

הכינוס ה-16 של האגודה הישראלית לפסיכומטריקה

ט"ז בשבט תש"פ 11.2.2020



כפר המכביה, רמת גן



בתמיכת:



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

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

מטרות האגודה:

- לעודד את הפעילות בתחום הפיתוח והיישום של מבחנים סטנדרטיים ושיטות להערכת ביצוע
- לקדם את המחקר והעשייה בתחומי המדידה החינוכית והפסיכולוגית
- לשכלל את שיטות ההערכה וכלי המדידה ולבסס את יישומם המושכל
- להתוות, להעריך ולהפיץ כלים מתודולוגיים עדכניים ושיטות הערכה חדשות
- להתריע על שימוש לקוי במבחנים ועל פרשנות מוטעית של תוצאותיהם
- לעודד קשר מקצועי, חילופי מידע ושיתוף פעולה בין העוסקים בתחום המדידה וההערכה

תכנית

התכנסות, רישום וארוחת בוקר	8:30 – 9:15	
דברי פתיחה וברכות יו"ר אפי: צור קרליץ יו"ר הכינוס: אימאן עואדיה, עינת נוטע-קורן	9:15 – 10:00	
<p>הענקת פרס אפי והצגת העבודה הזוכה יו"ר ועדת הפרס: נתנאל סילברשטיין</p> <ul style="list-style-type: none"> ■ פיתוח מיומנות לחשיבה אסטרטגית במתמטיקה בגישות ההוראה המפורשת והבלתי מפורשת בקרב תלמידים מוכשרים מהמועדון המתמטי "קידומטיקה" עבודת דוקטורט של ילנה פורטנוב נאמן ■ האם ניתן למדוד החלמה מהפרעות אכילה? תיקוף שאלון החלמה מהפרעות אכילה ככלי קליני ■ עבודת מאסטר של מיכל לינצקי 	ריומן מרכז	
מושב א' מושב פתיחה יו"ר: אימאן עואדיה	10:00 – 11:15	
<ul style="list-style-type: none"> ■ על מודלים של ניבוי במוסדות להשכלה גבוהה – הצגה של רעיון בפיתוח ניצן וינטראוב, עינת נוטע-קורן ■ שילוב מאפייני משחק ואינטראקטיביות בכלי הערכה: דוגמא ממחקר פיתוח מבחן מדעים לחטיבת ביניים ■ מירב אריאלי-אטאלי ■ זיהוי קשרים בין שאלות בסביבת למידה ממוחשבת באמצעות כריית נתונים טניה נזרצקי, שרה הרשקוביץ, גיורא אלכסנדרון 	ריומן מרכז	
הפסקת קפה	11:15 – 11:35	
מושב ב' 1 (מושב מקביל) שיטות סטטיסטיות יו"ר: יוסי מחלוף	11:35 – 12:50	
<ul style="list-style-type: none"> ■ שקילות רגרסיה לוגיסטית ■ גיא עשירי-פרוסנר ■ הליך גיבוש ובחירת השיטה לחישוב ציוני הנורמות במבחני וודקוק ג'ונסון בישראל ■ דביר קלפר, אימאן עואדיה, אסטלה מלמד, חגית גליקמן, יואל רפ, נעמי גפני ■ בחינת הרגישות של אומדני פרמטרים של IRT לגודל המדגמים של תגובות הנבחנים ■ מיכל באומר, יואל רפ 	ריומן מזרח	

<p style="text-align: center;">מושב ב' 2 (מושב מקביל) מבחנים ממוחשבים יו"ר: צור קרליץ</p> <ul style="list-style-type: none"> ■ אתגרים בפיתוח מבחן ממוחשב בשפה דבורה באנגלית וממצאים פסיכומטריים ראשונים מהפיילוט איה להב, ענבל רון-קפלן, רון ברצלבסקי, נורית ליפשטט ■ מבחן אדפטיבי בשלבים: סוגיות בפיתוח ובניתוח פריטים – המקרה של מבחן עברית לדוברי ערבית בעז צור, הדיל קורדחגי, רוני ברצלבסקי, נורית ליפשטט, ענבל רון-קפלן, צור קרליץ, מירב אריאלי-אטאלי ■ מה ניתן ללמוד בניתוח אוטומטי על תלמידים העונים על שאלות רב-ברירה מדוגמאות תומכות המצורפות לתשובותיהם? שי אולשר 	רימן מרכז	
ארוחת צהריים		14:00 – 12:50
<p style="text-align: center;">מושב ג' 1 (מושב מקביל) עיבוד נתונים ומחקר פסיכומטרי יו"ר: נועם פישמן</p> <ul style="list-style-type: none"> ■ פערי מגדר בבחינת הכניסה הפסיכומטרית לאוניברסיטאות יואל רפ, כרמל אורן, דוד מעגן ■ מגמות בציוני הבגרות על פני זמן כרמל אורן, דוד מעגן, יואל רפ ■ סינון ראשוני בבחירת תחום לימודים ומקצוע: מה מאפיין רשימת מקצועות מבטיחים טובה? טוני גוטנטיג, אביבה שמעוני, איתמר גתי 	רימן מרכז	15:15 – 14:00
<p style="text-align: center;">מושב ג' 2 (מושב מקביל) מחקרים בנושאים שונים יו"ר: אבי אללוף</p> <ul style="list-style-type: none"> ■ הבסיס הפסיכומטרי לחוסר היעילות של מדדי יעילות אייל גמליאל, איל פאר ■ כשפחות שווה יותר: הבנה אינטואיטיבית של משתנה מדכא בתהליכי מיון חגי רבינוביץ, יואלה ברבי-מאיר, דוד בודסקו ■ ניתוח רשתות: קווים מנחים ליישומי מחקר ענת רוטשטיין 	רימן מזרח	16:00 – 15:15
אסיפת חברי אפי		

עבודה דוקטורט זוכת פרס אפי 2020

פיתוח מיומנות לחשיבה אסטרטגית במתמטיקה בגישות ההוראה המפורשת והבלתי מפורשת בקרב תלמידים מוכשרים מהמועדון המתמטי "קידומטיקה"

ילנה פורטנוב נאמן

אוניברסיטת בן גוריון בנגב, קיי - המכללה האקדמית לחינוך

חשיפת ילדים לאסטרטגיות בפתרון בעיות מאפשרת להם להתמחות בהן, להתאימן לסיטואציות שונות ולפעול ביעילות מרבית. בכדי להטמיע אסטרטגיות, יש להיעזר בגישות הוראה שיכולות לספק פידבק לאורך הלימוד ולעקוב אחר התקדמות הלימודית של התלמידים. ללימוד אסטרטגיות קיימת גם השפעה חיובית על המיומנות לחשיבה אסטרטגית. בקרב עובדים ומנהלים נמצא כי שימוש באסטרטגיות מעלה רמת המיומנות לחשיבה אסטרטגית. מטרת עבודת המחקר הייתה לבחון כיצד יש ללמד אסטרטגיות בפתרון בעיות וכיצד ניתן למדוד את רמת המיומנות לחשיבה אסטרטגית בקרב תלמידים.

שתי גישות הוראה נבחנו במחקר: הוראה מפורשת (Explicit Teaching) והוראה בלתי מפורשת (Implicit Teaching). הגישות יושמו על תוכנית לימודים שכללה אסטרטגיות מתמטיות לפתרון בעיות: פתרון מהסוף להתחלה, רקורסיה, הוכחה על דרך השלילה, ניסוי וטעייה. התכנית כללה מאגרי בעיות, מבדקים ומחוונים שהעריכו את יכולת התלמידים לשלוט באסטרטגיות לאורך תוכנית הלימוד. בכדי להעריך כיצד תלמידים פותרים בעיות פותח מודל בשם H.I.B.E.L. המודל מורכב מראשי תיבות של השלבים שנועדו לסייע בעת פתרון בעיות מתמטיות: חשיבה מערכתית הוליסטית – 'Holistic System Thinking', 'זיהוי שלבי הבעיה' – 'Identification', 'בניית תכנית פעולה' – 'Building a Plan', 'ביצוע תכנית הפעולה' – 'Execution', 'בחנית הפתרון' – 'Looking Back'. לאור שלבי מודל ובשילוב הגדרות שונות לחשיבה אסטרטגית מתחומי מנהל עסקים, תורת המשחקים, והקשרים צבאיים, פותחה הגדרה אופרטיבית למיומנות של חשיבה אסטרטגית. בהתאם להגדרה נבנה כלי גנרי שמוודד את רמת המיומנות לחשיבה אסטרטגית לפי מרכיבים מטה-קוגניטיביים של תכנון, בקרה תהליכית והערכה ובדיקה.

אוכלוסיית המחקר מנתה 234 תלמידים מוכשרים במתמטיקה מכיתות ו-ז (גילי 12-13) אשר השתתפו בתוכנית העשרה במתמטיקה 'קידומטיקה'. מתוכם נבדקה רמת המיומנות לחשיבה אסטרטגית בקרב 70 תלמידים בסוף תכנית הלימוד. שלושה ממצאים עיקריים נמצאו במחקר זה:

1. הגישה המפורשת הובילה ליכולות תכנון ובקרה תהליכית טובות יותר מאשר גישה הבלתי מפורשת.
2. נמצא שתלמידים אינם משקיעים מספיק זמן בבדיקת התשובה ובהערכתה.
3. תלמידים משתי הקבוצות, הראו רמה גבוהה של מיומנות לחשיבה אסטרטגית בפתרון בעיות.

המודל וכלי המדידה שפותחו לבחינת מיומנות החשיבה האסטרטגית הצליחו לאפיין את חשיבת התלמידים בעת פתרון בעיות. באמצעותם ניתן היה למדוד את רמת המיומנויות של תכנון הבעיה, בקרה על תהליך ובדיקת הפתרון. נראה כי מיומנויות אלו יכולות להתפתח בעזרת תרגול, עידוד והדרכה. למודל ולכלי המדידה קיימת תרומה חשובה למורים, שכן הם מאפשרים לקבל תמונת מצב מקיפה על יכולותיו של כל תלמיד בכיתה ובעזרתם ניתן לסייע ולקדם את למידת התלמיד בכל תחום במתמטיקה.

