק? השפעת צמצום מספר התחנות במבחן הסמכה מבוסס סימולציה לאחיות דורון שגיא, רן עמרם CogniMentor: הערכת מדדי זיכרון עבודה ויזואלי-מרחבי ובדיקת שינויים לאורך זמן יהודית לירון, דוד צוריאלי, עודד מאירון 	<p>מושב ד'1</p> <p>יו"ר: ענת בן סימון 14:50 – 15:50</p>
<p>סוגיות ואתגרים פסיכומטריים (מושב מקביל)</p> <ul style="list-style-type: none"> מדד חוזק האקלים הארגוני (ICC): כלי משלים לסטיית התקן להבנת לכידות והסכמה תפיסתית ולדירוג יחידות במשטרה אולגה אנטל קטייב מעבר ממבחן בשיטה הקלאסית לתת"פ: העברת תטאות לסולם יע"ל הרב-שנתי שאול שלפק, מיכל באומר, צור קרליץ מדידה אדפטיבית באמצעות שאלונים רב-מימדיים – בחינה אמפירית ראשונית בשני שאלונים איתמר גתי, אלעד פאוסט 	<p>מושב ד'2</p> <p>יו"ר: אבי אללוף 14:50 – 15:50</p>
<p>אוריינות מידע בעידן ה-AI (מליאה)</p> <ul style="list-style-type: none"> מבחן "תנופה למחר": הערכת מיומנויות גנריות בסביבה פתוחה – אתגרים וממצאים ראשוניים תמי סבג שושן, אימאן עואדיה, לימור בורוט, יואב ברגנר, מירב אריאלי-אטאלי, נדב כספי, לירון גביסון שאלון ה-AIR: פיתוח ותיקוף של שאלון למדידת מוטיבציות לקריאה בסיועה של בינה מלאכותית יוצרת עדי ברן, שיר אתגר, יעל סידי 	<p>מושב נעילה</p> <p>יו"ר: גילי גולדצויג 16:00 – 16:40</p>
<p>אסיפת חברי אפי</p>	<p>16:40 – 17:00</p>



לחשוב מהר, לחשוב לאט בבינה מלאכותית

גל צ'צ'יק

אוניברסיטת בר-אילן, חברת NVIDIA

בדומה למערכות חשיבה ביולוגיות, מערכות AI, נדרשות לבצע הסקה בכמה מתווים שונים מאוד זה מזה: זיהוי דוגמאות מוכרות, הכללה לדוגמאות דומות, ופתרון בעיות חדשות לחלוטין. מערכות AI בנויות על גבי שני סוגי רכיבים: אלגוריתמים שתוכננו בידי אדם ואלגוריתמים הלומדים מתוך נתונים. הראשונים פותרים בעיות ביעילות אך ללא הסתמכות על ניסיון קודם, והאחרונים מתבססים על התפלגות נתונים אך מתקשים לנצל ביעילות תכונות של בעיה חדשה. בהרצאה אתאר כיצד ניתן לשלב את שתי הגישות למערכות שמבצעות חישוב יחודי עבור קלט נתון, על גבי מודל שאומן מראש. בדומה לשילוב "System 1" עם "System 2" במערכת הקוגניטיבית האנושית, מדובר בתהליך הסקה שיכול להיות איטרטיבי ואיטי, והוא כולל אופטימיזציה לא טריוויאלית מעבר להסקה פשוטה שמודל מאומן מבצע. ההרצאה תסקור את הגישות העדכניות בבניית מודלי AI לשילוב המערכות, ותתאר ישומים של גישה זאת בעיקר במודלים גנרטיביים מטקסט לתמונה, שמאפשרים עריכת תמונות, התאמה לבקשות המשתמש ויצירת קטגוריות נדירות.



הערכת חשיבה מסתעפת באמצעות מודלי שפה גדולים: בחינה פסיכומטרית של תוקף ומהימנות במבחן השימושים האלטרנטיביים

ערן הדס, ארנון הרשקוביץ
אוניברסיטת תל-אביב

הערכת יצירתיות, ובפרט חשיבה מסתעפת, מהווה אתגר פסיכומטרי מתמשך בשל אופייה הרב-ממדי, ההסתמכות המסורתית על שיפוט אנושי, והקושי בהבטחת עקביות ואובייקטיביות בניקוד. מבחן השימושים האלטרנטיביים (Alternative Uses Test; AUT) משמש כלי מרכזי למדידת חשיבה מסתעפת, וכולל ארבעה ממדים: שטף (Fluency), גמישות (Flexibility), מקוריות (Originality) ואלאבורציה (Elaboration). בעוד שמדדי שטף ואלאבורציה ניתנים לחישוב אוטומטי פשוט, הערכת גמישות ומקוריות נשענה עד כה לרוב על קידוד ידני עתיר משאבים. בהרצאה זו יוצגו ממצאים משני מחקרים אמפיריים הבוחנים את התאמתם של מודלי שפה גדולים (Large Language Models; LLMs) ככלי מדידה פסיכומטריים להערכת ממדים אלו.

המחקר הראשון בחן תוקף קריטריון, תוקף תוכן ומהימנות של ניקוד אוטומטי מבוסס LLM בהשוואה לניקוד אנושי. המדגם כלל שלוש אוכלוסיות גיל: ילדים בכיתות ג'-ד' (n=106), מתבגרים בכיתה ט' (n=119), ומבוגרים (n=618). הנתונים נאספו בעברית ובאנגלית, כאשר התגובות בעברית תורגמו לאנגלית באמצעות תרגום אוטומטי שעבר אימות סמנטי והראה מתאמים גבוהים לתרגום ידני. במסגרת מחקר זה, מדד הגמישות נבנה בתהליך דו-שלבי של הפקת קטגוריות סמנטיות וסיווג תגובות, בעוד שמדד המקוריות נוקד בסולם בן חמישה ציונים באמצעות מודלים כלליים (GPT-4o ו-GPT-4.1) ומודל מכוון ייעודי. הממצאים העלו מתאמים גבוהים בין הניקוד האוטומטי לניקוד האנושי עבור הגמישות ($r=0.87$) ועבור המקוריות ($r=0.73-0.80$), לצד חפיפה גבוהה בין הקטגוריות שנוצרו. מהימנות בין הרצות של אותו מודל נמצאה גבוהה מאוד ($ICC > 0.90$), ממצא המעיד על יציבות פנימית גבוהה. ניתוח השוואתי הראה כי בעוד שמודלים כלליים שמרו על רמת תוקף עקבית בין קבוצות גיל, מודלים מכוונים הפגינו ירידה ביכולת הכללה מחוץ לאוכלוסיית האימון.

המחקר השני הרחיב את הבחינה להיבטים של תוקף ומהימנות לאורך זמן. במסגרת מחקר אורכי שנמשך עשרה שבועות, השתתפו 157 תלמידי כיתה ט' שביצעו סדרת מבחני AUT חוזרים. הניקוד האוטומטי שמר על מתאם גבוה עם הניקוד האנושי לאורך כל תקופת המדידה ואפשר ניתוח מגמות שינוי. תוצאות המחקר הראו כי תרגול חוזר תרם לעלייה בשטף ובגמישות, אך לא לשיפור מובהק במקוריות. זהו ממצא התואם את ההבחנה הפסיכומטרית המקובלת בין מדדי כמות למדדי איכות.

הממצאים מצביעים על כך שניקוד אוטומטי מבוסס LLM עשוי לשמש כלי מדידה תקף ומהימן להערכת חשיבה מסתעפת בקנה מידה רחב, תוך שמירה על עקביות פנימית ויכולת הכללה בין אוכלוסיות. השימוש במודלים אלו מאפשר צמצום משמעותי של זמן ועלויות ניקוד, ומספק תשתית למדידה חוזרת ורגישה לשינוי לאורך זמן. מעבר להיבטים המדידתיים, ממצאי המחקר מצביעים על פוטנציאל לשילוב הערכה אוטומטית במערכות למידה אדפטיביות, שבהן ניטור מתמשך של מדדי חשיבה מסתעפת עשוי לאפשר הצעת רמזים סמנטיים מתונים כתמיכה בתהליך הלמידה. בהקשר זה, מודלי שפה גדולים עשויים לשמש גשר מתודולוגי בין מחקר קוגניטיבי למדידה חינוכית מבוססת נתונים, ולאפשר בחינה שיטתית של תהליכים מנטליים מורכבים תוך שמירה על עקרונות פסיכומטריים של תוקף ומהימנות.



יישום תיאוריה פסיכומטרית לזיהוי פריטים המבדילים בין בני אדם לצ'אטבוטים

ליכל זיינפלד, אלונה סטרוגצקי, זיוה בר-דב, רון בלונדר, שלי רפ, גיורא אלכסנדרון
מכון ויצמן למדע

האימוץ המואץ של כלי בינה מלאכותית יוצרת (GenAI) בחינוך מביא הזדמנויות וסיכונים: כלים אלו ובפרט מודלי שפה גדולים (LLMs), יכולים לספק לומדים הסברים, משוב ותמיכה מותאמים אישית, אך השימוש הגובר בהם עשוי לערער את התוקף של כלי הערכה. מכאן, עולה התובנה שכדי לייצר כלי הערכה תקפים ומהימנים בעידן הבינה המלאכותית היוצרת נדרש להבין את היכולות והמגבלות של LLMs ביחס למגוון של משימות הערכה, ובפרט, היכן אילו שונים מלומדים אנושיים.

בפסיכומטריקה, תפקוד דיפרנציאלי של פריט (DIF) מתרחש כאשר פריט הערכה מתפקד באופן שונה עבור תתי-קבוצות של לומדים המובחנות לפי מאפיין שאינו קשור ליכולת הנמדדת. במונחי DIF, הקבוצה שעליה תפקוד הפריטים ידוע מכונה קבוצת הייחוס, וקבוצת המיקוד היא הקבוצה שעליה אנו מעוניינים לבחון האם הפריטים מתפקדים באופן שונה. במחקר זה, שאלנו האם שיטות DIF אפקטיביות לזיהוי פריטים המתפקדים באופן שונה עבור לומדים אנושיים (קבוצת הייחוס) ו-LLMs (קבוצת המיקוד).

השאלה הראשונה בה התמקדנו היא האם טכניקות DIF יכולות לזהות פריטים המתפקדים באופן שונה עבור לומדים אנושיים ועבור צ'אטבוטים? לצורך כך פיתחנו שיטה המשלבת שיטות DIF סטנדרטיות כמו LR-MH ו-LR, בשילוב טכניקות ניתוח פריטים (item-total correlation) וטכניקות סטטיסטיות לבחינת יציבות הממצאים. המחקר התבצע על שני מערכי נתונים שכללו מעני לומדים למבחן הכנה לבגרות בכימיה, ולפרק הכמותי מהמבחן הפסיכומטרי. זאת בשילוב של מענה של גרסאות פרימיום של מודלי שפה שונים למבחנים אלו, ביחס של 1:10 בין צ'אטבוטים ללומדים אנושיים.

בבחינת יציבות הממצאים בכל אחת מהשיטות (LR-MH), נמצא כי מה-HM התקבלו תוצאות רועשות כאשר פריטים רבים סומנו כ-DIF גם כאשר מעני ה-LLMs הושמטו לחלוטין (Negative Control Analysis). שיטה ששילבה LR-DIF עם ניתוח item-total correlation נמצאה כאפקטיבית בזיהוי פריטים בעלי תפקוד שונה בין לומדים אנושיים לבוטים והיא השיטה שנבחרה.