עבודה מאסטר זוכת פרס אפי 2020

האם ניתן למדוד החלמה מהפרעות אכילה? תיקוף שאלון החלמה מהפרעות אכילה ככלי קליני

מיכל לנצקי
המרכז האקדמי רופין

רקע: הפרעות אכילה הן בין ההפרעות העיקשות והממושכות ביותר מבין ההפרעות הנפשיות ושיעורי התמותה שלהן הם מהגבוהים ביותר. חוקרים שונים בתחום מגדירים החלמה מהפרעות אכילה באופן שונה ומשתמשים בהגדרות שונות ומדדים שונים, דבר אשר מוביל בתורו למסקנות סותרות וחסרות עקביות בתחום. מחקר זה מציע את שאלון ההחלמה מהפרעות אכילה (Eating Disorders Recovery Questionnaire; EDRQ) כמדד קליני מקיף, מבוסס דיווח עצמי, להחלמה מהפרעות אכילה.

שיטה: במחקר השתתפו 866 נשים, 261 עם הפרעות אכילה בהווה (93 אנורקסיה נרבוזה, 46 בולימיה נרבוזה, 42 הפרעת אכילה בולמוסית, 80 אחר), 171 שסבלו מהפרעות אכילה בעבר (79 אנורקסיה נרבוזה, 39 בולימיה נרבוזה, 15 הפרעת אכילה בולמוסית, 62 אחר) ו-434 בקבוצת הביקורת ללא היסטוריה של הפרעות אכילה. הנבדקות מלאו את שאלון ה-EDRQ בעל 28 פריטים וכן שאלונים נוספים הקשורים להפרעות אכילה: תסמיני הפרעות אכילה (EDEQ), חוויה חיובית מהגוף (DKB-35), אפקט חיובי ושילי (PANAS), תסמיני דיכאון (PHQ-9), שביעות רצון מהחיים (SWLS) ואכילה חיובית (PES).

ממצאים: נמצא לשאלון ה-EDRQ מהימנות גבוהה (קרונברך $\alpha = 0.91$). בנוסף, השאלון נמצא במתאם בינוני עד גבוה עם שאלוני המחקר האחרים. זוהו בעזרת ניתוח גורמים מגשש ארבעת תתי הסולם: מאפיינים רגשיים וחברתיים, בריאות פיזיולוגית, קבלת העצמי והגוף והתנהגות סימפטומטית. כמו כן, ציוני ה-EDEQ, PHQ-9 והאפקט השילי (PANAS) היו הגבוהים ביותר בקבוצת המתמודדות עם הפרעות אכילה בהווה והנמוכים ביותר בקבוצת הביקורת. לקבוצת הנבדקות אשר התמודדו עם הפרעות אכילה בעבר היו הציונים באמצע הטווח בין שתי הקבוצות. ציוני ה-EDRQ, ה-SWLS, ה-DKB-35, ה-PES והאפקט החיובי (PANAS) היו בדפוס של כיווניות הפוכה. בנוסף, נשים עם היסטוריה של הפרעות אכילה אשר קיבלו ציון גבוה ביותר ב-EDRQ (N=30, 18%) הראו פרופיל פסיכולוגי בריא יותר באופן משמעותי מנשים ללא היסטוריה של הפרעות אכילה.

מסקנות: התוצאות תומכות בהיותו של שאלון ה-EDRQ כלי קליני תקף למדידת החלמה מהפרעות האכילה השונות מבחינת בריאות פיזיולוגית, התנהגות סימפטומטית, קבלת העצמי ומאפיינים רגשיים-חברתיים.

על מודלים של ניבוי במוסדות להשכלה הגבוהה - הצגה של רעיון בפיתוח

ניצן וינטראוב, עינת נוטע קורן
המחלקה להערכה, האוניברסיטה הפתוחה

בשנים האחרונות נצפית ירידה בשיעור המסיימים את הלימודים לתואר ראשון בארצות הברית. מנתונים משנת 2017 עולה שרק 60% מהסטודנטים שנכנסים לאוניברסיטאות הציבוריות רוכשים בסופו של דבר את התואר. בשל ההשלכות הכלכליות והחברתיות שעלולות לנבוע מהתופעה, קיים לחץ גדול על המוסדות להשכלה גבוהה בארצות הברית לעשות מאמץ ולפעול להקטנת תופעת הנשירה.

כדי לחתור לפעילות אפקטיבית וככלי להקצאת משאבים, יותר ויותר מוסדות אקדמיים החלו לאתר סטודנטים בסיכון לנשירה באמצעות מודלים מבוססי למידת מכונה (Machine Learning). לקבוצת/קבוצות היעד שאותרה/אותרו מבצעים התערבויות מתאימות. בסקרים מ-2015 עלה כי כ-40% מהקולג'ים בארה"ב עושים שימוש במודלים של ניבוי לצרכים של קבלת החלטות.

האוניברסיטה הפתוחה בישראל, אף היא, מתאפיינת בשיעורי נשירה גבוהים יחסית, בשל המבנה הייחודי שלה והקבלה הפתוחה. אי לכך, היא משקיעה משאבים רבים שמטרתם לסייע לסטודנטים להתמיד ולסיים את התואר. היא עושה זאת, למשל, באמצעות העמדת מערך ייעוץ ייחודי ורחב היקף שמלווה את הסטודנטים בדרכם האקדמית ובאמצעות יחידה שפונה לסטודנטים באופן ייזום בצמתים, שהוגדרו מראש, ומציעה להם שירותים.

מאפיינים אלו של האוניברסיטה הפתוחה: השאיפה לשפר את ההתמדה באמצעות העמדת מנגנוני תמיכה מפותחים ולצידם מערך טכנולוגי ומחקרי חזק, הופכים אותה למוסד שיכול להתאים לשימוש במודל חיזוי לצורך התערבויות מותאמות. נוסף על כך, עצם זה שהאוניברסיטה שוקדת זה שנים על בניית מאגרי נתונים גדולים ומסודרים, הופך אותה במובנים רבים לפרי בשל למהלך כזה.

כדי לסייע לאתר סטודנטים ברמות שונות של סיכון לנשירה, פיתחנו מודלים שונים על בסיס נתונים (חלקיים) הקיימים על הסטודנטים. מודלים אלו הם ניסיוניים וטרם נעשה בהם יישום כלשהו. בהרצאה נציג מודל לניבוי הצלחה בקורס שאליו רשום הסטודנט, מתוך נתוני העבר שלו. המודל מתאים רק לסטודנטים שניגשו לפחות למבחן אחד באוניברסיטה (כלומר, לא לסטודנטים חדשים) ומלמד על הסיכוי להצליח בקורס הנוכחי כבר בתחילת הסמסטר, עוד בטרם נלמד הקורס.

לשם כך הגדרנו 5 קבוצות של משתנים: משתנים דמוגרפים, משתני הצלחה, קצב לימוד, משתנים לימודיים נוספים (סיווג אנגלית, בגרות, תחום לימודים, קבלת ייעוץ, קבלת מלגות) ומדדי הקורסים שנלמדו בשנים הקודמות. המשתנים לא כללו עקבות שהשאיר הסטודנט בתוך אתרי הלימוד של הקורסים. לצורך בחירת המודל המתאים ביותר השתמשנו במספר משפחות של אלגוריתמים (למידת מכונה ML, ולמידה עמוקה DL) כגון:

Random forest (RF), k-Nearest Neighbors algorithm (k-NN), Artificial Neural Network (ANN).

כל המודלים נבנו תוך הפרדת הנתונים לנתוני אימון, נתוני ולידציה (תיקוף) ונתוני מבחן. כמו כן, לצורך כיול הפרמטרים ובחירת המודל הטוב ביותר עשינו הליך של תיקוף צולב (Cross Validation). בחירת המודל ה"טוב ביותר" התבססה על מגוון מדדי דיוק סימולטנית.

המודל שהניב את התוצאות הטובות ביותר הוא Random Forest שלאחר כיול הניב דיוק מאוזן הגבוה מ-80%.

שימוש במודלים מסוג זה אינו חף מביקורות שעוסקות בסוגיות כבדות משקל שיש לדון בהם בטרם נעשה בהם שימוש. למשל, חשש מהטייה מובנית במודלים, חשש משימוש לרעה (abuse) או שימוש לא נכון בתוצרי המודל (misuse), העדר שקיפות, פגיעה בפרטיות וכו'. בהרצאה נדון בקצרה באתגרים ובדרכי התמודדות שהועלו בספרות.

שילוב מאפייני משחק ואינטראקטיביות בכלי הערכה: דוגמא ממחקר פיתוח מבחן מדעים לחטיבת ביניים

מירב אריאלי-אטאלי

המרכז לטכנולוגיה חינוכית (מטח)

Fordham University, New York & ACTNext by ACT, Iowa

ההתקדמות הטכנולוגית מביאה עימה אפשרויות לחדשנות בעולם המבחנים והמדידה, ובפרט מאפשרת לפתח מבחנים אינטראקטיביים ולהוסיף אלמנטים של משחק ("גמיפיקציה"). אולם, עדיין אין מספיק מחקר המראה את ההשפעה של חידושים אלו על ביצועים במבחן. הרצאה זו תציג מחקר פיתוח של כלי הערכה במדעים לחטיבת הביניים ששילב מאפייני משחק ואינטראקטיביות, ובחן את האפקט של תוספות אלו על ציוני התלמידים ועל דיווחי המעורבות והעניין שלהם (engagement).