בשלב השני של המחקר יישמנו את השיטה לזיהוי פריטי DIF לשאלון בכימיה. פריטים אלו הועברו למומחי תוכן אשר זיהו את הגורמים המרכזיים שמאפיינים פריטים שבהם צ'אטבוטים מפגינים ביצועים דיפרנציאליים ביחס ללומדים אנושיים. הפריטים שנמצאו כבעלי DIF חיובי שיקפו חוזקות של הצ'אט במשימות המאופיינות כמכילות כללים ברורים ובהירות מבחינה מושגית, בעוד שהפריטים בעלי ה-DIF השלילי שיקפו קושי של הצ'אטבוטים בפריטים שכללו התייחסות למוסכמות מרומזות, מידע חזותי, ופרוצדורות רב-שלביות.



מה יידרש מבינה מלאכותית כדי להיות שותפה בבניית טיעון הערכתי? לקחים מפתרון בעיות שיתופי

יואב ברגנר

הרשות הארצית למדידה והערכה (ראמ"ה), NYU

אחת מתרומותיו המכוננות של מיסלבי (Mislevy et al, 2003) למדידה חינוכית הייתה פיתוח המטאפורה "הערכה כטיעון ראייתי". בעבודות מאוחרות (Mislevy 2013; 2018) יותר הגדיר זאת כ"טיעון עיצוב/פרשנות", ובכך הרחיב את השיח אל עבר המחוזות הסוציולוגיים. ההחלטה מה נחשב לראיה מערבת מגוון סגולות אנושיות קולקטיביות. לנוכח פיתוחם של מודלי בינה מלאכותית (AI) בעלי יכולות משופרות, הרצאה זו שואלת: מה נדרש כדי שבינה מלאכותית תהפוך לשותפה אפקטיבית בתהליך ה"הערכה כטיעון"?

תפקיד ה-AI במדידה חינוכית אינו חדש, אך הוא עודו מוגבל. כיום, למידת מכונה משמשת באופן שגרתי למתן ציונים אוטומטי ויצירת פריטים מבוססת תבניות. עם זאת, בעוד ששימויות אלו נסמכות על זיהוי דפוסים (pattern-matching). "הערכה כטיעון" מחייבת יכולת שונה מהותית: יכולת הסקה (Reasoning). מוגבלות ה-AI בהסקה ממשית נחשפת מול משימות לא שגרתיות החורגות מנתוני האימון. הרצאה זו תציג סדרת מחקרים העוסקים בפתרון בעיות שיתופי (CPS) במתמטיקה, החושפים את מגבלות ההסקה של מודלי ה-AI המתקדמים ביותר. במחקרים אלו, CPS הוגדר באמצעות תבניות פריטים המבוססות על תלות הדדית (interdependence).

המחקר הראשון בחן את יכולתם של מודלי שפה גדולים (LLMs) לסווג פריטים לפי קריטריונים של תלות הדדית. הממצאים העלו כי רמת ההסכמה מול שיפוט אנושי הייתה בינונית לכל היותר, וכי אסטרטגיות הנחיה (prompting) מורכבות הניבו שיפור זניח. הדבר מצביע על כך שהסקה לגבי תלות הדדית היא אתגר מבני עמוק עבור המודלים. בהמשך, פותח מערך מרובה-סוכנים ליצירת מחוונים (rubrics) ויישומם בהערכה. אף שהשקיפות השתפרה, לא נרשמו רווחים עקביים בדיוק הסיווג. המודלים נטו להדגיש תקשורת כללית תוך החמצת אילוצים קריטיים לפתרון הבעיה.

ניסיון נוסף כלל סימולציה של צמדי סטודנטים וירטואליים הפותרים משימות CPS. הסימולציה הציפה כשלי תכנון שהוחמצו בהערכה סטטית, אך המהימנות הכללית נותרה נמוכה. מחקר מסכם סינתז את התובנות לתהליך איטרטיבי של הערכה ותיקון. בעוד שהתהליך שיפר פריטים פגומים בסבב הראשון, סבבים נוספים הובילו לעלייה בשיעור החיוביים השגויים (false positives) - עדות לסיכון של "סחף דורי" שבו תיקון עצמי חוזר פוגע באיכות המוצר.

לסיכום, סדרת מחקרים זו מדגימה כי מודלי שפה יכולים להוות כלי עזר חלקי בבקרת איכות, אך אינם בשלים לשמש כמעריכים עצמאיים. לאור זאת, אדון ביכולות ההסקה המינימליות הנדרשות מבינה מלאכותית כדי להפוך לשותפה מועילה בבסוס טיעוני עיצוב ופרשנות - לאו דווקא בינה מלאכותית כללית (AGI), אלא יכולת הסקה מבנית מספקת שתאפשר לה לקחת חלק משמעותי בדיאלוג המקצועי.



אותו תוכן, שיפוט שונה: כיצד זהותו של נותן המשוב מעצבת את עמדות הסטודנטים כלפי משוב שנוצר על-ידי בינה מלאכותית לעומת משוב אנושי

טניה נזרצקי, פאולה מג'יה-דומנסאין, ויניטרה סוואמי, ג'יבריל פרג', טניה קיזר
EPFL, שוויץ

למשוב מעצב תפקיד מכריע בלמידה, שכן הוא מסייע ללומדים להבין את ביצועיהם ולשפרם. עם זאת, מתן משוב מותאם אישית ובזמן בהשכלה הגבוהה מציב אתגר, בשל אוכלוסיית הסטודנטים הגדולה והמגוונת, דבר שלעיתים קרובות מוביל למשוב מאוחר ולא מותאם אישית. התפתחויות אחרונות בתחום הבינה המלאכותית הגנרטיבית מציעות פתרון פוטנציאלי למתן משוב מעצב מידי ובקנה מידה רחב. עם זאת, מעט ידוע על תפיסות הסטודנטים כלפי משוב הנכתב על ידי בינה מלאכותית.

במאמר זה אנו בוחנים כיצד זהותו של נותן המשוב משפיעה על תפיסות הסטודנטים, תוך התמקדות בהשוואה בין משוב שנוצר על-ידי בינה מלאכותית לבין משוב שנכתב על-ידי בני אדם. הגישה שלנו מבוססת על הערכת משוב על-ידי סטודנטים בסביבות למידה אותנטיות, לפני ואחרי חשיפת זהותו של נותן המשוב, במטרה לבחון את השפעת מידע זה על תפיסתם. המחקר, שנערך בקרב 457 סטודנטים מתחומי לימוד ומרמות אקדמיות שונות, מעלה כי יכולתם של הסטודנטים להבחין בין משוב שנוצר על-ידי בינה מלאכותית למשוב אנושי תלויה במשימה הנבחנת. כאשר זהותו של נותן המשוב לא נחשפה לסטודנטים הם נטו להעדיף את המשוב שלדעתם נכתב על-ידי בני אדם. חשיפת זהותו של נותן המשוב משפיעה על העדפות הסטודנטים, ומובילה להעדפה מוגברת למשוב אנושי ולהפחתה בהערכה של משוב שנוצר על-ידי בינה מלאכותית. בנוסף, סטודנטים העריכו את בני האדם כאמינים יותר באופן מובהק בתור נותני משוב מן הבינה המלאכותית. נטיות אלו היו דומות בין רמות אקדמיות שונות, מגדרים ותחומי לימוד שונים.

ממצאים אלו מדגישים את המורכבות הכרוכה בשילוב בינה מלאכותית במערכות משוב חינוכיות, ואת החשיבות שבהתחשבות בתפיסות הסטודנטים בעת אימוץ משוב שנוצר על-ידי בינה מלאכותית בהשכלה הגבוהה.



נחישות (Grit) כגורם ממתן בקשר בין יכולות קוגניטיביות והישגים בלימודי STEM בתיכון

הילה סגל, מאור כלפון-חכמיגרי המכללה האקדמית תל אביב-יפו

לימודי STEM (מדעים, טכנולוגיה, הנדסה ומתמטיקה) מהווים תשתית מרכזית להכשרת כוח אדם מקצועי בעידן של חדשנות טכנולוגית וצמיחה כלכלית. מחקרים רבים מצביעים על תרומתן המשמעותית של יכולות קוגניטיביות לניבוי הצלחה אקדמית בתחומים אלו, אולם ממצאים מצטברים מעידים כי יכולות אלו אינן מסבירות לבדן את מלוא השונות בהישגים, במיוחד במסלולים תובעניים הדורשים התמדה לאורך זמן. לפיכך, גוברת ההתעניינות בתרומתם של גורמים לא קוגניטיביים לניבוי הצלחה בלימודי STEM.

המחקר הנוכחי מתמקד במושג הנחישות (Grit, Duckworth et al., 2007) המוגדר כתשוקה והתמדה למען מטרות ארוכות טווח, וכולל שני מרכיבים מרכזיים: התמדה במאמץ ועניין עקבי. הסולם של "התמדה במאמץ" מתייחס ליכולת להמשיך ולהשקיע מאמצים למרות כישלונות והתקדמות איטית, בעוד שהסולם של "עניין עקבי" מתאר שמירה על עניין יציב במטרות ארוכות טווח ללא שינוי תדיר בכיווני העניין. בעוד שחלק מהמחקרים מצאו כי נחישות תורמת לניבוי הישגים אקדמיים ואף עשויה לפצות על פערים קוגניטיביים, מטא-אנליזות עדכניות מצביעות על תרומה מתונה בלבד של נחישות לציונים, ועל שונות בממצאים לגבי תפקידם היחסי של שני מרכיביה (Credé et al., 2017; Lam & Zhou, 2022).

מטרת המחקר הייתה לבחון האם נחישות תורמת לניבוי הישגים בלימודי STEM בתיכון מעבר ליכולות קוגניטיביות, והאם היא משמשת גורם ממתן בקשר בין פוטנציאל קוגניטיבי להישגים. בנוסף, נבחנה לראשונה התאמתו הפסיכומטרית של שאלון הנחישות בעברית כחלק מתהליך מיון, בדגש על מהימנות והיבטים של רצייה חברתית.

המדגם כלל 814 תלמידי תיכון מצטיינים בלימודי STEM, שהשתתפו בתהליך מיון לתוכנית עתודה יוקרתית בתחומי ההנדסה והמדעים המדויקים בשלוש אוניברסיטאות בישראל. במסגרת התהליך נאספו ציוני התיכון במתמטיקה, פיזיקה ומדעי המחשב, ציוני פסיכומטרי, ציוני מבחני יכולת קוגניטיביים ייעודיים, וציוני שאלון נחישות (Grit-S). בוצעו מודלי רגרסיה עם מיתון, שבחנו את תרומת הפוטנציאל הקוגניטיבי, הנחישות והאינטראקציה ביניהם לניבוי הישגים בלימודי STEM.

הממצאים הצביעו על מהימנות נאותה של שאלון הנחישות בעברית. כצפוי, יכולות קוגניטיביות ניבאו באופן מובהק הישגים בלימודי STEM, ונחישות תרמה נוספת אך מתונה לניבוי. הממצא המרכזי היה אפקט מיתון מובהק: ככל שרמת הנחישות הייתה גבוהה יותר, כך נחלש הקשר בין פוטנציאל קוגניטיבי להישגים. תלמידים בעלי נחישות גבוהה הצליחו לצמצם פערים בהישגיהם במקצועות ה-STEM בתיכון, גם כאשר יכולותיהם הקוגניטיביות היו מתונות יחסית.

מסקנות המחקר מדגישות את חשיבות השילוב בין מדדים קוגניטיביים ולא-קוגניטיביים בהערכה ובמיון. הממצאים תומכים בשילוב מדידת נחישות בתהליכי מיון ל-STEM, ומציעים תרומה יישומית ופסיכומטרית להבנת מנגנוני ניבוי הצלחה במסלולים אקדמיים תובעניים.



ידע משנה חקירה – שינויים קוגניטיביים ביכולת שאילת והשפעתם על הערכה אקדמית

תובל רז, יועד קנת
הטכניון

שאילת שאלות היא מנגנון קוגניטיבי מרכזי המניע למידה, סקרנות ופיתוח של ידע חדש. תהליך זה, אשר כולל איתור של פערים בידע, גיבוש שאלה מתאימה ובחירת כיווני חקירה, מהווה יסוד מרכזי בהתפתחות קוגניטיבית ובחינוך, ונחקר רבות במסגרת ניסויי מעבדה קצרי טווח. עם זאת, אנו יודעים מעט מאוד על האופן שבו איכות השאלות, אשר מבוססת על ידי מקוריותן ומורכבותן, משתנה לאורך זמן, ובמיוחד בהקשר של ההשפעה של רכישת ידע רלוונטי לתחום השאלה. המחקר הנוכחי ביקש לענות על פער זה באמצעות בחינה שיטתית של האופן שבו ידע שנרכש במהלך סמסטר אקדמי משפיע על יכולת שאילת שאלות, וכן כיצד מדדים אלה מנבאים הישגים אקדמיים שונים.

במחקר השתתפו 68 סטודנטים בקורס "מבוא לפסיכולוגיה", אשר הוערכו בתחילת ובסוף הסמסטר בשתי מטלות שאילת שאלות: מטלה כללית של שאילת שאלות – מטלת השאלות האלטרנטיביות (AQT) וגרסה ייעודית לתחום הקורס (AQT – Domain Specific (DS)). בכל אחת מן המטלות נמדדו שלושה מדדים מרכזיים: שטף (fluency) מקוריות (originality) ומורכבות (complexity) השאלה, אשר מבוססת על טקסונומיית בלום. מדדים אלו הוכחו בעבר בתור מנבאים חיוביים של יצירתיות, פתרון בעיות פתוחות וסקרנות. במקביל, נאספו נתוני ביצוע והישגים אקדמיים: הצלחה בפרויקט קבוצתי פתוח (ניסוח של השערת מחקר והרצת ניסוי) וציון בבחינה אמריקאית סגורה. מבנה המחקר אפשר בחינה אקולוגית ורב-ממדית של הקשר בין שאילת שאלות, ידע רלוונטי שנרכש לאורך זמן והערכה אקדמית פתוחה וסגורה.

ממצאי המחקר הראו כי בעוד שבמטלת ה-AQT-DS הספציפית חל שיפור מובהק בכל המדדים – שטף, מקוריות ומורכבות שאלה, הרי שבמטלת ה-AQT הכללית נצפתה יציבות ואף ירידה במקוריות לאורך זמן. דפוס זה מצביע על כך שרכישת ידע תחומי מאפשרת שאילת שאלות מורכבות ובעלות יצירתיות גבוהה יותר בתוך תחום הידע הרלוונטי, אך אינה מרחיבה בהכרח יכולות כלליות של שאילת שאלות.

ניתוחי המשך חשפו כי מדדי המורכבות והמקוריות, הן בתחילת הסמסטר והן בסופו, היו קשורים באופן שיטתי לביצועים בסוף הקורס, אך בצורה דיפרנציאלית. בעוד שמדדים אלו ניבאו באופן שלילי את ציוני הבחינה הסגורה, הם ניבאו באופן חיובי את הישגי הפרויקט הפתוח בקבוצות. ממצא זה תואם תיאוריות מוכרות בתחום היצירתיות וחשיבה המבחינות בין חשיבה מתבזרת (divergent) המתאימה למשימות פתוחות וקשורה ביצירתיות, לבין חשיבה מתכנסת (convergent) הנדרשת לבחינות סגורות וקשורה לאינטליגנציה. בנוסף, ניתוח רגרסיה היררכית הראה כי מדדי שאילת שאלות מסוף הסמסטר תרמו תרומה ייחודית לניבוי הישגים, מה שמצביע על כך שהצטברות הידע ולא רק יכולות ראשוניות היא המעצבת את איכות השאלות ואת תרומתן להצלחה.

מבחינה תאורטית, ממצאים אלה מדגישים כי שאילת שאלות היא יכולת דינמית הנוצרת מתוך אינטראקציה בין ידע, סקרנות ומבני חשיבה יצירתיים. מורכבות השאלות משקפת את האופן שבו ידע חדש מוטמע במערכת הייצוגית של הלומד, ומספקת עדות לכך שתהליכי יצירתיות ושאילת שאלות חולקים מנגנונים קוגניטיביים משותפים.

מבחינה חינוכית, המחקר מצביע על חשיבות טיפוח אוריינות שאילת שאלות כמיומנות העשויה לקדם הצלחה במטלות פתוחות ויצירתיות, לצד הצורך בהתאמת כלי הערכה שונים כדי ללכוד מגוון יכולות קוגניטיביות. בכך, המחקר מסמן כיוונים עתידיים להתערבויות פדגוגיות ולקידום למידה המבוססת על חקר, הבנה עמוקה ופיתוח חשיבה גבוהה.



ציוני מבחן פיזה כמנבאים הישגים לימודיים עתידיים בקרב תלמידים בישראל – מחקר אורך

נחום בלס	זמירה מברך
מרכז טאוב לחקר המדיניות החברתית בישראל	אוניברסיטת בר-אילן
DOD מעגן	יואל רפ
למ"ס	מרכז ארצי לבחינות ולהערכה (מאל"ו)

מחקר פיזה הנערך על ידי ארגון ה-OECD מאז שנת 2000, נחשב למחקר החינוך החשוב והמקיף ביותר בעולם. המחקר נועד להעריך את רמת האוריינות בקריאה, במתמטיקה ובמדעים של תלמידים בני 15 במדינות משתתפות (כיום כ-80 מדינות שכמחציתן מדינות ה-OECD) כדי לבחון באיזו מידה תלמידים מוכנים להשתלב כבוגרים בחברה ובכלכלה של מדינות מפותחות. הוא נערך אחת לשלוש שנים, כאשר בכל מחזור מחקר ניתן דגש לאחת האורייניות. ב-2009 עמדה במרכז המחקר אוריינות קריאה. בכל מדינה נבחר מדגם מייצג של אוכלוסיית הלומדים באותה מדינה. השילוב בין מדגמים גדולים ומייצגים ובין תכנון מוקפד של כלי ההערכה יוצר תשתית איתנה לדיווח על תוצרי חינוך של המדינות המשתתפות ומסגרת להשוואה בין-לאומית. כך, ניתנים כלים למקבלי החלטות לעיצוב מדיניות חינוכית בכל מדינה ואינדיקציה לכלכלנים לגבי איכות ההון האנושי העתידי ולפוטנציאל הצמיחה הכלכלית. עורכי המחקר מחלקים את סולם הציונים באוריינות קריאה לשבע רמות בקיאות, וטוענים כי תלמידים בעלי רמת בקיאות נמוכה מ-2 נמצאים ברמת סיכון גבוהה וספק לגבי יכולתם להשתלב בבגרותם בחברה ובכלכלה.

המחקר הנוכחי הוא מחקר אורך, העוקב אחר נבחני פיזה 2009 בישראל במשך קרוב ל-15 שנים ובודק באיזו מידה הם הצליחו בשלבי חינוך עתידיים. האם ניגשו לבחינות הבגרות ולבחינה הפסיכומטרית, ועד כמה הצליחו בהן? האם נכנסו להשכלה גבוהה והאם השלימו תואר ראשון? האם, כפי שנטען, למבחני פיזה יכולת חזוי טובה של הישגים חינוכיים ורכישת השכלה, מדדים שעשויים להצביע בעקיפין על נתוני הצמיחה הכלכלית של כל מדינה משתתפת?

בדיקות הצלחתם של תלמידים על פי פיזה נעשיתה במחקר זה באמצעות ניתוח נתונים תיאורי בכל שלב חינוך (בגרות, פסיכומטרי, השכלה גבוהה) על פי רמות אוריינות הקריאה ומודלים רב-משתניים בהם נבדקה השפעת הציונים על מדדי החינוך העתידיים תוך פיקוח על הרקע החברתי-כלכלי, המגדר והמגזר השפתי (עברית/ערבית) של התלמידים.

הממצאים מצביעים על קשר חזק בין רמת אוריינות הקריאה בפיזה להישגים לימודיים עתידיים ומספקים מספר נתונים מפתיעים בנוגע להישגיהם של תלמידים בעלי רמת קריאה נמוכה במבחן הפיזה ולנבחנים דוברי ערבית. בכנס נציג את הממצאים, נדון בהשתמעותם, במגבלות המחקר ונציע כיוונים למחקרים עתידיים.



סימפוזיון

ניבוי התפתחות אוריינית בשפות שמיות: מחקר אורך מהגן ועד כיתה ד'

תמי קציר

מרכז אדמונד י. ספרא לחקר המוח ולקויות למידה, אוניברסיטת חיפה

שאלה מרכזית במחקר האוריינות עוסקת במידה שבה יכולות קוגניטיביות-לשוניות מוקדמות מנבאות הישגי קריאה מאוחרים יותר. יכולות יסוד אלה מתחילות להתפתח בשנות הגן, לפני כניסה להוראה פורמלית, והן חיוניות להצלחה לימודית; כאשר הן אינן מתפתחות כראוי, צפויים קשיים בקריאה (Landerl et al., 2022). זיהוי והתערבות מוקדמים הם קריטיים, שכן מחקרים מראים באופן עקבי תוצאות טובות יותר כאשר תמיכה מתחילה בגן או בכיתה א' ולא לאחר שהקשיים התבססו (Gaab et al., 2019; Lovett et al., 2017; Ozernov-Palchik & Gaab, 2016).

סימפוזיון זה מציג את מחקר האורך הראשון הישראלי באוריינות המשתרע על פני חמש שנים, ועוקב אחר ילדים דוברי עברית וערבית מהגן ועד כיתה ד'. בחינת העברית והערבית מספקת תובנות ייחודיות לגבי מנגנוני למידה אוניברסליים ותלויי-שפה כאחד, ומתייחסת לייצוג החסר של אוכלוסיות דוברות שפות שמיות במדע ההתפתחותי העולמי. עוצמה מתודולוגית מרכזית היא תקופת המעקב הממושכת. בעוד שרוב מחקרי האורך מגיל הגן עוקבים אחר התפתחות קריאה רק במהלך הכיתות הראשונות, מעטים נמשכים עד לשנים מתקדמות יותר בבית הספר היסודי כדי ללכוד את המעבר המלא מפענוח ראשוני לקריאה מיומנת.

במסגרת פרויקט זה, פיתחנו ותיקפנו בטריית הערכה התפתחותית מקיפה שתוכננה במיוחד עבור מערכות כתב שמיות, עם מהימנות גבוהה ($\alpha = .74-.97$) ונפרדת לכל שלב התפתחותי. בחנו מספר מטלות עבור כל מבנה (מודעות פונולוגית, ידע אותיות, קריאה מהירה, ידע אורתוגרפי, מודעות מורפולוגית, אוצר מילים), מה שצמצם שגיאת מדידה ואפשר זיהוי של מסלולי התפתחות מגוונים בקרב קוראים טיפוסיים וקוראים מתקשים כאחד.

הסימפוזיון מציג שלושה מחקרים: (1) ניבוי של התפתחות קריאת מילים בעברית מהגן, דרך כיתה א' ועד כיתה ד'; (2) ניבוי מהגן של התפתחות קריאה ואיות בערבית; (3) מנבאים מוקדמים להבנת הנקרא דרך מסלולים ישירים ומתווכים. במחקר הראשון (515 ילדים), מודעות פונולוגית ניבאה קריאה בכיתה א', בעוד שמודעות מורפולוגית ניבאה קריאה באופן ייחודי רק בכיתה ד' (תומכת במעבר לכתב לא מנוקד). במחקר השני (523 ילדים), קריאה ואיות בערבית הראו בסיסים משותפים מוקדמים (מודעות פונולוגית וידע אותיות) בעוד שבשלבים מאוחרים נמצא כי קריאה תלויה יותר במהירות עיבוד (RAN) בעוד שאיות מתבסס על דיוק אורתוגרפי. במחקר השלישי (540 ילדים), מודעות מורפולוגית הראתה מסלול התפתחותי ייחודי, לא ניבאה קריאה בכיתה א' אך ניבאה באופן משמעותי הבנת נקרא בכיתה ד' דרך מספר מסלולים.