מטלות אינטראקטיביות ומשחקיות יכולות לזמן עבור תלמידים פעילות אותנטית עם משוב כחלק אינטגרלי של הפעילות, דבר שיכול להעלות את מידת המעורבות בביצוע המטלה, וכתוצאה מכך צפוי גם להביא לביטוי טוב יותר של יכולות התלמידים. מחקרים מראים כי הוספת אינטראקטיביות למטלות למידה מעלה מעורבות בכך שהיא מחייבת את התלמידים להיות אקטיביים (ולעיתים אף קונסטרוקטיביים ויצירתיים) כדי לקדם את הפעילות (Chi, 2009). אפשר להעלות מעורבות ועניין גם על ידי הוספת מאפייני-משחק. הוספת מאפייני-משחק, או גמיפיקציה, אינה משנה את הפעילות עצמה, אלא מוסיפה רובד משחקי דרך הוספת תחרות (ניקוד), אלמנטים ויזואליים או אודיטורים, הוספת סיפור מסגרת דרך דמויות ודיאלוגים, וכדומה. ישנן עדויות מחקריות שלהוספת מאפייני משחק למערכות למידה יש פוטנציאל לעורר סקרנות ואתגר, הזדהות ודמיון, ו/או תחושת שליטה ותחרותיות, וכתוצאה מכך להעלות את המוטיבציה והעניין של התלמידים (Malone, 2000; Ryan & Deci, 2000; McNamara, Jackson & Graesser, 2010; 1981). אולם, הממצאים אינם חד משמעיים כשמדובר במערכות מבחן, שכן למאפייני המשחק יכול גם להיות אפקט מפריע/מסיח (distraction) (e.g., Attali & Arieli-Attali, 2015). במחקר זה יצאנו לבדוק מהו האפקט של הוספת מאפייני משחק שונים למטלה דמוית-מבחן, והאם יש להם אפקט נוסף בהשוואה לאפקט של הוספת אינטראקטיביות בלבד.

לצורך בדיקת שאלות המחקר, פיתחנו ארבע גירסאות של מבחן מדעים: (1) גירסת המקור (קבוצת ביקורת) - מבחן קונוונציונאלי הכולל שאלות רב-ברירה שהועלה על המחשב; (2) מבחן אינטראקטיבי - אותן שאלות שעברו התאמה להיות אינטראקטיביות עם אפשרות להתנסות; (3) מבחן אינטראקטיבי עם סיפור מסגרת - גירסה 2 בתוספת נרטיב המוצג דרך דמויות מצוירות ודיאלוג ביניהן; ו- (4) מבחן אינטראקטיבי עם נקודות - גירסה 2 בתוספת מתן משוב על תשובות, וניקוד מצטבר עם בונוס על מהירות, מלווה באלמנטים של קול ותמונות נעות.

ארבע הגירסאות הועברו ל-245 תלמידי כיתות ז' בבית ספר במרכז ארצות-הברית במסגרת שיעור מדעים. לאחר סיום היחידה, התלמידים התבקשו לענות על שאלון משוב קצר לגבי החוויה שעברו.

במחקר נבדק האפקט של האלמנטים הנוספים על ביצועי התלמידים ועל מידת העניין והמעורבות שלהם, תוך השוואת התוצאות של ארבע הגירסאות. השערות המחקר היו שתלמידים יגלו מעורבות רבה יותר ויבצעו טוב יותר בגירסא האינטראקטיבית, ואף יותר מכך בגירסאות עם מאפייני המשחק ("גירסת סיפור מסגרת", או "גירסת הנקודות"). התוצאות הראו שאכן הביצוע היה גבוה יותר בשלוש הגירסאות האינטראקטיביות בהשוואה לגירסה הקונוונציונאלית (קבוצת הביקורת), אולם לא נמצא הבדל בין הגירסאות האינטראקטיביות, בפרט, הוספת מאפייני משחק לא שיפרו את הביצוע במבחן מעבר לשיפור שנמצא בעקבות הוספת האינטראקטיביות. בהרצאה אדון בתוצאות ובהשלכותיהן להערכה מעצבת, ובפרט אדבר על כיוונים נוספים שבכוונתינו לבדוק לאור ממצאים אלו.

זיהוי קשרים בין שאלות בסביבת למידה ממוחשבת באמצעות כריית נתונים

טניה נזרצקי

מכון ויצמן למדע

שרה הרשקוביץ

המרכז לטכנולוגיה חינוכית (מטח)

גיורא אלכסנדרון

מכון ויצמן למדע

לומדות אדפטיביות (Intelligent Tutoring Systems) מבוססות בדרך כלל על אפיון של תחום הדעת כאוסף מיומנויות והקשרים ביניהן (למשל, מיומנות א' היא תנאי מקדים למיומנות ב'), ומיפוי של פריטי למידה למיומנויות שהתלמיד מיועד לרכוש. לפיכך, מיפוי זה הוא חלק הכרחי וחשוב בתכנון לומדות כאילו, הנעשה לרוב על ידי מומחי תוכן.

עם זאת, תיוג של מאגרים גדולים של פריטים על ידי מומחי תוכן דורש זמן ומשאבים רבים. בנוסף, תהליך רכישת המיומנויות בפועל על ידי הלומדים אינו תמיד עולה בקנה אחד עם מפת המיומנויות שאופיינה על ידי המומחים, כפי שמעידים מחקרים רבים. לפיכך, לשיטות אלגוריתמיות המסייעות בתהליך זיהוי מבנה הידע וסיווג הפריטים לרכיביו יש פוטנציאל גדול לייעל ולדייק את תהליך התיוג.

קיימות שתי גישות לאפיון של המיומנויות ומיפוי הפריטים אליהן מתוך נתוני למידה. הגישה הראשונה מבוססת על זיהוי מודל סמוי, והשנייה על חישוב הדמיון בין הפריטים. מחקר זה מתמקד בגישה השנייה. מחקרים קודמים הראו שניתן להשתמש במדדי דמיון סטנדרטיים כמו Cohen's Kappa, Pearson או Yule לצורך מיפוי פריטים למיומנויות מתוך נתוני למידה. עם זאת יישומים אלה מסתמכים (גם אם בצורה לא מודעת) על ההנחה שרמת השליטה של הלומדים במיומנויות קבועה. הנחה זו סבירה בהקשר של הערכה מסכמת, אך אינה מתקיימת בהקשר של הערכה מעצבת, בה שליטת הלומדים במיומנויות משתנה במהלך פעילות הלמידה.

כדי לטפל בפער זה אנו מציעים מדד דמיון חדש, המכונה Kappa Learning, אשר נועד למדוד דמיון בין פריטי הערכה תחת ההנחה שהלומדים רוכשים שליטה במיומנויות שעומדות בבסיס הפריטים בעת שהם פותרים אותם. מדד זה מבוסס על התאמת מדד Cohen's Kappa על ידי מידול מפורש של תהליך השיפור ברמת השליטה של הלומדים. הערכת המדד החדש התבצעה על נתונים אמתיים ממערכת למידה מקוונת להוראת שברים של המרכז לטכנולוגיה חינוכית, ועל סימולציה ממוחשבת של נתוני לומדים.

התוצאות מראות שביצועי השיטה המוצעת, עבור תחומי תוכן מובנים ותחת ההנחה של למידה, עולים על הביצועים של שיטות המבוססות על מדדי דמיון נפוצים סטנדרטיים. לפיכך, אנחנו מאמינים שמדד Kappa Learning מתאים יותר מהמדדים הסטנדרטיים למיפוי אוטומטי של פריטי למידה למיומנויות עבור תחומי תוכן מובנים (כגון מתמטיקה לחינוך יסודי) ובסביבות למידה מקוונות.

שקילות רגרסיה לוגיסטית

גיא עשירי-פרוסנר

האוניברסיטה העברית, מרכז ארצי לבחינות ולהערכה (מאל"ו)

כאשר אנו בונים כלי פסיכומטרי, אנו עושים זאת באמצעות שימוש במדגם מחקרי מתוך אוכלוסיה מסוימת. האם נוכל להשתמש בו גם עבור אוכלוסיה אחרת? האם כלים שנבנו על בסיס נתוני אוכלוסיות שונות הם שקולים, וניתן להשתמש רק באחד מהם? שאלה זו של שקילות בין מודלים אינה נובעת מקושי סטטיסטי או חישובי אלא מקושי מדעי-פסיכומטרי: היכולת להכליל מודל ניבוי מסוים לאוכלוסיות שונות מעלה את התוקף שלו, והצורך במודלים שונים עבור אוכלוסיות שונות מחליש את המודל התאורטי.

בעיה זו של בחינת שקילות מודלים בין אוכלוסיות הנה סוגיה יסודית במדעי החברה, וניתנים לה פתרונות שונים עבור סוגי מודלים שונים. בין הפתרונות המוכרים ניתן למנות הערכת תפקוד פריטים דיפרנציאלי (DIF) בענפים שונים של תורת המבחנים ושקילות מדידה (Measurement Invariance) בהערכת שקילות מודלים של ניתוח גורמים בין אוכלוסיות. עד כה, לא ניתן מענה מקיף לשאלת שקילות מודלים של רגרסיה לוגיסטית לניבוי, המבוססים על אוכלוסיות שונות. שאלה זו היא העומדת בבסיס מחקר זה. המניע למחקר הוא בחינת הצורך במודלים נפרדים לניבוי עבור אוכלוסיות שונות של פונים לאבחון ממוחשב של לקויות למידה במערכת מת"ל, בחלוקה על פי שפת אם או מגדר.

כאשר בוחנים הבדלים והפרשים בין אוכלוסיות, מקובל להשתמש במבחני מובהקות עם השערת האפס כי לא קיים הבדל בין האוכלוסיות. בבחינת השערות מסוג זה, מדגמים קטנים מאופיינים לרוב בשונות גבוהה ולפיכך ההסתברות שלא לדחות את השערת האפס גדולה יותר בשימוש במדגם קטן. כך נוצר תמריץ שלילי (מבחינה מדעית) לשימוש במדגמים קטנים על מנת שלא לדחות את השערת האפס. בעבודה זו אנו מציעים לבחון השערות באמצעות מבחני שקילות, בהם השערת האפס היא כי קיים הפרש בגודל מוגדר מראש בין האוכלוסיות. על ידי שימוש במבחני שקילות נוצר תמריץ להגדלת המדגם ככל הניתן.

בעבודה זו אנו מציגים שלוש שיטות לבחינת שקילות בין מודלים של רגרסיה לוגיסטית: (א) שקילות תיאורית, המושגת על ידי בחינת השקילות בין וקטורי המקדמים של מודלי הרגרסיה; (ב) שקילות ניבוי פרטנית, המושגת על ידי בחינת השקילות בין יחסי הסיכויים המופקים בשימוש במודלים השונים; (ג) שקילות ניבוי כללית, המושגת על ידי בחינת השקילות של ציוני ברייר (Brier) של המודלים השונים. לכל אחת משלוש השיטות אנו מספקים הסבר כללי, הנחיות שימוש, הוכחת יעילות מול גודל מדגם והוכחת יעילות יתרה לעומת שיטות מקובלות (באמצעות הדמיות). לסיום, אנו מדגימים מקרים של מודלים שקולים ושאינם שקולים באמצעות נתונים מחקריים ממערכת מת"ל.

הליך גיבוש ובחירת השיטה לחישוב ציוני הנורמות במבחני וודקוק ג'ונסון בישראל

דביר קלפר¹, אימאן עואדיה², אסטלה מלמד¹, חגית גליקמן², יואל רפ^{1,2}, נעמי גפני¹

¹מאל"ו - מרכז ארצי לבחינות ולהערכה

²ראמ"ה - הרשות הארצית למדידה והערכה בחינוך

סוללת מבחני וודקוק-ג'ונסון 3 לבדיקת כשרים קוגניטיביים של ילדים ומבוגרים (Woodcock, McGrew, & Mather, 2001) תורגמה והותאמה לעברית ולעברית בישראל על ידי ראמ"ה - הרשות הארצית למדידה והערכה בחינוך, בשיתוף עם גופים נוספים (תרגום והתאמה - מט"ח, המרכז לטכנולוגיה חינוכית; מחקר הנורמות - מאל"ו, המרכז הארצי לבחינות ולהערכה, העברת המבחנים במחקר הנורמות - חברת טלדור) ובליווי מומחים לפסיכומטריקה.

קיימת ספרות פסיכומטרית דלה יחסית העוסקת בשיטות לחישוב נורמות במבחני אינטליגנציה, ובמיוחד כזו הדנה בסוגיות הנוגעות בשאלה של מה הן השיטות הנכונות יותר ובאלו תנאים. כמו כן, חברות מסחריות או ארגונים אקדמיים שמפתחים מבחנים כאלו ממעטים לפרט את השיטות המדויקות שעל פיהן עבדו ועל פיהן פיתחו את הנורמות. בעוד שהם מספקים תיאור מפורט של המדגם בו השתמשו הם מסתפקים במידע מעורפל וכללי בלבד אודות שיטות החישוב שבהן בחרו.

בהרצאה זו יתואר שלב פיתוח השיטה לחישוב נורמות גילאיות עבור סוללת המבחנים בישראל, לגרסה העברית והערבית. כלומר, שיטה להמרת ציוני הגלם לציוני נורמה בכל מבחן בסוללה, ושיטה בה ימוזגו ציוני הנורמה במבחנים הבודדים לציוני נורמה של מצרפים שונים של מבחנים ובפרט ציון אינטליגנציה כללית. המחקר יציג באופן סדור את הליך גיבוש פיתוח השיטה שבה בוצעו בסופו של דבר חישובי הנורמות הגילאיות למבחני וודקוק-ג'ונסון בישראל בעברית ובערבית. השיטה מותאמת לנתונים שנאספו, ופתרה במידה משביעת רצון את הבעיות והסוגיות הקיימות בשיטות השונות האחרות שנבדקו.

ראשית יוצגו שתי מחלקות מרכזיות של שיטות לפיתוח נורמות גילאיות עבור מבחן בודד: שיטות בדידות (קרי, חישוב נורמות לקטגוריות גיל נפרדות) ושיטות רציפות (לכל גיל באשר הוא לאורך כל רצף הגילאים הרלבנטי). לצורך ההחלטה על השיטה הוגדרו עקרונות מנחים כללים אותן הנורמות צריכות לקיים, ובהתאם לכך יוצגו היתרונות והחסרונות של כול אחת משיטות אלו.

בסופו של דבר הוחלט ליישם שיטה רציפה ולצורך זה הוצעו 4 מודלים שונים של חישוב אשר כולם מתבססים על התיזה של אוסטרויס (2018, פרק 5). בכנס יוצג בהרחבה המודל הראשון שמהווה בסיס לכול שלושת האחרים, ויוסברו בקצרה השינויים שהוכנסו במודל הבסיסי לצורך התאמת שלושת המודלים האחרים. גם מודלים אלו נמצאו לא מספקים לצרכינו ובסופו של דבר פותח על ידנו מודל נוסף המבוסס על קו שבור (ספלייני).

לבסוף, יוסבר בקצרה כיצד עוברים מציונים של מבחנים בודדים לציוני מצרף ואת הבעיות והדילמות שנתקלנו בהן גם בהליך זה.

בחינת הרגישות של אומדני פרמטרים של IRT לגודל המדגמים של תגובות הנבחנים

מיכל באומר, יואל רפ
מרכז ארצי לבחינות ולהערכה (מאל"ו)

בימים אלה שוקד המרכז הארצי לבחינות על פיתוח מבחן ידע חדש באנגלית כשפה זרה. המבחן יכלול בדיקה של מיומנויות חדשות שלא נבדקו בבחינות האנגלית עד כה כמו הבנת הנשמע וכתיבה בנוסף להבנת הנקרא והוא צפוי להיות מועבר באמצעות מחשב למועמדים להשכלה גבוהה באוניברסיטאות ובמכללות ברחבי הארץ בהרבה מועדים לאורך כל השנה. המבחן יהיה מבחן אדפטיבי וידרוש פיתוח מואץ של פריטים, בדיקה שלהם והפעלת הליכי אמידת פרמטרים תכופים (לפי תאוריית התגובה לפריט – IRT – Item Response Theory) שכל אחד מהם מסתמך על מספר קטן יחסית של תגובות נבחנים.

לשם אמידת פרמטרים של IRT נדרשים בדרך כלל תגובות של נבחנים רבים. בנושא זה קיימת ספרות ענפה, אך עם המעבר להליכי אמידה עם מספר קטן יחסית של נבחנים, יש חשיבות רבה לניתוח רגישות הנועד לבחון עד כמה מושפעים הפרמטרים מגודל המדגם המשמש לאמידתם.

המטרה המרכזית של מחקר זה הייתה לבדוק את השפעת מספר התגובות המשמשות לאמידת פרמטרים של פריטי פרק האנגלית בבחינה הפסיכומטרית על הדיוק של הפרמטרים של הפריטים ולמצוא מהו מספר תגובות הנבחנים המינימלי שאפשר ורצוי לבסס עליו את הליך אמידת הפרמטרים, בתנאי ההבחנות הצפויים ובמגבלות הצפויות.

לצורך המחקר השתמשנו בנתוני התגובות של נבחנים בפרקי האנגלית של שני נוסחים של המכפ"ל (מבחן הכניסה הפסיכומטרי לאוניברסיטאות) שהועברו לנבחנים דוברי עברית ודוברי ערבית ב-2016. הנתונים כללו תגובות ל-132 פריטים שהועברו יחדיו לכ-17000 נבחנים.

בעזרת נתונים אלה נערכו מספר הליכי אמידה של הפרמטרים במודל IRT הלוגיסטי התלת-פרמטרי הכולל שלושה פרמטרים בעבור כל פריט: פרמטר האבחנה (a), הקושי (b) והניחוש (c). ראשית נלקחו כל הנתונים שעמדו לרשותנו ונאמדו הפרמטרים כאשר אלו שימשו בהמשך כפרמטרים "האמיתיים", אליהם הושוו אומדני הפרמטרים בשלבים הבאים. לאחר מכן נלקחו מדגמים מקריים מתוך כלל הנבחנים, בגדלים הולכים וקטנים בין 1000 ל-50. כאמור, את היתכנות תוצאות האמידה בדקנו בכל המקרים על פי המתאמים בין סדרות אומדני הפרמטרים בכל אחד מהליכי האמידה לבין האומדנים שהתקבלו בהליך האמידה שהתבסס על כלל הנבחנים.

ככל וכצפוי, ממצאי המחקר הראו כי ככל שמספר הנבחנים שעליהם מבססים את הליכי האמידה קטן, כך ההתאמה של האומדנים עם האומדנים שהתקבלו בהליך האמידה הראשי, שהתבסס על כלל הנבחנים, קטנה. בפרט קטנה ההתאמה בעבור אומדני פרמטר האבחנה (a) ועוד יותר מכך בפרמטר הניחוש (c). לעומת זאת אומדני פרמטר הקושי (b) שמרו על תוצאות דומות למדי בכל גדלי המדגמים. בנוסף לכך, גם אומדני ציוני הנבחנים (ציוני התטה) לא נמצאו רגישים לירידה בגודל המדגם.

אתגרים בפיתוח מבחן ממוחשב בשפה דבורה באנגלית וממצאים פסיכומטריים ראשונים מהפיילוט

איה להב, ענבל רון-קפלן, רון ברצלבסקי, נורית ליפשטט
הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

השפה הדבורה היא אופנות שפתית המשמשת להבעה עצמית, להעברת מידע וליצירת אינטראקציה בין-אישית. במערכת החינוך בישראל נהוגה מדידה סטנדרטית של מיומנות שפה דבורה באנגלית בסיום התיכון. בשנים האחרונות משרד החינוך שם דגש על מיומנות הדיבור לאורך כל שנות הלימוד ומקצה משאבים לתוכניות ייעודיות לעידוד השפה הדבורה ולכן התעורר הצורך בעריכת מדידה ארצית סטנדרטית כבר בחטיבת הביניים.

השפה הדבורה מאופיינת בכך שהתוצר שלה מועבר בזמן אמתי ומתפוגג מיד לאחר הפקתו. לכן הדרך האופטימלית לקיים מדידה מהימנה וסטנדרטית בהיקף רחב היא באמצעות הערכה ממוחשבת שבה התלמיד מקליט את תשובתו. אומנם יש יתרונות למבחן ממוחשב – סטנדרטיות, נוחות שימוש, מגוון מטלות אפשריות, אך יש לו גם מגבלות שמרחיקות אותו מהמצב הטבעי. הדיבור מול מחשב לעיתים מביך ומרגיש לא טבעי ובנוסף הוא מוגבל בממד האינטראקציה, שהוא חלק מהותי מהשפה דבורה.