יחד, מחקרים אלה מדגימים את הגורמים הקוגניטיביים-לשוניים המוקדמים אשר תורמים להתפתחות אוריינות בשפות שמיות, וכיצד הערכה מוקדמת מקיפה מאפשרת ניבוי ארוך טווח של הישגי אוריינות. הממצאים מקדמים הן תיאוריה חוצת שפות והן פרקטיקה מבוססת מחקר לתכנון תוכניות זיהוי מוקדם והתערבות ממוקדת יעילות.



סימפוזיון

מנבאים מוקדמים לקריאה בעברית: מודל רב-מרכיבי מהגן עד כיתות א' ו-ד'

רותם ינון¹, דנה טל¹, שלי שאול¹, יניב קנט-מימון², תמי קציר¹
¹ מרכז אדמונד י. ספרא לחקר המוח ולקויות למידה, אוניברסיטת חיפה
² בית הספר ברוך איבצ'ר לפסיכולוגיה, אוניברסיטת רייכמן

מטרה: רוב מחקרי האורך על מנבאים טרום-אורייניים מתמקדים במעבר מהגן לשלבים מוקדמים של רכישת הקריאה (כיתות א'-ב'), דבר המותיר פערים בהבנת תוקף הניבוי ארוך הטווח של מדדים מוקדמים ככל שילדים מתקדמים מפענוח פונולוגי לזיהוי מילים שוטף. מחקר אורך זה של חמש שנים הוא הראשון בשפה העברית הבוחן ניבוי מהגן דרך כיתה א' ועד כיתה ד', ומטרתו להעריך את התרומה הדיפרנציאלית והמבנית של מיומנויות קוגניטיביות-לשוניות בגיל הגן למסלולי התפתחות של דיוק ושוטף קריאת מילים.

שיטה: 515 ילדים דוברי עברית הוערכו בגן (גיל 5.9), בכיתה א' (גיל 6.8) ובכיתה ד' (גיל 9.7). מדידות הגן כללו מודעות פונולוגית, שיום מהיר, ידע אותיות ומודעות מורפולוגית. יכולות דיוק ושוטף קריאת מילים בכיתות א' ו-ד' נמדדו באמצעות כלים מותאמים התפתחותית: מילים בניקוד קמץ בלבד לקוראים מתחילים בכיתה א', ומילים מנוקדות ולא מנוקדות לקוראים מיומנים בכיתה ד'. מודלים של משוואות מבניות (SEM) בחנו ניבוי מקביל (גן → כיתה א') וניבוי לטווח ארוך (גן → כיתה ד', בשליטה על יכולות הקריאה בכיתה א'), תוך הבחנה בין תרומות ייחודיות של מרכיבים חופפים.

תוצאות: מודעות פונולוגית נמצאה כמנבא החזק ביותר לדיוק ושוטף הקריאה בכיתה א' ($\beta = .35-.40$), כאשר ידע האותיות ($\beta = .16-.20$) ושיום מהיר ($\beta = -.27$) תרמו אף הם בשלב מוקדם זה. מעבר לכך, מיומנויות טרום-אורייניות בגן המשיכו לנבא דיוק ושוטף קריאה בכיתה ד' גם לאחר שליטה ביכולות הקריאה בכיתה א'. באופן ספציפי, ידע האותיות ($\beta = .17-.21$) ושיום מהיר ($\beta = -.15$ to $-.21$) הציגו תרומות ניבוי עקביות, בעוד שמודעות מורפולוגית נמצאה כמנבא ייחודי לשלב התפתחותי זה ($\beta = .12-.21$).

מסקנות: הממצאים מדגימים כי מיומנויות טרום-אורייניות ניתנות למדידה מהימנה כבר בגיל הגן, וכי למרכיביו השונים תוקף ניבוי דיפרנציאלי ויציב לאורך התפתחות הקריאה. דפוסי הניבוי תומכים במודל רב-מרכיבי והיררכי של התפתחות הקריאה, ומדגישים את חשיבות ההתאמה ההתפתחותית של כלי מדידה בבחינת ניבוי אורייני ארוך-טווח. ממצאים אלו מצביעים על הפוטנציאל של מדדים טרום-אורייניים כבסיס לפיתוח כלי הערכה והתערבות מוקדמים ומקיפים.



סימפוזיון

מסלולי התפתחות של קריאה ואיות בערבית מהגן ועד כיתה ד'

אפנאן ח'ורי-מטאניס¹, אבי קרני¹⁻²

¹ מרכז אדמונד י. ספרא לחקר המוח ולקויות למידה, אוניברסיטת חיפה

² בית הספר לרפואה ע"ש הרטה ופול עמיר, אוניברסיטת חיפה

מטרה

מחקר זה עוקב אחר המסלולים ההתפתחותיים המקשרים בין כישורים קוגניטיביים-לשוניים בגן לבין קריאה ואיות בערבית מכיתה א' עד כיתה ד'. המחקר מתמקד באוריינות לקסיקלית, קריאת מילים בודדות ואיות, מכיוון שבערבית המורכבות החזותית של הכתב והמורפולוגיה העשירה נשענות על תהליכים חופפים בשני התחומים. שאלות המחקר: (1) כיצד מנבאים מהגן: מודעות פונולוגית, שיום מהיר אוטומטי (RAN), מודעות מורפולוגית וידע אותיות, משפיעים ישירות ובעקיפין על תוצאות בכיתה א' וד' בקריאת מילים ואיות; (2) כיצד משתנים קשרי הניבוי הללו לאורך ההתפתחות.

שיטה וניתוח

עוקבה אורכית של 523 ילדים דוברי ערבית מ-73 גני ילדים בצפון ישראל (גיל ממוצע בבסיס = 70 חודשים) הייתה במעקב עד כיתה ד'. בגן נמדדו מודעות פונולוגית, RAN, ידע אותיות ומודעות מורפולוגית. התוצאות היו דיוק בקריאת מילים ואיות בכיתה א', ודיוק ושטף בכיתה ד'. שני מודלים של משוואות מבניות (קריאה; איות) השתמשו בדיוק כיתה א' כמתווך מהגן לכיתה ד'. התאמת המודל הייתה מצוינת.

תוצאות

קריאה: דיוק בכיתה א' הוסבר בעיקר ממודעות פונולוגית ($\beta=.469$) עם תרומה נוספת מידע אותיות (RAN), $\beta=.20$ מהיר יותר קשור לדיוק טוב יותר ($\beta=-.13$). בכיתה ד', דיוק כיתה א' ניבא דיוק מאוחר ($\beta=.29$) ו-RAN- תמך באופן ייחודי בשטף ($\beta=-.19$). המודל הסביר כ-44% מהשונות בכיתה א' ו-34% משטף כיתה ד'.

איות: דיוק בכיתה א' נובא ממודעות פונולוגית ($\beta=.28$) וידע אותיות ($\beta=.16$) בכיתה ד', ידע אותיות היה המנבא הישיר החזק ביותר לדיוק איות ($\beta=.27$) ודיוק כיתה א' נשאר יציב ($\beta=.26$). השונות המוסברת הייתה נמוכה יותר (~25% בכיתה א'; ~20% בשטף כיתה ד'). המתאמים בין קריאה לאיות ירדו מ- $r=.75$ (כיתה א') ל- $r=.54$ (כיתה ד'), מה שמעיד על התמחות התפתחותית.

מסקנות

אוריינות בערבית מראה יסודות משותפים מוקדמים, מודעות פונולוגית וידע אותיות התומכים בדיוק בשני התחומים, ולאחר מכן צמיחה ספציפית לתחום: קריאה תלויה יותר ויותר במהירות עיבוד ואוטומטיות (RAN) לשטף, בעוד שאיות מגבש דיוק אורתוגרפי המעוגן בידע דפוסי אותיות. הממצאים מצביעים על כך שאיות מיומן בערבית נשען על מנגנונים נוספים מעבר למודעות פונולוגית/RAN/מודעות מורפולוגית/ידע אותיות, כולל זיכרון אורתוגרפי, ייצוגים מורפו-אורתוגרפיים יציבים, והבחנה חזותית עדינה של צורות אותיות דומות.

השלכות פדגוגיות: הוראה שיטתית של מודעות פונולוגית והכרת האלפבית צריכה להתחיל בגן ולהימשך בכיתות הנמוכות; בכיתות ג'-ד' ההוראה צריכה להתמקד - בניית שטף ופענוח קצוב לקריאה, ועבודה מפורשת על זיכרון אורתוגרפי (תרגול איות חוזר, דפוסים מורפו-אורתוגרפיים, וקשב חזותי לאלוגרפיה וסימנים דיאקריטיים) לאיות.



סימפוזיון

מודל התפתחותי רב-מרכיבי של הבנת הנקרא בעברית מכיתות א'

עד ד'

לירון חזן כהן, תמי קציר, שחר דותן, רותם יינון
מרכז אדמונד י. ספרא לחקר המוח ולקויות למידה

מטרה

על אף התמיכה האמפירית החזקה למודל הפשוט לקריאה (Simple View of Reading; Gough & Tunmer, 1986), נותרו פערים משמעותיים בהבנת האופי ההתפתחותי והמבנה הרב-ממדי של כל אחד מרכיביו. מרבית המחקרים בחנו את רכיבי הקריאה כמבנים סטטיים בנקודת זמן אחת, ומחקרי האורך הקיימים בעברית עקבו אחר ילדים לתקופות קצרות יחסית, מבלי ללכוד את המסלולים ההתפתחותיים המלאים במהלך שנות הקריאה המכריעות. המחקר הנוכחי נועד למלא פערים אלה באמצעות התמקדות בשלושה היבטים: ראשית, אימוץ נקודת מבט התפתחותית על ידי מעקב אורכי מכיתה א' עד כיתה ד', תקופה המאופיינת במעבר קריטי מכתב מנוקד (מאופיין בשקיפות אורתוגרפית) לכתב לא מנוקד (מאופיין בעומק אורתוגרפי) ובהתפתחות המודעות המורפולוגית; שנית, בחינת המבנה הרב-מרכיבי של זיהוי מילים והבנה לשונית תוך מיפוי מסלולים ישירים ועקיפים להבנת הנקרא; ושלישית, התייחסות למאפיינים הייחודיים לשפה העברית.

שיטה

המחקר עקב אחר 540 ילדים מכיתה א' ועד כיתה ד'. מדדי כיתה א' כללו מודעות פונולוגית, שיום מהיר ואוטומטי של אותיות (RAN), ידע אורתוגרפי, מודעות מורפולוגית, אוצר מילים ושטף קריאת מילים. בכיתה ד', נבדקו שטף קריאת טקסט, אוצר מילים והבנת הנקרא. המחקר השתמש במודל משוואות מבניות (SEM) על מנת לבחון מסלולים התפתחותיים ישירים ועקיפים בין המשתנים לאורך זמן.