בשנת תשע"ח החלו בראמ"ה בפיתוחו של מבחן ממוחשב בשפה דבורה באנגלית המיועד לתלמידי כיתות ח ו-ט. הפיתוח כלל אתגר כפול: התמודדות עם האתגרים של פיתוח מבחן בשפה דבורה, בפעם הראשונה בראמ"ה, והתמודדות עם האתגרים הנלווים לפיתוח של מבחן ממוחשב.

פיתוח המבחן הובל על ידי הצורך ליצור מדידה סטנדרטית של שפה דבורה, תוך רתימת היכולות הייחודיות של המערכת הממוחשבת ליצירת חוויה שמוודדת שפה דבורה בצורה טבעית יחסית ותקפה. המבחן נבנה בפורמט של שיחה שבה דמות ממוחשבת פונה אל התלמיד ושואלת אותו שאלות, במטרה לזמן דיבור. הפורמט הממוחשב כולל גם מידע קולי וויזואלי – סרטונים ותמונות – על מנת ליצור תחושה של שיחה טבעית שיש בה גם ממד ויזואלי – תנועות גוף והבעות פנים.

בניית המבחן לוותה בהתלבטויות לגבי הדרך הנכונה להעריך שפה דבורה. הערכה של תשובה דבורה דומה במידה רבה להערכה של תשובה כתובה, אולם לשפה דבורה נוסף גם רובד אינטראקטיבי-חברתי. הרובד הזה העלה התלבטויות הייחודיות לשפה הדבורה. למשל, כיצד מתייחסים למצבים בהם הנבחן מראה הבנה של השאלה אך מתחמק מלענות עליה – האם היכולת לשנות נושא מעידה על היכולת ליישם שפה דבורה בצורה פונקציונאלית? מה עושים כאשר הנבחן לא הבין נכון את השאלה – האם היכולת להבין את בן השיח היא חלק ממהות האינטראקציה המאפיינת שפה דבורה?

תגובות התלמידים נבדקו על ידי שמונה מורים – מספר שווה של דוברי עברית וערבית, כ-50% מהתגובות נבדקו פעמים על ידי שני מעריכים שונים לצורך בדיקת מהימנות. מערך הפיילוט והבדיקה הכפולה איפשר לענות על שאלות לגבי איכות המבחן והמבחן לפני היציאה למחקר רחב היקף. מערך הבדיקה הכפולה איפשר גם לבחון שאלה ייחודיות לבדיקה של שפה דבורה: האם קיימת השפעה של שפת האם של הנבדק והמעריך – האם למבטא של הדובר (מאפיין שאינו מוערך במבחן) יש השפעה על מידת המובנות שלו וכפועל יוצא גם על הציון?

פיתוח המבחן הממוחשב בשפה דבורה לווה בלא מעט התלבטויות ואתגרים. הפיתוח לווה בהתייעצות עם מומחים מתחום השפה וסקירה של פריטים ומבחנים ממבחינים ממוחשבים של שפה דבורה הקיימים בעולם. תוצאות מבחן הפיילוט מראות שהמבחן שפותח מתפקד כראוי מבחינה פסיכומטרית ויכול להוות כלי מדידה שישמש ליצירת נורמות ארציות ואיתור מוקדי קושי בעתיד.

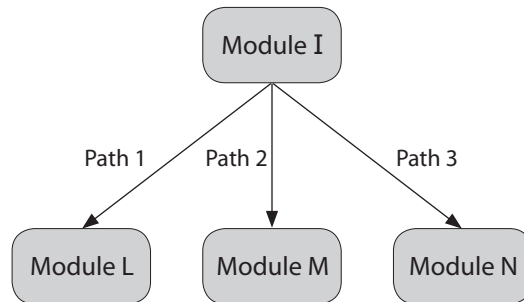
מבחן אדפטיבי בשלבים: סוגיות בפיתוח ובניתוח פריטים - המקרה של מבחן עברית לדוברי ערבית

ענבל רון-קפלן, רוני ברצלבסקי, נורית ליפשוט
הרשות הארצית למדידה והערכה בחינוך (ראמ"ה)

מירב אריאלי-אטאלי, בעז צור, הדיל קורדחגי
המרכז לטכנולוגיה חינוכית (מטח)

צור קרליץ, שאול שלפק, נעם כאהן, איתמר כהן
מרכז ארצי לבחינות ולהערכה (מאל"ו)

בשנים האחרונות עולה הצורך וההכרה בפיתוח מבחנים אדפטיביים בשלבים (Yan, von Davier & Lewis, 2016). בעיקר במסגרת הערכה מעצבת, אך יותר ויותר גם בהערכה מסכמת. למבחן אדפטיבי יתרונות רבים הן מנקודת המבט של הנבחנים והן מנקודת המבט של מפתחי המבחנים. מבחינת הנבחנים הדבר בא לידי ביטוי בחוויית ההיבחנות - נבחנים מקבלים פריטים ברמת קושי המתאימה להם, דבר שיכול לחסוך תסכול מנבחנים מתקשים, ושעמום מנבחנים חזקים (Wainer, 2000). מבחינת מפתחי המבחנים היתרון הוא ביעילות המבחן - ניתן לקצר את אורך המבחנים בצורה משמעותית שכן רמת האינפורמציה שמתקבלת מכל פריט גבוהה יותר, וכן לקבל אינפורמציה על כל הטווח הרחב של הפריטים מבלי שכל הנבחנים יקבלו את כל הפריטים. תרשים 1 מדגים מבנה של מבחן אדפטיבי דו-שלבי, שבו לאחר השלב הראשון מנותבים הנבחנים ליחידת המשך על פי קריטריון מסוים. דוגמא לניתוב על פי קושי היא כאשר הנבחנים עונים לחלק הראשון שהוא ברמת קושי בינונית, ולפי הביצוע בחלק הראשון הנבחנים מנותבים לחלק שני - קל, בינוני או קשה.



תרשים 1: דוגמא למבנה של מבחן אדפטיבי דו-שלבי
(מתוך Yan, von Davier & Lewis, 2016, עמ' 5)

שימוש במערך אדפטיבי, שבו לכל אחד מחלקי המבחן מוקצה אוכלוסייה שונה (על בסיס הביצוע בשלב הראשון), מקשה על ניתוח המבחן באמצעות תורת המבחנים הקלאסית. ניתוח כזה עלול להטעות ולתת תמונה מעוותת של רמת הקושי ומידת ההבחנה של הפריטים, וכמובן אי אפשר להשוות את הציון הממוצע בכל חלק כדי לדעת את היחס בין רמות הקושי של החלקים השונים. ניתוח התוצאות באמצעות תיאורית התגובה לפריט (IRT) אמור להתגבר על בעיה זו ולהפיק מדדים פסיכומטריים על סולם מדידה אחיד לכל חלקי המבחן (IRT; Lord, 1960). כפי שנראה במחקר זה, פתרון זה הוא חלקי ואינו מספיק לקבלת התמונה המלאה.

בהרצאה נדון בסוגיה זו דרך המקרה הפרטי של פיתוח מבחן ממוחשב בעברית לדוברי ערבית לכיתות ט. פרק הבנת הנקרא במבחן כלל חלק ראשון ממיין, שבעקבותיו נותבו התלמידים לחלק קל, בינוני או קשה, במערך משולב אדפטיבי ורנדומאלי, כלומר הניתוב לחלק השני נקבע על-פי הביצוע בחלק הממיין בשילוב הסתברות שנקבעה מראש. למרות המרכיב הרנדומאלי, החלק הקל של המבחן הועבר ברובו לתלמידים מתקשים. אנו נציג את תוצאות המחקר ונדון בסוגיות העולות ממערך זה - עד כמה אפשר להסתמך על ממצאי ניתוח הפריטים הקלאסי, מה "התיקון" שמאפשרים ממצאי ניתוח ה-IRT, ומה נשאר בעייתי גם אחרי תיקון זה.

מה ניתן ללמוד בניתוח אוטומטי על תלמידים העונים על שאלות בחירה מרובה מדוגמאות תומכות המצורפות לתשובותיהם?

שי אולשר
אוניברסיטת חיפה

שאלות בחירה מרובה הן הבחירה הטבעית עבור מערכות להערכה אוטומטית מבוססת מחשב. באופן אידאלי, בחירה בתשובה צריכה להיות מבוססת על ידע והנמקה. למרות זאת, מחקרים מראים כי לעיתים קרובות טכניקות שונות (למשל ניחוש) הן דרכי פעולה נפוצות. עם התפתחות הטכנולוגיה, נפתחות הזדמנויות לעיצוב משימות אשר מאפשרות לתלמידים ליצור מספר דוגמאות בעזרת גיאומטריה דינאמית וסביבות אחרות של כלים טכנולוגיים. מחקרים מציעים ניתוח מתמטי ומאפיינים דידקטיים בהצגת דוגמאות של תלמידים, דבר המאפשר לסייע למורים בהערכה מעצבת ובקבלת החלטות בזמן אמת בכיתה המבוססות על נתוני התלמידים. אחת מהסביבות האלו היא המרא"ה (הערכה מעצבת – רואים את התמונה), סביבת הערכה מעצבת ממוחשבת במתמטיקה אשר מפותחת באוניברסיטת חיפה. במסגרת עבודת התלמידים בסביבה זאת הם משתמשים ביישומונים מקוונים אשר מאפשרים מניפולציות שונות על אובייקטים מתמטיים באמצעות הזנת ועדכון נתונים, בנייה, גרירה, ועוד. במסגרת המשימות התלמידים חוקרים את מרחב הדוגמאות האפשרי במסגרת הדיאגרמה, ויוצרים דוגמאות אשר מדגימות דרישות אשר מצוינות במסגרת המשימה. המחקר הנוכחי בדק האם וכיצד דוגמאות הנוצרות בידי הלומדים, כאשר אלו נדרשות מהן על מנת לתמוך בבחירה שלהם בתשובה לשאלות בחירה מרובה, ניתנות לניתוח אוטומטי ומספקות תובנות לגבי הבנת הלומדים מעבר לבחירה עצמה בלבד.

המחקר בוצע במסגרת תכנית הטמעה של סביבת המרא"ה בבתי ספר, וכחלק ממחקר ארוך-טווח אשר מטרתו לבחון דרכים בהן הערכה אוטומטית יכולה לספק תובנות על הבנת תלמידים. המשתתפים הם תלמידי כיתות יא' משני בתי ספר בישראל. והמשימות במחקר זה עוצבו כך שהמערכת יכולה לציין האם הדוגמה תואמת את התנאים המתמטיים המצוינים בטענה שנבחרה. יחידת הניתוח היא כל טענה והדוגמה המצורפת אליה. בשלב הראשון אותרו אי-התאמות בין הטענה הנבחרת לבין הדוגמה שצורפה, כדי לקבוע האם הדוגמאות יכולות להוסיף לנו עוד מידע לגבי הנבדק מעבר לבחירה שלו. בשלב הבא קודדו אי התאמות בהתאם לקטגוריות שהוכנו מראש כדי ללמוד על המאפיינים השונים שניתנים להערכה אוטומטית וכוללים מאפיינים הקשורים לנושא הלימוד כמו גם מאפיינים כלליים של מיומנויות מתמטיות.

תוצאות המחקר מצביעות על כך שדרישה להדגים נכונות של טענה אכן מספקת יותר מידע לגבי עבודת התלמידים, וכך מאפשרת להבחין בצורה יותר טובה בין עבודות של תלמידים אשר בחרו את אותן הטענות כנכונות. ניתוח אוטומטי של דוגמאות אלו אפשרי, ומספק מידע מעמיק בלא צורך בהערכה ידנית נוספת. יחד עם זאת, חלק מהנתונים אשר מוגשים על ידי התלמידים (למשל סימונים ביד חופשית) יכולים לספק נתונים משמעותיים נוספים אך עדיין מצריכים ניתוח ידני של מורה.

הדרישה לספק דוגמה תומכת לטענה נבחרת מוסיפה עוד שכבת הנמקה של הנבדק, והיכולת לנתח אותה באופן אוטומטי מאפשרת העמקה של הערכת העבודה בצורה יעילה. החידוש העיקרי של מחקר זה הוא בכך שהדוגמאות ניתנות, כמו שאלות הבחירה המרובה, לניתוח באופן אוטומטי, ואינן מצריכות בדיקה ידנית כדי לקבל מאפיינים אשר לרוב אינם זמינים בצורת הבחנות כזאת.

פערי מגדר בבחינת הכניסה הפסיכומטרית לאוניברסיטאות

יואל רפ, כרמל אורן

מרכז ארצי לבחינות ולהערכה (מאל"ו)

דוד מעגן

הלשכה המרכזית לסטטיסטיקה (למ"ס)

פערי ההישגים בין המגדרים במבחנים סטנדרטיים ועתירי סיכון מושכים עניין מחקרי רב בהקשר של הוגנות ושוויון חברתי. לאורך שנות קיומה של הבחינה הפסיכומטרית נמצא כי הציון הממוצע של הגברים הנבחנים בה גבוה מזה של הנשים בכ-40 נקודות בשנה נתונה. פער זה מתפרש בציבור כחוסר הוגנות הבחינה כלפי נשים. מנגד, שעור הנשים הנבחנות בבחינה גדול משעור הגברים בערך פי 1.4. כמו כן, גדול יותר שיעור הנשים לעומת הגברים בקרב הלומדים בהשכלה הגבוהה בתארים הראשון והשני.

מחקר זה עוסק בפערים הבין-מגדריים בבחינה הפסיכומטרית ומנסה להבין את מקורותיהם. השערת המחקר היא כי שיעורם הנמוך יחסית של גברים שנבחנים בפסיכומטרי אינו מקרי אלא נובע מהליכי סלקציה עצמית וחברתית שונים המתקיימים בקרב שני המינים. לפי השערת המחקר, גברים בעלי יכולת נמוכה נוטים פחות מנשים בעלות יכולת נמוכה לגשת לבחינה הפסיכומטרית. הבדלים אפרויריים אלו בין המינים ברמת היכולת של הנבחנים בפועל, מביאים לידי כך שבממוצע, הגברים מצליחים יותר בבחינה מאשר נשים, מה שעשוי להסביר חלק ניכר מהפערים המתקבלים בבחינה הפסיכומטרית.

מנתוני הלמ"ס עולה כי ציוני הפסיכומטרי מתואמים עם הישגיהם הלימודיים של תלמידים בבית הספר (ציוני המיצ"ב בכיתה ח'). בעזרת ציוני המיצ"ב, ובצירוף מדד לרקע החברתי-כלכלי שלהם, ניבאנו לכל נבחני המיצב משני שנתונים (2008 ו-2009) את הציון בבחינה הפסיכומטרית, ובכך השלמנו נתונים חסרים לאלה מהם שלא ייגשו לימים לבחינה הפסיכומטרית. היתרון בשימוש בנתוני המיצ"ב טמון בהקיפם (כמעט) את כל בני מחזור נתון (להבדיל למשל, מנתוני הבגרות שאף בהם טמונה סלקציה). השווינו את הפערים המגדריים במיצ"ב של כלל הנבחנים בו ב-2008 ו-2009 לפערים המקבילים בקרב אלו שניגשו לבחינה הפסיכומטרית. לבסוף, אמדנו את הפער הבין-מגדרי בפסיכומטרי אילו הגברים והנשים הניגשים לבחינה היו בעלי רקע לימודי וחברתי-כלכלי דומה.

השוואת הפערים הבין מגדריים במבחני המיצ"ב בכיתה ח' ב-2008 ובמדד הסוציו-אקונומי העלתה כי למרות שבאוכלוסייה הכללית קיים פער קטן עד בינוני במבחני המיצ"ב לטובת בנות (ולא קיים פער במדד החברתי-כלכלי), בקרב הניגשים לבחינה הפסיכומטרית נמצא פער לטובת בנים הן בציוני המיצ"ב והן במדד החברתי-כלכלי (והפער לטובת בנות הצטמצם בשפת אם).

ממצא זה אישש את השערת המחקר הראשונית בדבר קיומם של תהליכי סלקציה עצמית-חברתית שונים בין גברים לנשים בהחלטה לגשת לבחינה הפסיכומטרית. ניתוח צורת התפלגויות של כלל הבנים והבנות נבחני המיצ"ב לעומת ניגשי הפסיכומטרי מביניהם תמך אף הוא בהשערת המחקר: בנים "חלשים" נוטים להעדר מן הבחינה, וכך גם בנות "חזקות". לבסוף, ניתוחי רגרסיה וזקיפה העלו כי אילו התפלגויות הרקע הלימודי-חברתי-כלכלי של הניגשים לבחינה הפסיכומטרית היו זהים בקרב שני המינים – הפערים ביניהם במגזר דוברי העברית היו קטנים בכ-20-10 נקודות.

עמידה על גודל הפער בין המינים בפסיכומטרי לאחר ניכוי השפעת תהליכי הסלקציה הללו מאפשרת להבין טוב יותר את סוגיית הפערים המגדריים. מחקרי המשך דרושים כדי להעמיק ולהבין כמה מן הפער הנצפה הנותר נובע מהבחינה עצמה או מהבדלים בין המינים במאפיינים נוספים שעשויים להשפיע כמו מידת המוטיבציה וההישגיות, חרדת מבחנים וכו'. חשוב לזכור גם כי קיימים הבדלים הפוכים (לטובת תלמידות) בציוני הבגרות – המהווה קריטריון מרכזי נוסף ומאזן בהליך המיון והקבלה הכללי לאוניברסיטאות.

מגמות בציוני הבגרות על פני זמן

כרמל אורן

מרכז ארצי לבחינות ולהערכה (מאל"ו)

דוד מעגן

הלשכה המרכזית לסטטיסטיקה (למ"ס)

יואל רפ

מרכז ארצי לבחינות ולהערכה (מאל"ו)

מטרת עבודה זו היא לעמוד על מגמות שינוי בציוני הבגרות של בוגרי מערכת החינוך בארץ לאורך כשלושה עשורים – משנות התשעים המוקדמות ועד סוף העשור הנוכחי. ציוני הבגרות אינם מכילים לסולם אחיד (רמת הקושי אינה מפוקחת באופן סטטיסטי) ולכן לא ניתן לקשור באופן סיבתי שינויים בממוצעי הציונים השנתיים עם שינויים ברמת ההוראה או ברמת הידע האמיתי של התלמידים.

קירוב לפרקטיקות של כיוול, השתמשנו בציוני בחינת הכניסה הפסיכומטרית לאוניברסיטאות ("הפסיכומטרי") ובמשתני רקע נוספים לצורך תיקון סטטיסטי של רמת היכולת. ציוני הפסיכומטרי מכילים בין נוסחים, מועדים ושפות בחינה, ולפיכך בני השוואה בין מועדים. הפסיכומטרי מתאים במידה רבה למטרת המחקר גם מכיוון שתחומי הדעת הנבחנו בו מתואמים עם המקבילים להם בבגרות.

המדגם שנבדק כולל את נבחני הבגרות בחינוך העברי ובחינוך הערבי ב-26 השנים 1991–2017. בעבור ציון הבגרות הממוצע הכללי ובכל אחת מהבחינות במתמטיקה, אנגלית ושפת אם, בחינוך העברי ובחינוך הערבי בנפרד, בוצע חישוב השינוי השנתי, תוך בקרה ("תיקון") באמצעות משתני היכולת: הפסיכומטרי ותחומיו, ומשתני הרקע: רמת ההיבחנות, מגדר, והשכלת אם. התיקון נעשה באמצעות מודלים לינאריים, שבהם הציונים הפסיכומטריים ומשתני הרקע, וכן אינטראקציות נבחרות ביניהם, משמשים משתנים תלויים, ואילו ציון הבגרות הכללי, והציון בכל מקצוע מהשלושה הנבדקים, הם המשתנה התלוי בכל מודל.

ללא פיקוח על משתני יכולת ורקע: לאורך השנים 1991–2017, ניכרת מגמה ברורה של עלייה בציוני הבגרות הממוצעים – הן בחינוך העברי והן בערבי (0.20 ו-0.33 נקודות בשנה, בהתאמה). בחינוך העברי, עליית ציוני הבגרות אינה מלווה בעלייה דומה בציוני הבחינה הפסיכומטרית. בחינוך הערבי, מגמת העלייה בציוני הבגרות מתמתנת החל משנת 2003, אך חשוב מכך – גם הציונים הפסיכומטריים מצויים במגמת עלייה, אף אם מתונה יותר, לאורך התקופה כולה.