תוצאות

מודל ה-SEM חשף מסלולים התפתחותיים נפרדים לרכיבי המודל הפשוט לקריאה. במסלול זיהוי המילים, מודעות פונולוגית, ידע אורתוגרפי ו-RAN בכיתה א' ניבאו את שטף קריאת המילים, שבתורו תרם להבנת הנקרא בכיתה ד' הן במישורין והן בעקיפין דרך שטף קריאת טקסט ואוצר מילים בכיתה ד'. במסלול ההבנה הלשונית, מודעות מורפולוגית בכיתה א' ניבאה הבנת נקרא בכיתה ד' באופן ישיר ודרך מסלולים עקיפים בתיווך שטף קריאת טקסט ואוצר מילים בכיתה ד', בעוד שאוצר המילים בכיתה א' ניבא הבנת נקרא רק דרך מסלול עקיף. המודל הסופי הסביר 40.9% מהשונות בהבנת הנקרא בכיתה ד'.

מסקנות

המודל המדורג שפותח במחקר זה חושף את הטבע הרב-ממדי של רכיבי המודל הפשוט לקריאה, הפועלים דרך מסלולים התפתחותיים משולבים זה בזה. ממצא מרכזי מעיד על הופעה מאוחרת של המודעות המורפולוגית כמנבא להבנת הנקרא: למרות שלא תרמה ליכולות קריאה בכיתה א', היא ניבאה הבנת נקרא בכיתה ד' הן במישורין והן בעקיפין. תופעה זו משקפת את המעבר ההדרגתי מתלות בניקוד לקריאה בכתב לא מנוקד. לממצאים השלכות חינוכיות בהבנת הצורך בגישה משולבת להוראת הבנת הנקרא, המתחשבות בהתפתחות ההדרגתית של רכיביה ובזיהוי חוזקות וחולשות בכל רכיב. המודעות המורפולוגית המוקדמת מזוהה כמיומנות קריטית התומכת הן בהרחבת אוצר המילים והן בהתפתחות הקריאה בכיתות הגבוהות. שטף קריאת המילים בכיתה א' מהווה מנבא אמין ליכולות קריאה והבנת הנקרא לאורך זמן.



שיפור תהליך המיון להתמחויות כירורגיות ברפואה: הערכה של כישורים טכניים, יכולות קוגניטיביות ומאפייני אישיות באמצעות כלים מבוססי מציאות מדומה וגיימיפיקציה

עבודת הדוקטורט של נועה גזית, כלת פרס אפי האוניברסיטה העברית, מרכז ארצי לבחינות ולהערכה (מאל"ו)

מיון מועמדים להתמחויות כירורגיות הוא שלב קריטי להבטחת בטיחות ואיכות הטיפול הרפואי. המטרה היא לזהות את המועמדים בעלי הפוטנציאל הגבוה ביותר להצליח במהלך ההתמחות ובהמשך כרופאים מנתחים. עם זאת, ברוב תוכניות ההתמחות בעולם, המיון מתבסס על כלי הערכה לא אפקטיביים (כגון קורות חיים, הישגים אקדמיים, מכתבי המלצה וראיונות לא מובנים), שלהם מתאם נמוך עם ביצוע עתידי.

העבודה הנוכחית כללה שלושה שלבים שבהם נערכו ארבעה מחקרים. בשלב הראשון בוצע ניתוח עיסוקים מקיף, במטרה לזהות את הכישורים, היכולות והמאפיינים הדרושים לרופא מנתח במאה ה-21, מעבר לתחומים הכירורגים השונים (מחקר 1). הניתוח התבסס על ראיונות עם 104 רופאים מנתחים מומחים, שאפשרו לזהות 24 כישורים, יכולות ומאפיינים: חמישה כישורים טכניים, שש יכולות קוגניטיביות ושלושה עשר מאפייני אישיות. הרשימה תוקפה באמצעות שאלון שהועבר ל-1,102 מנתחים ומתמחים.

בשלב השני פותחו, בהתבסס על ממצאי ניתוח העיסוקים, שני מבחנים חדשניים: (1) מבחן להערכת יכולת טכנית באמצעות סימולטור לפרוסקופיה; (2) מבחן ממוחשב מבוסס גיימיפיקציה להערכת יכולות קוגניטיביות ומאפייני אישיות. מבחנים אלו משלבים מציאות מדומה ואלמנטים משחקיים, ומציעים יתרונות כגון מוטיבציה גבוהה של הנבחנים, איסוף נתונים התנהגותיים ברזולוציה גבוהה והערכה אוטומטית ולא מוטה.

בשלב השלישי נאספו עדויות תוקף לשני המבחנים שפותחו בשלושה מחקרים (מחקרים 2-4) על פי ההמשגה העדכנית של מושג התוקף המתוארת בסטנדרטים לפיתוח מבחנים חינוכיים. העדויות כללו עדויות לתוכן, לתהליכי תגובה, למבנה הפנימי ולקשרים עם משתנים אחרים, לצד בחינת היבטי ישימות, קבלה (acceptability) והוגנות. במחקרים אלו השתתפו קבוצות רופאים ברמות שונות של ותק וניסיון: 30 מומחים ו-152 סטאז'רים (מחקרים 2-3), ו-76 סטאז'רים, 75 מתמחים ו-65 מומחים (מחקר 4).

באופן כללי, ממצאי העבודה סיפקו עדויות משמעותיות לתוקף, לישימות, לקבלה ולהוגנות של המבחנים, ותמכו בשימוש במבחנים אלו עבור מיון מועמדים להתמחויות כירורגיות. ממצאים אלו מעידים על כך שלמבחנים אלו יש פוטנציאל לשפר את תהליך המיון להתמחויות כירורגיות, מה שבסופו של דבר יוביל לשיפור באיכות כוח האדם במחלקות הכירורגיות ולשיפור באיכות הטיפול שמקבלים מטופלים.



שימושי בינה מלאכותית במערכות תומכות למידה

ביאטה בייגמן קלבנוב
ETS Research Institute Edusoft

שפה הינה אמצעי של למידה, לכן ניתוח ופיתוח אוטומטיים של תוצרי שפה שמלווים למידה, כגון טקסטים ותרגילים שמהם לומדים, קטעי צ'ט או שיח שמתקיימים בין תלמידים או בין תלמיד למורה, וכן תשובות של התלמיד לשאלות, עשויים לסייע בארגון, ייעול, והערכה של למידה. בהרצאה אסקור כעשור של מחקר שנעזר בבינה מלאכותית ועיבוד שפה טבעית כדי לתמוך בתהליכי למידה ואתן דוגמאות ממחקר ופיתוח שהובילי של מערכת שתומכת ברכישת שטף קריאה. בנוסף, נתבונן ביחד קדימה כדי לשרטט כיווני התפתחות מבטיחים שמשמשים ביכולותיה הייחודיות של בינה מלאכותית יוצרת למלא אחר הוראות מילוליות ולפתח שיח טבעי ופתוח במגוון נושאים.



מחקר תיקוף חוזר של מת"ל (מערכת תפקודי למידה) לאבחון לקות למידה והפרעת קשב במבוגרים

יערה לדור ויצמן, ענבל חת, ענת בן-סימון
מרכז ארצי לבחינות ולהערכה (מאל"ו)

מת"ל היא מערכת ממוחשבת של מבחנים קוגניטיביים ושאלונים, המיועדת לאבחון לקות למידה או הפרעת קשב בקרב מועמדים להשכלה הגבוהה וסטודנטים. מטרת האבחון היא להמליץ על זכאות להתאמות בבחינה הפסיכומטרית ובבחינות האקדמיות, וכן לגבש המלצות לסיוע לסטודנטים במוסדות להשכלה גבוהה. המערכת נועדה לאבחן שלוש לקויות למידה: לקות קריאה (דיסלקסיה), לקות כתיבה (דיסגרפיה), לקות בעיבוד נומרי (דיסקלקוליה), וכן להעריך את הסבירות לקיומה של הפרעת קשב. בדוח אבחון מת"ל מדווחים שני סוגי אבחנה: (1) 'אבחנה ממוחשבת', המבוססת על מבחנים קוגניטיביים ומופקת על-ידי אלגוריתם; ו-(2) 'אבחנה קלינית', הנקבעת על-ידי מאבחנת, ומבוססת על אינטגרציה של תוצאות אבחון מת"ל עם מקורות מידע נוספים. האבחנה הקלינית היא האבחנה הקובעת.

מת"ל משמשת מזה קרוב לשני עשורים כלי מרכזי להערכת לקויות למידה והפרעת קשב בקרב בוגרים בישראל. הגרסה הראשונה של מת"ל הושקה ב-2007, וב-2013 בוצע מחקר תיקוף שכלל גם ניתוח גורמים. ב-2018 הושקה גרסה מקוונת של מת"ל, שכללה לראשונה שימוש בעשרה גורמי תפקוד לחישוב האלגוריתם המשמש להפקת האבחנה הממוחשבת. הגורמים ששימשו להפקת האבחנה הממוחשבת של רמת התפקוד ומשקלוליהם היו אלה שהניבו את ההבחנה המדויקת ביותר בין קבוצת הפונים עם לקות לבין קבוצת הפונים ללא לקות. משום שמספר הפונים לאבחון המת"ל הולך ועולה בשנים האחרונות, ובשל ההשלכות המרחיקות לכת של תוצאות האבחון על מתן זכאות להתאמות בבחינות, יש חשיבות עליונה לתיקוף חוזר של כלי האבחון, ובכלל זה בעדכון משוואות הניבוי להפקת אבחנה ממוחשבת המסייעת למאבחנות בביצוע אבחנה קלינית.

בתום למעלה מעשור מביצוע מחקר התיקוף החוזר, שבמהלכו חלה בהלימה בין האבחנה הממוחשבת לאבחנה הקלינית, הוחלט לבצע מחקר תיקוף חוזר, תוך שימוש במדגם של 15,296 צעירים שהשלימו אבחון מת"ל בעברית בשנים 2024-2022. מטרת המחקר הן: לבחון שינויים בשכיחות ארבע הלקויות המאובחנות על-ידי מת"ל ובהתפלגות הביצוע של פונים עם כל אחת מארבע הלקויות במדדי ובגורמי הביצוע של מת"ל, לערוך תיקוף חוזר של מדדי הביצוע של מבחני מת"ל והגורמים המופקים מהם, ולעדכן את משוואות הניבוי כדי לשפר את האבחנה הממוחשבת, ולצמצם הפער בינה לבין האבחנה הקלינית.

תוצאות המחקר מצביעות על שינויים בשכיחות של ארבע הלקויות המאובחנות על-ידי מת"ל שהבולטים ביניהם הם: עלייה ניכרת בשכיחות של אבחנה של הפרעת קשב וירידה בשיעורי תחלואה משותפת של לקויות, על חשבון אבחנה יחידה של הפרעת קשב. מתוך ניתוח ההתפלגות של מדדי הביצוע במבחני מת"ל עולות כמה תופעות בהשוואה למחקר התיקוף הקודם, ועיקרן בתחום הכתיבה, קצב העיבוד והקשב. תוצג ההלימה בין האבחנה הקלינית לבין האבחנה הממוחשבת, בהתייחס לכל אחת מארבע הלקויות. יוצג ניתוח הגורמים העדכני, בו זוהו שמונה גורמים, וכן ההבדלים העיקריים בין ניתוח הגורמים שבוצע לפני כעשור לבין הניתוח העדכני. בהתבסס על תוצאות ניתוח הגורמים המעודכן הופקו משוואות לניבוי כל אחת מארבע הלקויות באמצעות גרסיה לוגיסטית, תוך שימוש במדדי הביצוע ובמשקלולות שהניבו את הניבוי האופטימלי לכל אחת מקבוצת הלקות. תוצג בדיקה של רמת ההלימה בין האבחנה הממוחשבת לאבחנה הקלינית כתוצאה משימוש במשוואות ניבוי עדכניות.