בדיקת הנתונים תוך פיקוח על משתני יכולת ורקע, מעלה ספק באיזו מידה עליית ממוצעי הבגרות משקפת עליה אמיתית ביכולת. בחינוך העברי, העלייה השנתית ה"אינפלציונית" (תוך פיקוח על כל משתני הרקע) ב-26 השנים שנבדקו בציון הבגרות הכללי (ללא בונוסים) היא 0.35 נקודות, שהן כ-10 נקודות לאורך התקופה כולה. קצב העלייה מתמתן עם הזמן: בשנים 2005–2017 שיעור העלייה האינפלציונית השנתית הוא 0.16 נקודות. בחינוך הערבי: מעבר לכל התקופה, שיעור העלייה האינפלציונית השנתית בממוצע הבגרות הוא 0.14, אך בשנים 2005–2017 המגמה נעצרת ונרשמת ירידה זעומה בשיעור של 0.11 נקודות לשנה (שהן 1.1 נקודות מעבר ל-12 השנים הנבדקות).

ברור כי הגדרת השינוי בציוני הבגרות כאמיתי או אינפלציוני במחקר זה היא הגדרה אופרטיבית הנסמכת על קירובים מתוך העדויות המצויות בידינו, ולא בהכרח שלמה. ניתן כמובן להעשיר את חקר הסוגיה בעזרת עדויות נוספות, למשל כאלה הנסמכות על ניתוחים השוואתיים של תוכן וקושי מבחני הבגרות לאורך השנים, אך אלה מצויים מעבר למגבלות המחקר הנוכחי.

Prescreening in Career Decision Making: The Quality of the List of Promising Alternatives

סינון ראשוני בבחירת תחום לימודים ומקצוע:
מה מאפיין רשימת מקצועות מבטיחים טובה?

Tony Gutentag¹, Aviva Shimoni^{1,2}, Itamar Gati¹

¹ The Hebrew University

² Beit Berl College

In many career decisions where the number of alternatives to choose from is large (e.g., choosing a college, a major, a job), the first step is prescreening the alternatives and compiling a list of promising alternatives worth in-depth exploration. The present research developed and tested indicators for assessing such list's quality: list length, the degree of similarity among the alternatives, and the number of career clusters represented by the occupations on the list. With respect to list length, a list of 7 (+/-2) is regarded as optimal (Miller, 1956; "Magic number 7"). With respect to similarity among the promising alternatives, a homogenous list is preferred, as it makes it easier to compare the occupations on the list, decreasing regret from unsatisfactory choice (Sagi & Friedland, 2007). With respect to the number of career clusters represented by the occupations on the list in terms of Holland's (1997) RIASEC types or Roe's (1956) occupational fields, a low number of clusters indicates that the individual has a better sense of future direction. The similarity among occupations is similar but not identical to the number of career clusters; for example, similarity can be fairly great (e.g., accountant and economist), but at the same time the occupations can represent different career clusters (CSI and IAS in terms of Holland's codes).

The goal of Study 1 was to refine and test a model for list quality, using the judgments of 20 career counselors and 103 graduate students. Higher ratings of list quality given by the graduate students were predicted by fewer career clusters, a list length closer to 7 occupations, and greater similarity among the occupations on the list. Among career counselors, only the number of career clusters predicted list quality. Study 2 focused on the association between individuals' career preference cohesiveness (i.e., well-defined and focused preferences; Shimoni, Gutentag, & Gati, 2019) and the quality of the lists of promising occupations compatible with their preferences. As hypothesized, higher preference cohesiveness was associated with better list quality, as rated by both career counselors and graduate students, and as estimated using the aggregate of the three list quality indicators developed in Study 1.

These findings have implication for theory and practice. With respect to theory, the present research validates a parsimonious and plausible model of list quality. With respect to applied implications, list quality indicators can be integrated into self-directed computerized programs, that could prompt users to seek out professional counseling in cases where the quality of lists is suboptimal. In addition, career counselors can use the indicators to detect suboptimal lists, and work with counselees to fine-tune their promising list until achieving an optimal one.

הבסיס הפסיכומטרי לחוסר היעילות של מדדי יעילות

אייל גמליאל
 המרכז האקדמי רופין
 איל פאר
 האוניברסיטה העברית

בספרות מתועדות הטיות בשיפוטיות של חיסכון והטיות בשיפוטיות של ממוצעים שנגרמות בשל השימוש במדדי יעילות כגון מהירות (מדד של יעילות זמן) או יעילות דלק (בישראל: קילומטרים לליטר; בארה"ב: מיילים לגלון). ההסבר שניתן להטיות היה שאנשים סבורים שיש קשר ליניארי בין מדדי יעילות לבין מדדי צריכה, בעוד שהקשר הוא בפועל קורווי-ליניארי. במחקר זה אנו מבקשים לבחון הסבר פרסימוני יותר, אשר לא דורש התייחסות למדדי הצריכה. הסבר זה גורס שאנשים מייחסים למדדי יעילות מאפיינים פסיכומטריים של רמת מדידה של סולם מנה, ובהתאם לכך מעניקים משמעות (שגויה) להפרשים בין ערכים של מדדי יעילות, ומחשבים ממוצעים פשוטים (במקום ממוצעים הרמוניים) של מדדי יעילות. פרשנות זו שגויה, היות ונוסחת החישוב של מדדי היעילות מציבה את מדדי הצריכה במכנה. כך, שבעוד שלמדדי הצריכה יש משמעות ברמת מדידה של סולם מנה, הרי שלמדדי היעילות יש מאפיינים לא אינטואיטיביים: מדדי היעילות משמרים משמעות של מנות בין אובייקטים, אך אינם משמרים את המשמעות של ההפרשים. זאת, בניגוד למדדי הצריכה ולמדדים שכיחים אחרים (כגון משקל וגובה) אשר משמרים את שתי המשמעויות.

על מנת לבחון הסבר זה, התבקשו 71 סטודנטים לדרג חמש אפשרויות של (א) חיסכון בזמן; (ב) חיסכון בדלק; (ג) זוגות ערכים של יעילות זמן (מהירות); ו (ד) זוגות ערכים של יעילות דלק. הערכים שהוצגו למשתתפים בכל המטלות היו כאלה שהענקת משמעות להפרשים בין הערכים הובילה לדירוג הפוך בדיוק לדירוג הנכון.

הממצאים לגבי מטלות החיסכון בזמן הראו ש 88% מהמשתתפים דירגו את האפשרויות בהתאמה להפרש המהירויות ורק 3% דירגו נכון את האפשרויות. לגבי מטלת החיסכון בדלק, 73% דרגו את האפשרויות לפי הצפוי על סמך הפרש היעילויות ורק 10% דירגו נכון את האפשרויות. לגבי ממוצע המהירויות, 81% דירגו את האפשרויות בהתאמה לממוצע האריתמטי ורק 9% דירגו נכון את האפשרויות. לגבי מטלת ממוצע יעילות הדלק, 49% דרגו את האפשרויות לפי הצפוי על סמך ממוצע פשוט ו 21% דירגו נכון את האפשרויות.

ממצאים אלה עקביים עם השערת המחקר שאנשים מייחסים למדדי היעילות מאפייני רמת מדידה של סולם מנה, שבאים לידי ביטוי בחישוב הפרשים ובחישוב ממוצעים אריתמטיים פשוטים. ניתן להתמודד עם הטיות אלה באמצעות הסבר של הטעויות בפירוש מדדי היעילות כאילו שהיו ברמת מדידה של סולם מנה. לחלופין, ניתן להחליף, גם אם באופן הדרגתי, את מדדי היעילות במדדי צריכה. בניגוד לשימוש המקובל בארה"ב ובישראל במדדי יעילות דלק, ישנן מדינות בהן נהוג השימוש במדדי צריכת דלק (כגון כמות ליטרים שרכב צורך ב 100 ק"מ נסיעה). כמו כן, ספורטאים משתמשים במדד צריכה של זמן (קצב; כגון כמה דקות לוקח לאצן לגמא קילומטר) ולא במדד יעילות זמן (מהירות). יתכן והספורטאים למדו שאין הם יכולים לפרש הפרשים של מהירויות ואין הם יכולים לחשב ממוצעים אריתמטיים של מהירויות, ולכן הם משתמשים במדדי צריכת זמן. ממצאי מחקרים מראים כי בהינתן מדדי צריכה, אנשים מחשבים ממוצעים אריתמטיים פשוטים ומחשבים הפרשים בין ערכים, אלא שכעת חישובים אלה הם נכונים ולגיטימיים. החלפת מדדי היעילות במדדי צריכה פועלת כמעין הינד (nudge) שמאפשר לאנשים לעשות שימוש בחשיבה מהירה ואינטואיטיבית, אך להגיע לשיפוטיות נכונים ולקבל החלטות נכונות.

כשפחות שווה יותר: הבנה אינטואיטיבית של משתנה מדכא בתהליכי מיון

חגי רבינוביץ'

אוניברסיטת בן גוריון

יואלה ברבי-מאיר

אוניברסיטת בן גוריון

דוד בודסקו

אוניברסיטת פורדהם

דמיינו שאתם צריכים לבחור בין שני מועמדים למשרת מורה להיסטוריה. על מנת להעריך את הידע שלהם בנושא אתם מעבירים להם מבחן בכתב, המושפע ממהירות הקריאה שלהם. שני המועמדים השיגו ציונים זהים כאשר הראשון קורא מהר יותר מהשני. במי תבחרו? למי ידע רב יותר בהיסטוריה?

פעמים רבות בחירה בין מועמדים עשויה להיות מורכבת מכפי הנדמה, בייחוד כאשר כלי המיון מושפע ממשתנה שאינו רלוונטי לקריטריון. מודל סטטיסטי שינבא ידע בהיסטוריה מתוך ציון המבחן וציון מהירות הקריאה, ינכה את ההשפעה של מהירות הקריאה ויניב נקי יותר עבור ידע בהיסטוריה, כך שלמעשה למועמד בעל מהירות הקריאה הנמוכה יותר, ידע רב יותר. במקרה זה, משתנה מהירות הקריאה יקבל מקדם שלילי במשוואת הרגרסיה וניתן יהיה להתייחס אליו כאל משתנה מדכא (suppressor variable).

פעמים רבות אנו נדרשים לקבל החלטות מסוג זה ללא היכולת להיעזר במודל סטטיסטי. המחקר הנוכחי בחן האם אנשים מקבלים החלטות התואמות את המודל הנורמטיבי כאשר הם נדרשים להחליט באופן אינטואיטיבי, מבלי להיעזר במודל מסוג זה.

בשלושה ניסויים ($N = 357$) הוצגו לנבדקים סיטואציות בהן הם נדרשו לבחור בין שני מועמדים למשרה מסוימת על בסיס ציון במבחן מיון. המבחן תואר כמודד יכולת מסוימת (למשל, ידע בנהיגה ובטיחות בדרכים למשרת נהג משאית), וכמושפע ממשתנה שאינו רלוונטי להצלחה בתפקיד (למשל, שפת אם). הציון שהשיגו המועמדים במבחן תופעל (ציון זהה / שונה) כמו גם העוצמה בה המבחן הושפע מהמשתנה הלא רלוונטי (ללא השפעה / חלשה / חזקה). בנוסף, תופעל סוג המשתנה הלא רלוונטי (מאפיין של המועמד כגון שפת אם / מאפיין של הסיטואציה כגון משך זמן שניתן להשלמת המבחן).

בהתאם להשערותינו, נמצא כי כאשר המשתנה הלא רלוונטי היה מאפיין של המועמד, נבדקים קיבלו החלטות השונות מהמודל הנורמטיבי ובכך הציגו בחירה שאינה אופטימאלית. ניתן לשייך זאת לאופי הלא אינטואיטיבי הנדרש על מנת לנכות את השפעת המשתנה הלא רלוונטי ("להעניש" מועמד כיוון שהוא בעל מהירות קריאה גבוהה או דובר עברית שפת אם). עם זאת, כאשר המשתנה הלא רלוונטי היה מאפיין של הסיטואציה, תהליך התיקון עולה בקנה אחד עם ההבנה האינטואיטיבית ונבדקים אכן קיבלו החלטות תואמות למודל הנורמטיבי. עוד נמצא כי סגנון חשיבה ויכולת נומרית מעורבים גם הם בתהליך קבלת ההחלטות, אם כי במידה מסוימת בלבד.

ממצאי המחקר מראים כי ההבנה האינטואיטיבית של אנשים אינה תואמת את המודל הנורמטיבי בכל הקשור להבנת השפעתם של משתנים מדכאים. לממצאים אלו חשיבות רבה כיוון שבסיטואציות רבות בחיי היום-יום אנשים מקבלים החלטות דומות מבלי להיעזר במודלים סטטיסטיים. הבנת התופעה וכיצד ניתן להתמודד עימה תסייע למועמדים מקבוצות שונות באוכלוסייה לקבל הזדמנות הוגנת יותר, לא כתוצאה של אפליה מתקנת אלא בזכות מדידה נקייה יותר של יכולותיהם האמיתיות.

ניתוח רשתות: קווים מנחים ליישומי מחקר

ענת רוטשטיין
אוניברסיטת חיפה

הצגת נתונים תוך שימוש בעזרים חזותיים נתפסת כברורה, מקצועית, משכנעת ומעניינת יותר. הצגה חזותית של נתונים מתאפשרת, בין היתר, באמצעות ניתוח רשתות פסיכומטרי. ניתוח זה ממפה דפוסים סטטיסטיים מורכבים ומציג את הנתונים כרשת. מתוך רשת הנתונים ניתן לזהות את המבנה של מושגים תאורטיים שונים ולמפות את הדינמיקה בין רכיבי המושג. ניתוח רשתות משמש לקיבוץ משתנים באופן שונה מניתוח גורמים, היות ואין הנחה אודות גורם היררכי סמוי. בנוסף, ניתוח רשתות משמש למיפוי הקשרים בין משתנים המרכיבים מושג תיאורטי נתון, מבחינת עוצמה, אופי הקשר (חיובי או שלילי) ומובהקות. ניתוח רשתות מאפשר גם לזהות היבטים מרכזיים במושג תאורטי נתון באמצעות חישוב תרומתו של כל משתנה או רכיב. ניתוח זה בוצע בשנים האחרונות על מגוון של מושגים כגון אינטליגנציה, איכות חיים ואוטיזם.

רשת נתונים מורכבת משני אלמנטים בסיסיים. רכיבי המושג התיאורטי, שהם המשתנים עצמם (למשל פריטים בשאלון), מיוצגים חזותית כעיוגולים (Nodes), ואילו הקשרים הדו-כיווניים בין אותם משתנים, מיוצגים חזותית כקווים המחברים בין העיוגולים (Edges). קשרים חיוביים מיוצגים ע"י קווים ירוקים וקשרים שליליים ע"י קווים אדומים. עוצמת הקשר מיוצגת באמצעות עובי הקו. מיקום העיוגולים ברשת נקבע לרוב על סמך עוצמת הקשרים בין המשתנים.

קיימים מספר מדדים סטטיסטיים המאפיינים את הרשת כולה. ראשית, ניתן להעריך את חוזק הרשת, או עד כמה המשתנים שברשת מחוברים זה לזה. שנית, ניתן לחשב את הקרבה בין כלל המשתנים ברשת על סמך מספר הקשרים הישירים המינימלי שקיים בין המשתנים. שלישית, ניתן לאפיין את הרשת באמצעות מדדים סטטיסטיים להערכת התקבצות של משתנים (Clustering). קבוצות של משתנים בתוך הרשת מחושבות באמצעות מגוון אלגוריתמים כאשר כל מערך של קבוצות משתנים מדורג בעזרת ציונים ייעודיים.

בנוסף, כל רשת ניתנת לאפיין באמצעות מדדים סטטיסטיים המחושבים ברמת המשתנה (Centrality Scores). מדדים אלו מעריכים את גודל התרומה של כל משתנה לרשת, ביחס לשאר המשתנים. חוזק המשתנה (Strength Centrality) מבוסס על עוצמת הקשרים הקיימים בין משתנה מסוים לשאר המשתנים ברשת. השפעתו העקיפה של המשתנה (Betweenness Centrality) בוחנת עד כמה המשתנה משפיע על משתנים אחרים ברשת גם באופן עקיף. קרבתו של המשתנה למשתנים אחרים (Closeness Centrality) ברשת מושפעת בעיקר מכמות קשריו הישירים של המשתנה עם משתנים אחרים ברשת.

מעבר לאפיין של רשת בודדת, ניתן להשוות בין רשתות שונות באמצעות מדדים סטטיסטיים ייעודיים. השוואה זו מאפשרת בחינה של הבדלים הנוגעים למושג תיאורטי נתון, בין אוכלוסיות שונות ונקודות זמן שונות. הבדלים בין רשתות שונות יכולים להתקיים במבנה, בחוזק הכולל, בעוצמה הממוצעת של הקשרים החיוביים והשליליים שבהן, ובעוצמת הקשר בין שני משתנים מסוימים.

ביצוע ניתוח רשתות פסיכומטרי מתאפשר בשפת R. ספריית qgraph מספקת ממשק להצגה חזותית של נתונים באמצעות ניתוח רשתות. ספרייה זו פותחה בהקשר פסיכומטרי, תוך כוונה לאפשר ייצוג של מושגים תאורטיים ע"י מיפוי המשתנים המרכיבים אותם. באמצעות קוד בסיסי ניתן לבצע ניתוח רשתות ע"י שימוש בספרייה זו עבור רוב סוגי הנתונים. בנוסף, ניתן באמצעותה לחשב מדדים ברמת הרשת וברמת המשתנה הבודד. השוואה בין רשתות מתאפשרת באמצעות ספרייה אחרת בשפת R התומכת בקוד זה.

אנשי קשר ומציגים בכינוס אפי 2020

דוא"ל	שם	
tzur@nite.org.il	צור קרליץ	יו"ר אפי
iawadie.rama@education.gov.il	אימאן עואדיה	יו"ר הכינוס
einatno@openu.ac.il	עינת נוטע-קורן	
netanel@nite.org.il	נתנאל סילברשטיין	יו"ר ועדת הפרס
yelenapo@post.bgu.ac.il	ילנה פורטנוב נאמן	זוכת פרס אפי
michal.linetzky@gmail.com	מיכל לינצקי	זוכת פרס אפי
iawadie.rama@education.gov.il	יו"ר - אימאן עואדיה	מושב א' מושב פתיחה
nizanwa@openu.ac.il	ניצן וינטראוב	
meiravar@cet.ac.il	מירב אריאלי-אטאלי	
tanya.nazaretsky@weizmann.ac.il	טניה נזרצקי	
ymachluf.rama@education.gov.il	יו"ר - יוסי מחלוף	מושב ב' 1 סטטיסטיקה בשירות הפסיכומטריקה
guy@nite.org.il	גיא עשירי-פרוסנר	
dvir@nite.org.il	דביר קלפר	
michalb@nite.org.il	מיכל באומר	
tzur@nite.org.il	יו"ר - צור קרליץ	מושב ב' 2 מבחנים ממוחשבים
aya.lahav@gmail.com	איה להב	
boazt@cet.ac.il	בעז צור	
olshers@edu.haifa.ac.il	שי אולשר	
noamfi@openu.ac.il	יו"ר - נועם פישמן	מושב ג' 1 עיבוד נתונים ומחקר פסיכומטרי
yoel@nite.org.il	יואל רפ	
carmel@nite.org.il	כרמל אורן	
gutentag.tony@mail.huji.ac.il	טוני גוטנטג	
avi@nite.org.il	יו"ר - אבי אללוף	מושב ג' 2 מחקרים בנושאים שונים
eyalg@ruppin.ac.il	אייל גמליאל	
hagaira@gmail.com	חגי רבינוביץ'	
rotstein.anat@gmail.com	ענת רוטשטיין	
ispaisrael@gmail.com	מארגן הכינוס - צור קרליץ	ארגון
metal@nite.org.il	מיטל איילון	

כסימל



כ'א/א

כ"א