פחות אבל מספק? השפעת צמצום מספר התחנות במבחן הסמכה מבוסס סימולציה לאחיות

דורון שגיא

רן עמרם

מסר – המרכז הארצי לסימולציה רפואית

מרכז ארצי לבחינות ולהערכה (מאל"ו)

מבחן ההסמכה העל-בסיסי באחיות הוא מבחן מעשי בסיכון גבוה, המיועד לאחיות ואחים מוסמכים המבקשים להתמחות באחד מהמקצועות להם נדרשת התמחות נוספת, כפסיכיאטריה, טיפול נמרץ, חדר ניתוח וכדומה. במבחן זה הנבחנים עוברים בין 6-7 תחנות, מתוך 4-5 תחנות סימולציה שבהן עליהם להדגים יכולות ומיומנויות הנדרשות כדי לטפל במטופל (המדומה על ידי שחקן או דגם / סימולטור), ו-2 תחנות ניהול מקרה מול מחשב. המצב הרפואי בכל תחנה הוא "מקרה מייצג" של תחום אחד או יותר מתוך עולם הידע של המקצוע הנבחן. הבוחנים הקובעים את ההערכה בתחנות המעשיות הם מומחים, אנשי מקצוע מהמקצוע הנבחן.

יתרונו הגדול של המבחן המעשי הוא הערכת מיומנויות ויכולות (ולא רק ידע) בסביבה מעשית תוך שמירה על התנאים הנוקשים הנדרשים ממבחני סיכון גבוה. אולם קיימים קשיים ומגבלות למבחן זה, שהמרכזיים שבהם הם דגימת תוכן, המתבטאת במספר התחנות, והשקעת משאבים. מבחן מעשי רחב היקף דורש משאבים רבים, כתשתית פיזית, כוח אדם מקצועי רב, פיתוח נוסחי בחינה רבים ועוד, וככל שהדגימה רחבה יותר כן נדרשים משאבים רבים יותר. במבחנים מסוג זה מתקיים דיון קבוע לגבי נקודת האיזון הנכונה בין היקף הדגימה לבין היכולת לגייס את המשאבים הנדרשים לפיתוח המבחן ולהעברתו.

בשנים האחרונות ערכנו, מסר ומאל"ו, שלוש עבודות מחקר שעניינן לבדוק את ההשפעה של שינוי מספר התחנות במבחן על המדדים הפסיכומטריים שלו. שינויי המבנה בוצעו או הוצעו כדי לצמצם את המשאבים הנדרשים לקיום הבחינה, על רקע עלייה במספר הנבחנים וימי הבחינה.

שתי העבודות הראשונות התייחסו לביטול של תחנת ידע. כשלעצמה תחנה זו אינה תחנה מעשית אך היא מגדילה את היקף דגימת התוכן. בעבודה הראשונה, שבוצעה לפני ביטול התחנה, מצאנו כי הביטול צפוי להוריד רק במעט את תוקף הבחינה, שנבדק אל מול הציון במבחן ההסמכה הממשלתי, אותו מבצעת כל אחות בסיום לימודיה כדי לקבל תואר אחות מוסמכת. בעבודה השנייה, שבוצעה לאחר ביטול תחנת הידע, מצאנו כי השינוי הוריד באופן זניח את המהימנות הפנימית של הבחינה ואת הציון הממוצע.

בעבודה השלישית מצאנו כי ביטול תחנה מבוססת סימולטור, תחנה מעשית הדורשת משאבי תפעול רבים, צפוי להוריד רק במעט את המהימנות הפנימית של הבחינה ולהעלות באופן זניח את שיעור הנכשלים.

לסיכום, הממצאים מראים כי לביטול תחנה במבחן העל-בסיסי, מבחן מעשי ורחב היקף שמטרתו הערכת מיומנויות ויכולות, השפעה קטנה על מדדי הבחינה. תוצאות אלה נותנות בידי מקבלי החלטות כלי מבוסס לצמצום המשאבים הנדרשים לבחינה, מבלי לפגוע במדדים הפסיכומטריים שלה או בהשלכותיה. העבודה יכולה לשמש גופים אחרים המקיימים מבחנים או מרכזי הערכה בעלי אופי דומה.



CogniMentor: הערכת מדדי זיכרון עבודה ויזואלי-מרחבי ובדיקת שינויים לאורך זמן

יהודית לירון, דוד צוריאל, עודד מאירון
אוניברסיטת בר-אילן

ילדים עם הפרעת קשב (ADHD) מתמודדים עם קשיים בתפקודם היומי, כאשר ליקויים בזיכרון עבודה ויזואלי-מרחבי ובוויסות עצמי מהווים חלק ממאפייני ליבה שכיחים בקרב אוכלוסייה זו. בעוד שהאבחונים הסטנדרטיים מספקים מידע חיוני ומפורט לצוותים החינוכיים, דרכי ההתערבות המקובלות אינן תמיד מתייחסות לפוטנציאל הלמידה של התלמיד כמדד לבחינת יעילות ההתערבות.

מחקרים קודמים הראו כי יישום עקרונות האבחון הדינמי והמודל הטרנסאקציוני מוביל לשיפור במדדי זיכרון עבודה ויזואלי-מרחבי ואינטליגנציה פלואידית בקרב ילדים עם התפתחות תקינה. עם זאת, הכלים הפיזיים ששימשו למדידות אלו, הציבו אתגרים מתודולוגיים באיסוף מדדי תהליך מדויקים, דבר העשוי להשפיע על התוצאות ולהגביל את מהימנות היישום במצבים משתנים.

כמענה לאתגרים אלו, תוצג בהרצאה מערכת CogniMentor המפותחת במסגרת מחקר דוקטורט בפקולטה לחינוך בשיתוף הפקולטה למדעי המחשב באוניברסיטת בר-אילן. המחקר הבוחן את תרומת התיווך האנושי במהלך אימון ממוחשב של זיכרון עבודה ויזואלי-מרחבי בקרב תלמידים עם ADHD (כיתות ה'). המערכת מבוססת על עקרונות האבחון הדינמי, המודל הטרנסאקציוני ומבחן הרצפים (SQ-II) מתוך סוללת ה-CMB, בסביבה ממוחשבת להערכת זיכרון עבודה ויזואלי-מרחבי. מטרתה לספק פלטפורמה לאימון זיכרון עבודה ויזואלי-מרחבי רב-ממדי ולקידום כושר ההשתנות הקוגניטיבי. האימון מתבצע באמצעות משחק מחשב המשלב עקרונות מישחק לשימור המוטיבציה לאורך זמן.

ארכיטקטורת המערכת ומדדים פסיכומטריים

המערכת מורכבת משני ממשקים: ממשק למשתתף לביצוע האימון וממשק לחוקר לניהול ומעקב. מכיוון שהמערכת מיועדת בראש ובראשונה לילדים עם הפרעת קשב ובהמשך לילדים עם הפרעות נוירו התפתחותיות נוספות, תהליך הפיתוח התמקד במתן מענה לאתגרים הייחודיים המאפיינים אוכלוסייה זו. לשם כך, הושם דגש מיוחד על הפחתת העומס הוויזואלי, התאמת גופנים, שילוב מנגנונים למיקוד קשב וקביעת משך הזמן האופטימלי להצגת הודעות. עקרונות עיצוב אלו נועדו להבטיח שהמדדים הנאספים ישקפו נאמנה את יכולות העיבוד הקוגניטיביות של הילדים, ולא יושפעו לרעה מקשיי הקשב והעומס החושי.

לסיכום, מערכת CogniMentor תומכת במחזור הערכה והתערבות רב-שלבי לאורך זמן להעצמת תפקודים ניהוליים ומאפשרת אימון בו-זמני של זיכרון עבודה ויזואלי-מרחבי, ויסות עצמי ותכנון ופתרון בעיות. ייחודה של המערכת הוא ביכולתה להפיק פרופיל פסיכומטרי מדויק המבוסס על אבחון תפקודי זיכרון עבודה ויזואלי-מרחבי בעומסים משתנים. הפרופיל הפסיכומטרי המופק במערכת כולל מדדי התנהגות כמותיים כגון: זמני תגובה, דפוסי תנועה ודיוק דיפרנציאלי ומדדים קוגניטיביים-ניהוליים הקשורים לזיהוי מטרות ויזואליות-מרחביות. בנוסף, המערכת כוללת שאלון רגשות ומבחן למילוי הוראות, המאפשרים לקבל תמונה מקיפה על יכולותיו ומצבו של המשתתף בדגש על היבטים של לבקרה רגשית ותפקודים שפתיים.



מדד חוזק האקלים הארגוני (ICC): כלי משלים לסטיית התקן להבנת לכידות והסכמה תפיסתית ולדירוג יחידות במשטרת ישראל

אולגה אנטל קטייב
משטרת ישראל

העבודה מציגה שימוש מושכל במדד ה-Intraclass Correlation Coefficient (ICC) ככלי משלים למדדים תיאוריים מקובלים, לצורך העמקת ההבנה של חוזק אקלים ארגוני ביחידות משטרתיות. יודגש כי ה-ICC אינו מדד אקלים ארגוני בפני עצמו, אלא מדד פסיכומטרי להערכת מידת ההסכמה והמהימנות בין מדידות בתוך קבוצה, אשר יישומו בהקשר זה מאפשר העשרה פרשנית מעבר לחישוב ממוצעים וסטיית תקן בלבד.

מדדי אקלים ארגוני מקובלים מתמקדים לרוב בכיוון הכללי של העמדות (חיובי/שלילי), אך אינם בוחנים האם מתקיימת תפיסה משותפת בקרב חברי היחידה. בדרך כלל, מקובל להסתמך על ממוצעים וסטיית תקן כדי להעריך שביעות רצון, אמון בפיקוד או תחושת שייכות. עם זאת, מדדים אלו מתמקדים ברמת העמדה הממוצעת ואינם מאפשרים להבחין האם העובדים בתוך אותה יחידה חולקים תפיסה דומה, או שמא מתקיימים פערים פנימיים משמעותיים. סטיית התקן אמנם משקפת פיזור תשובות, אך היא מושפעת מגודל המדגם ואינה מבדילה בין שונות אקראית לבין דפוס קבוצתי עקבי. ה-ICC, לעומת זאת, בוחן את החלק היחסי של השונות בין יחידות לעומת השונות בתוך יחידות, ובכך מאפשר להעריך עד כמה חברי היחידה "רואים את המציאות הארגונית באותו האופן".

מדד ה-ICC מבוסס על פירוק השונות הכוללת לשני רכיבים: שונות בין יחידות ושונות בתוך יחידות. ערך המדד משקף את החלק היחסי של השונות שניתן לייחס להשתייכות ליחידה מסוימת. ערך ICC גבוה מעיד כי מרבית השונות נובעת מהבדלים בין יחידות, כלומר שקיימת הסכמה תפיסתית גבוהה בתוך כל יחידה. לעומת זאת, ערך ICC נמוך מצביע על כך שהשונות העיקרית היא בתוך היחידות עצמן, ומרמז על פערים פנימיים בין עובדים החולקים מסגרת ארגונית משותפת.

במסגרת ההרצאה תינתן הקדמה מושגית למדד ה-ICC, תוך הבהרת ההבדל בין שימוש ב-ICC, כמדד מהימנות טכני לבין שימוש בו ככלי פרשני להבנת חוזק אקלים. בנוסף, יובהרו יתרונות ה-ICC על פני שימוש בסטיית תקן בלבד, תוך הדגשה כי אין לראות במדדים אלו תחליפים אלא משלימים.

היישום האמפירי מבוסס על נתוני סקר האקלים הארגוני של משטרת ישראל (2025), הכולל מאות יחידות. עבור כל יחידה חושבו ממוצע, סטיית תקן ו-ICC בנושאים כגון אמון במפקדים, יחסים בין-אישיים, סביבת עבודה ותחושת שייכות. היחידות דורגו בשלוש גישות: דירוג לפי ICC בלבד, דירוג לפי סטיית תקן בלבד, ודירוג משולב של ממוצע ו-ICC לאחר נרמול.

הממצאים מראים כי שימוש ב-ICC, משנה את פרשנות הנתונים. כך למשל, יחידות בעלות ממוצע נמוך ו-ICC גבוה משקפות הסכמה רחבה על חוויה שלילית משותפת – ממצא המעיד על בעיה מערכתית ולא על פיזור אקראי. לעומתן, יחידות בעלות ממוצע גבוה ו-ICC, נמוך מצביעות על פערים פנימיים משמעותיים, גם כאשר התמונה הממוצעת נראית חיובית. הבחנה זו אינה מתאפשרת באמצעות סטיית תקן בלבד.

שילוב ה-ICC במדידת אקלים ארגוני מרחיב את ארגז הכלים האנליטי, מאפשר דירוג יחידות על בסיס חוזק ולכידות תפיסתית, ותורם להבנה עמוקה יותר של דפוסי תפקוד, מורכבות ואתגרים ארגוניים. שילוב המדד עשוי לסייע למקבלי החלטות בזיהוי מוקדם של יחידות במצוקה, ובפיתוח התערבויות מותאמות לחיזוק החוסן הארגוני.



מעבר ממבחן בשיטה הקלאסית לתת"פ: העברת תטאות לסולם יע"ל הרב-שנתי

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

מבחן יע"ל (יעלנט בגרסתו הממוחשבת) הוא מבחן בדיעת השפה העברית המועבר על ידי המרכז הארצי לבחינות ולהערכה (מאל"ו). המבחן מיועד למועמדים להשכלה גבוהה שנבחנו בבחינה הפסיכומטרית בשפה שאינה עברית. לאחרונה התעורר הצורך להגדיל באופן משמעותי את מגוון הפרקים שעוברים במבחן יעלנט הממוחשב והוחלט כי בניית הפרקים החדשים תיעשה באופן אוטומטי מתוך מאגר הפריטים הקיימים. לשם כך ולשם יצירת פרקים שיהיו ברי השוואה למבחן הקודם נדרש לסווג את הפריטים מבחינת המאפיינים הפסיכומטריים שלהם כמו קושי, מידת ההבחנה וכו'. הדבר בוצע באמצעות התאמת מודל תת"פ (IRT) תלת-פרמטרי ואמידת הפרמטרים של כ-2,700 פריטים קיימים.

בפרקים החדשים של יעלנט, ציוני הנבחנים יחושבו באמצעות אומדן היכולת החבויה (תטא) לפי מודל תת"פ, שכן הוא מספק מדד מהימן יותר מהציון הקלאסי ומאפשר השוואה טבעית בין נוסחים ללא צורך בכיול נוסף. עם זאת, כדי לשמור על רצף והקבלה לציונים שנמדדו בעבר, יש להמיר את ערכי התטא לסולם הציונים הרב-שנתי של יע"ל (150-50, ממוצע 100, סטיית תקן 20). המעבר בין סולם התטאות לסולם הציונים הרב-שנתי של יע"ל אינו טריוויאלי, משום שבשיטה הקלאסית כיול הציון השתנה מנוסח וכלל גם טרנספורמציות לא ליניאריות ("הברכה") לקצוות הסקאלה. האתגר המרכזי הוא קביעת טרנספורמציה שתשמר את מבנה הציונים הקיים ותבטיח שהמשמעות של ציוני יעלנט במתכונת החדשה תישאר זהה לזו של הציונים בעבר.

לצורך מציאת הטרנספורמציה האופטימלית נערך מחקר אשר התבסס על נתוני 1,853 פריטי בחינה אשר נענו על ידי 343,634 נבחנים במבחני יע"ל ויעלנט שנערכו בשנים 2010-2023. לכל נבחן חושבה תטא בשיטת הנראות המקסימלית (ML), בנוסף לציונו המקורי בסולם הרב-שנתי. לאחר מכן, נבדקו טרנספורמציות שונות של סולם התטאות, במטרה למצוא טרנספורמציה שמפיקה סולם הדומה ככל הניתן לסולם הרב-שנתי המקורי. דמיון זה נבדק על פי שני קריטריונים: א. גובה המתאם בין שני הסולמות ב. גודל ממוצע השאריות (ההפרש בין הציון המסולם לציון הרב-שנתי) בערך מוחלט.

נבדקו שלוש שיטות:

1. סילום לינארי פשוט - התאמת ממוצע וסטיית תקן (שיטת mean-sigma)
 2. גרגסיה לינארית - קביעת מקדמים לפי קו הרגרסיה בין שני הסולמות
 3. כיול אחוזונים - התאמת ההתפלגויות לפי מיקום אחוזוני של כל ערך בסולם
- בכל שיטה נבדקו גם טרנספורמציות עם הברכה (dog-legging) באחוזונים שונים במטרה לתחום את הערכים בקצות הסקאלה (50-150).

ניתוח הנתונים הצביע על הבדלים קטנים אך עקביים בין השיטות. הסילום הלינארי והסילום מבוסס הרגרסיה הניבו מתאם של קרוב ל- $r=0.977$ וממוצע שאריות סביב ארבע נקודות. שיטת האחוזונים הניבה את ההתאמה הגבוהה ביותר לסולם הקיים עם מתאם $r=0.980$ וממוצע שאריות של 3.8 נקודות. ניתוח התפלגות ההבדלים הראה כי עבור כ-75% מהנבחנים ההפרש בין הציונים החדשים לקודמים קטן מ-5 נקודות, ועבור 97% ההפרש קטן מ-10 נקודות.

התוצאות מצביעות על כך שכיוול האחוזונים הוא הדרך המתאימה ביותר להעברת סולם התטאות החדש לסולם יע"ל הרב-שנתי - שיטה זו שומרת על דמיון מרבי להתפלגות הציונים ההיסטורית, ובכך מבטיחה שסולם הציונים החדש יהיה בר-השוואה ועקבי עם הסולם המקורי של הבחינה.



מדידה אדפטיבית באמצעות שאלונים רב-מימדיים - בחינה אמפירית ראשונית בשני שאלונים

איתמר גתי, אלעד פאוסט
האוניברסיטה העברית בירושלים

מבחן אדפטיבי ממוחשב (Computerized Adaptive Test) מתאפיין בהתאמה אישית של רמת הקושי לנבחן, כך שהאלגוריתם מעדכן בזמן אמת את הפריט הבא בהתאם לתשובות הקודמות, תשובה נכונה או שגויה. מבחנים אדפטיביים יעילים יותר ממבחנים סטנדרטיים, הן בחיסכון בזמן והן בדיוק הערכת המשתנה הנמדד. אחד ממאפייני הצעירים מדור ה-Z דודר האלפא הוא קיצור טווח הקשב והסבלנות למסרים מילוליים ולטקסטים ארוכים. לכך יש השלכות, במיוחד לגבי הערכת משתנים באמצעות שאלונים מילוליים.

מטרת המחקר היא לבחון אפשרות לפיתוח שאלונים אדפטיביים, שבדומה למבחנים אדפטיביים יציגו את הפריטים המאוחרים בהתאם לתשובות המשיב לפריטים קודמים, ולהעריך את ההשלכות על מהימנות ויעילות המדידה. האתגר המרכזי הוא שבשאלונים אין תשובה "נכונה", ולכן לא ניתן לסדר פריטים מהקל אל הקשה. בנוסף, בעוד שבמבחנים ההערכה מבוססת על אינדיקטור יחיד המשקף רמת ביצוע בהתאם למספר התשובות הנכונות במשתנה הנמדד, כלומר ציון יחיד (למשל, ציון כושר המספרי בפסיכומטרי), הרי שבשאלונים רב-מימדיים מתקבל פרופיל ציונים המשקף את מידת ההסכמה עם ההיגדים המייצגים כל מימד. לכן נדרש למצוא דרך שונה ליישום העקרונות שבבסיס מבחנים אדפטיביים בשאלונים (שכולל הערכת ה"מחיר" במהימנות המדידה, והערכת ה"חיסכון" הנבדלת בין נבדקים).

במחקר התמקדנו בשאלון האוריינטציות להשכלה גבוהה, Higher Education Orientation questionnaire (HEO; Willner et al., 2023), שאלון רב-מימדי המודד חמישה מניעים ללימודים במוסד להשכלה גבוהה (רכישת מקצוע, לחץ חיצוני, הרחבת השכלה, סביבה חברתית ויוקרה). השאלון כולל 25 פריטים, 5 פריטים לכל סולם. בגרסה האדפטיבית שפיתחנו נכללים 2 עד 4 פריטים בכל סולם, בהתאם לאלגוריתם ממוחשב המתבסס על תשובות המשיב לפריטים הראשונים (האלגוריתם יוצג בהרצאה). נמצא מתאם גבוה בין הציונים המבוססים על הגרסה האדפטיבית המקוצרת לבין הציונים המקוריים בשאלון המלא: במדגם של 1212 סטודנטים משנה א' שמילאו את שאלון האוריינטציות להשכלה הגבוהה (Willner et al., 2020) נמצא שהמתאם החיצוני בתוך סטודנט (מעבר לחמשת סולמות השאלון) היה 0.98. (תחום בין רבעוני 95-99), כאשר בשאלון האדפטיבי ציוני הסולמות התבססו במוצע על 11.4 פריטים בלבד לעומת 25 פריטים בשאלון המקורי (קיצור של כ-50%), תוך אובדן מזערי של דיוק ציוני הסולמות. כיוון שבגרסת נייר ועיפרון לא ניתן ליישם את מודל השאלון האדפטיבי, בדקנו גם גרסה מקוצרת של שלושה פריטים לכל סולם. חציון המתאמים בתוך סטודנט בין גרסת שלושה פריטים לכל סולם לשאלון המלא המקורי היה 0.99. (תחום בין רבעוני 98-99). חציון המתאם בתוך סטודנט בין המדידה הראשונה למדידה החוזרת לאחר 32 חודשים (test-retest) היה 0.96. בשאלון המקורי ו-0.94. בשאלון האדפטיבי, ו-0.95. בשאלון המקוצר. המתאמים בין ציוני הסולמות בשאלון המלא והאדפטיבי נעו בין 0.86 ל-0.92. בנוסף, טווחי מתאמי test-retest בתוך סולם מעבר לסטודנטים היו דומים בין הגרסאות, ונעו בין 0.62 ל-0.73. בגרסה המלאה, בין 0.57 ל-0.71. בגרסה המקוצרת לשלושה פריטים, ובין 0.56 ל-0.70. בגרסה האדפטיבית.

אם יישאר זמן, נציג בקצרה ממצאים דומים לגבי הדמיון בין הגרסה המלאה לאדפטיבית של שאלון האוריינטציות לעבודה (Work Orientation Questionnaire; Willner et al., 2020; N=605), שאלון שגם בו חמישה ממדים עם 5 פריטים בכל סולם.



מבחן "תנופה למחר": הערכת מיומנויות גנריות בסביבה פתוחה - אתגרים וממצאים ראשוניים

תמי סבג שושן, אימאן עואדיה, לימור בורוט, יואב ברגנר
הרשות הארצית למדידה והערכה (ראמ"ה)

מירב אריאלי-אטאלי, נדב כספי, לירון גביסון
One Assessment (לשעבר טלדור אססמנט)

פיתוח מבחן "תנופה למחר", המפותח על ידי ראמ"ה (הרשות הארצית למדידה והערכה בחינוך) בשיתוף one assessment, נועד למלא פער יסודי וקריטי בתחום ההערכה, ובכך לשרת את תפיסת "דמות הלמידה המתחדשת". בעוד הערכת ידע בתחומי הדעת מבוססת היטב, והערכת מיומנויות בזיקה לתחומי דעת הולכת ומתבססת, ניכר פער משמעותי ביותר ביכולת למדוד מיומנויות קוגניטיביות (כגון חשיבה ביקורתית, יצירתית, אוריינות מידע ואוריינות דיגיטלית) בהקשר חוצה-תחומים. שילוב הערכת מיומנויות אלו בסביבה פתוחה המאפשרת שימוש בכלי בינה מלאכותית באמצעות הטלפון הנייד של התלמידים, מעצימה באופן ניכר את האתגרים הפסיכומטריים הכרוכים בה.

המבחן נשען על ארבע תפיסות יסוד עכשוויות: תפיסה הוליסטית של מיומנויות החשיבה, הרואה בחשיבה ביקורתית, יצירתית ובאוריינות דיגיטלית מערכת אחת הפועלת באופן משולב; תפיסת הידע כמקור וכתוצר, המתמקדת בתהליך העבודה עם מידע; גישה חוצת-תחומים הרואה במיומנויות קוגניטיביות יכולות כלליות המתפתחות בהקשרים אותנטיים; ועקרונות ההערכה בסביבה פתוחה, הרואה בבינה מלאכותית מרכיב אינטגרלי בתהליך החשיבה והלמידה.

המבחן מעריך אוריינות מידע במדיה דיגיטלית דרך שלושה ממדי ליבה:

1. איתור והערכת מידע
2. ארגון ועיבוד מידע
3. ייצוג והצגת מידע

כל ממד נבחן באינטגרציה עם הערכה של חשיבה ביקורתית ויצירתית. המבחן מציג מצבים אותנטיים מהחיים, והתלמידים נבחנים דרך טלפונים ניידים או מחשבים עם גישה מלאה לאינטרנט וכלי AI. המבחן מיועד לתלמידי כיתות ו' וט', והוא כולל ארבעה נוסחים מקבילים עם פריטים פתוחים וסגורים.

תהליך פיתוח המבחן נמשך כשנתיים ונבנה כהליך שיטתי, מדורג ומבוסס ראיות. תהליך זה הדגיש את הצורך להתמודד עם האתגרים המהותיים בביסוס התוקף המבני של המבחן, הנובעים מהאופי האינטגרטיבי של ההערכה. בחודש יוני 2025 הועבר המבחן לראשונה במתכונת דיגיטלית, פתוחה ומקוונת. במחקר השתתפו 1,160 תלמידים דוברי עברית מכיתות ט' שנדגמו לפי סוג פיקוח, מחוז, רמה סוציו-אקונומית ומגדר. תהליך הצינון בוצע עלידי מעריכות מוכשרות, עם בדיקת מהימנות בין-שופטים על כ-50% מהתגובות. נוסף על כך נערכת הצלבה עם ציונים במבחנים בשפת אם ומדעים, לבחינת תוקף מתכנס ומבחין.

ציוני המבחן בארבעת נוסחיו היו דומים, עם ממוצעים של 47-50 וסטיות תקן של 23-26. נמצאה מהימנות פנימית גבוהה מאוד של הפריטים ($\alpha=0.91-0.94$), וכן רמה גבוהה של מהימנות בין-שופטים בפריטים הפתוחים. לא נמצאו הבדלים מובהקים בין בנים לבנות או בין היבחנות במחשב לעומת היבחנות בטלפון נייד, אך נמצאו פערים מובהקים לפי סוג פיקוח ולפי רקע חברתי-כלכלי, בדומה למגמות מוכרות במבחנים תחומיים. ממצאי ניתוח תוקף הצביעו על מתאמים בינוניים עם מבחני שפת אם ומדעים, ומתאמים גבוהים יותר עם רכיבי חשיבה ביקורתית-יצירתית יותר מאשר עם רכיבי ידע תחומיים. ממצא זה תומך בתוקף מתכנס ומבחין ומעיד כי המבחן מודד בעיקר מיומנויות חשיבה כלליות וחוצות-תחומים.

הממצאים הראשוניים מעידים על פוטנציאל יישומי ומחקרי גבוה של המבחן, לצד אתגרים פסיכומטריים ומתודולוגיים משמעותיים.



שאלון ה-AIR: פיתוח ותיקוף של שאלון למדידת מוטיבציות לקריאה בסיועה של בינה מלאכותית יוצרת

עדי ברן¹, שיר אתגר², יעל סידי¹

¹ האוניברסיטה הפתוחה

² אוניברסיטת תל אביב

השימוש ההולך וגובר בכלי בינה מלאכותית יוצרת (GenAI) מעצב מחדש את האופן שבו אנשים לומדים, צורכים ידע, וקוראים. כלי בינה מלאכותית יוצרת יכולים לשפר את היעילות, ההבנה, והמעורבות של המשתמשים, אבל גם מציבים סיכונים רבים למשתמשים, כגון צריכה של ידע שגוי או מוטה. ספציפית לגבי קריאה באמצעות GenAI, חלק מהמחקרים מוצאים כי קריאה כזו עוזרת להבנת הנקרא, ואילו חלק אינם מוצאים בקריאה כזו יתרונות. אנו מניחות כי הממצאים הלא עקביים הללו נובעים מהמוטיבציות השונות שמובילות אנשים להסתייע ב-GenAI לקריאה. מאחר שלא היה קיים כלי למדידת המוטיבציות השונות הללו, המחקר הנוכחי שאף לבנות כלי כזה, על ידי פיתוח ותיקוף של שאלון המאפשר מדידה של מוטיבציות לקריאה בסיוע בינה מלאכותית. השאלון נקרא AIR: AI for Reading.

בתחילת המחקר, יצרנו רשימה של 52 מוטיבציות שונות לקריאה, עליהן למדנו ממקורות שונים: החל מספרות אקדמית, דרך טקסטים ברשתות חברתיות, ועד להנחות נפוצות לגבי מוטיבציות לשימוש ב-GenAI לקריאה. הפריטים ברשימה הזו הפכו לשאלות הראשוניות בהן השתמשנו במחקר הראשון.

במחקר הראשון, העברנו שאלון הכולל את המוטיבציות הללו ל-425 משתתפי פרוליפיק, ועל תשובותיהם ביצענו ניתוח גורמים מגשש עם רוטציית varimax. המודל הסופי היה מודל של ארבעה פקטורים שונים, 261 פריטים. מודל זה הסביר 65.91% מהשונות. ארבעת הפקטורים השונים – המהווים ארבע מוטיבציות שונות לשימוש ב-GenAI בקריאה – היו: (א) מוכוונות משימה – זוהי המוטיבציה להשתמש שימוש אסטרטגי בבינה מלאכותית יוצרת בכדי לתמוך במטלות קריאה שונות. שימוש זה מתרחש מתוך רצון ללמוד ולבצע את המשימה על הצד הטוב ביותר. (ב) מוכוונות הנאה – זו המוטיבציה להשתמש בבינה מלאכותית יוצרת בכדי להנות או להרגיש "חלק מהטרנד" (ג) מאמץ נמוך – המוטיבציה להסתמך על בינה מלאכותית יוצרת כדי להפחית את ההשקעה הקוגניטיבית במטלה (ד) תרגום – זוהי המוטיבציה להשתמש בבינה מלאכותית יוצרת בכדי לשפר הבנה של טקסטים בשפות שאינן שפת האם של האדם.

במחקר השני, 414 משתתפי פרוליפיק נוספים ענו על 26 הפריטים שנמצאו במחקר הראשון. לאחר מכן, ביצענו ניתוח גורמים מאשש בכדי לאשש את מבנה ארבעת הפקטורים שנמצאו לפני כן. כל ארבעת הפקטורים שנמצאו במחקר הראשון קיבלו אישוש במחקר השני. לאחר הסרת ארבעה פריטים שהיו בקורלציות גבוהות זה עם זה, המודל הראה מדדי התאמה מצויינים, וכל תת סקאלה הראתה מהימנות סבירה עד גבוהה מאוד.

בכדי לתקף את שאלון AIR, בחנו את הקורלציות של תתי הסקאלות שלו עם שאלון של מוטיבציות לשימוש ב-AI ללמידה (תוקף מתכנס), ועם מולטיטאסקינג ולקויות למידה (תוקף מבחין). כצפוי, היה קשר חזק בין תתי הסקאלות של שאלון AIR לשאלון של מוטיבציות לשימוש ב-AI ללמידה, ולא נמצאו קשרים חזקים עם מולטיטאסקינג ועם לקויות למידה.

לסיכום, שאלון ה-AIR, שפותח בכדי למדוד מוטיבציות לקריאה בסיוע כלי בינה מלאכותית, מצטייר כמדד מהימן ותקף. המחקר מציע כי קיימות מוטיבציות רבות ושונות לפעולה, הלכאורה פשוטה, של קריאה בסיועה של בינה מלאכותית יוצרת, ומציע לא לראות את כל המסתייעים ב-AI לקריאה כמקשה אחת.



אנשי קשר ומציגים בכינוס אפי 2026

שם	דוא"ל	
יור אפי יור הכינוס	Einat.notea@gmail.com	עינת נוטע-קורן
יור הכינוס	MeiravA@taldor.co.il	מירב אריאלי-אטאלי
מרצה אורח	gal.chechik@gmail.com	גל צ'צ'יק
מושב א' כשהערכה פוגשת בינה מלאכותית	tzur@nite.org.il	יור - צור קרליץ
	ehadas@tauex.tau.ac.il	ערן הדס
	licol.zeinfeld@weizmann.ac.il	ליכל זיינפלד
	yoavbe@education.gov.il	יואב ברגנר
	tanya.nazaretsky@epfl.ch	טניה נזרצקי
מושב ב' ממיומנויות להישגים	yosy.machluf@gmail.com	יור - יוסי מחלוף
	Hila.segal5@gmail.com	הילה סגל
	tuval.raz@campus.technion.ac.il	תובל רז
	Zemira.Mevarech@biu.ac.il	זמירה מברך
	yoel@nite.org.il	יואל רפ
מושב ב' סימפוזיון: התפתחות מיומנות קריאה	katzirta@gmail.com	יור - תמי קציר
	rotemwr82@gmail.com	רותם ינון
	afnan38@gmail.com	אפנאן ח'ורי-מטאניס
	katzirta@gmail.com	תמי קציר
מושב ג' הענקת פרס אפי והרצאה מוזמנת	MeiravA@taldor.co.il	יור - מירב אריאלי-אטאלי
	noag@nite.org.il	נועה גזית
	beata.beigman.klebanov@gmail.com	ביאטה בייגמן קלבנוב
מושב ד' מדידת תפקודים קוגניטיביים וביצועיים	anat_ben_simon@hotmail.com	יור - ענת בן-סימון
	yaara@nite.org.il	יערה לדור ויצמן
	Doron.sagi@sheba.health.gov.il	דורון שגיא
	judith.liron@gmail.com lironju@biu.ac.il	יהודית לירון
מושב ד' סוגיות ואתגרים פסיכומטריים	avialef@gmail.com	יור - אבי אללוף
	olgaentel24@gmail.com	אולגה אנטל קטייב
	shaul@nite.org.il	שאול שלפק
	ltamar.gati@huji.ac.il	איתמר גתי
מושב נעילה אורינות מידע בעידן ה-AI - מוטיבציות ואיכות	giligold@mta.ac.il	יור - גילי גולדצויג
	Tamisa@education.gov.il	תמי סבג שושן
	shiretgar@tauex.tau.ac.il	שיר אתגר







תגובות והצעות

בנושא האגודה הישראלית לפסיכומטריקה

פנו אלינו במייל | info@ispa.org.il

אתר אפי | www.ispa.org.il

עיצוב ועריכה גרפית | יעל פרץ | מאל"ו