



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

תוקף והוגנות מערכת המיון
לאוניברסיטאות בישראל

דוח מחקר



מאי 2021

תמר קנת-כהן | מאיר ברנרון | כרמל אורן

דוח מחקר

RR-21-05

978-965-502-221-6

תוקף והוגנות מערכת המיון לאוניברסיטאות בישראל

תמר קנת-כהן
מאיר ברנרן
כרמל אורן

מאי 2021

תוכן עניינים

2	תקציר
3	מבוא
3	תוקף
5	הוגנות
7	סיכום של מטרות המחקר וייחודו
8	תוקף מערכת המיון
8	שיטה
10	תוצאות
15	הוגנות מערכת המיון
15	שיטה
17	תוצאות: הוגנות ביחס לשפת היבחנות
21	תוצאות: הוגנות ביחס למגדר
24	משמעות ממצאי ההטיה בברירה במונחי הייצוג של קבוצת המוקד
28	סיכום ודיון
28	תוקף
30	הוגנות
34	מקורות
38	נספחים
38	נספח 1: הרכב החוגים בתחומי הלימוד
39	נספח 2: מקדמי מתאם נצפים

תקציר

במחקר זה נבדקו תוקף והוגנות כלי המיון לאוניברסיטאות בישראל, בהתבסס על ניתוח ההישגים בלימודי שנה א' של סטודנטים שהחלו את לימודיהם בשנים 2013/14-2015/16.

תוקפו של כלי המשמש בבירורה להשכלה הגבוהה מוערך בעיקר על פי מידת הצלחתו בניבוי ההישגים בלימודים ("תוקף ניבוי"). נמצא שתוקף הניבוי של ציון הסכם (43). גבוה מתוקף הניבוי של כל אחד משני מרכיביו בנפרד – הציון הפסיכומטרי (40) וממוצע הבגרות (35). במחקר זה נבדק גם תוקף הניבוי של שני ציונים כלליים בבחינה הפסיכומטרית שהוכנסו לשימוש באוקטובר 2011 – ציון בדגש מילולי וציון בדגש כמותי – ושל ציוני הסכם המבוססים עליהם. נמצא שבתחומי לימוד מילוליים יש יתרון, במונחי תוקף, לציונים בדגש מילולי ובתחומי לימוד כמותיים יש יתרון לציונים בדגש כמותי. ממצאים אלה מצדיקים את ההמלצה להשתמש בציונים בדגש מילולי בהחלטות הקבלה לתחומי לימוד מילוליים ובציונים בדגש כמותי בהחלטות הקבלה לתחומי לימוד כמותיים.

כדי לתת פשר למקדמי המתאם שדווחו, הוצגה, באמצעות הדמיה, התועלת המופקת משימוש בחזאי, כפונקציה של תוקף הניבוי של החזאי ויחס הברירה. ההדמיה המחישה שהרווח הנובע משימוש בחזאי עם תוקף נתון, ובפרט משיפור בתוקפו, גבוה יותר ככל שיחס הברירה נמוך יותר. לפיכך, החוגים והמוסדות אשר מפיקים את התועלת המרבית ממקסימיזציה של תוקף הניבוי הם הסלקטיביים יותר.

שאלת ההוגנות בבירורה נבדקה ביחס לשפת היבחנות בבחינה הפסיכומטרית (ערבית או עברית) ומגדר. בדיקת ההוגנות כללה בדיקה של הטיה בבירורה על פי שני מודלים – מודל הרגרסיה של קלירי והשוואת d-ים של כהן ברוח מודל היחס הקבוע של תורנדייק. בבדיקת הטיה בבירורה כלפי נבחני ערבית נמצא שציון הסכם מוטה לטובת נבחני ערבית, כשמידת ההטיה היא בין ההטיה הניכרת לטובה שמקורה בממוצע הבגרות להטיה לטובה שמקורה בציון הפסיכומטרי. בבדיקת הטיה בבירורה כלפי נשים נמצא, על פי המודל של קלירי, שממדי ההטיה הם זניחים. על פי המודל של השוואת d-ים נמצא שציון הסכם מוטה מעט לטובת נשים, הטיה שהיא תוצר של העובדה שממוצע הבגרות מוטה לטובתן והציון הפסיכומטרי מוטה לרעתן, אך מידת ההטיה שלו פחותה מזו של ממוצע הבגרות. משמעותם של ממצאי ההטיה (על פי המודל של השוואת d-ים) הומחשה, באמצעות הדמיה, כייצוג-יתר (הטיה לטובה) של נבחני ערבית ושל נשים כאשר הברירה מתבססת על ציון הסכם בהשוואה לייצוגם לו הברירה הייתה מתבססת על ההישגים בלימודים.

ממצאי התוקף וההוגנות תומכים בשימוש בכלי המיון הקיימים, ומחזקים את ההמלצה להשתמש בחזאים מרובים בתהליך הברירה להשכלה הגבוהה.

מבוא

המוסדות להשכלה גבוהה בישראל אינם יכולים לקבל ללימודים, לפחות בחלק מהחוגים, את כל מי שפונה אליהם, כיוון שמספר המועמדים עולה על מספר המקומות המוצעים בהם. בגישה המריטוקרטית לקבלת תלמידים, שהיא הגישה הרווחת במדינות העולם להקצאת משאבי השכלה גבוהה (גולדצויג, 2015; טשנר, 2013; Wikström & Wikström, 2020), ניתנת זכות קדימה בקבלה למועמדים שצפויים להיות בעלי הישגים הגבוהים ביותר בלימודים. כדי לחזות את הישגיהם של המועמדים, נעזרים מרבית המוסדות בציוני הבחינה הפסיכומטרית ובציוני תעודת הבגרות: הם מחשבים לכל מועמד ציון קבלה הנקרא 'ציון סְכָם', המבוסס על ציון הבחינה הפסיכומטרית ועל ציוני תעודת הבגרות. המועמדים לכל חוג לימודים מדורגים לפי ציון הסכּם שקיבלו, והגבוהים ביותר במדרג (שמספרם נקבע על פי מספר המקומות בחוג) מתקבלים ללימודים (לרצינול ולרקע היסטורי של מדיניות הקבלה ושל מרכיביה ראו Allalouf, Cohen, & Gafni, 2020).

מדיניות קבלה זו עולה לעיתים תכופות לדיון ציבורי, אשר בו מושמעות ביקורות שונות על השיטה הקיימת, ובפרט על השימוש בבחינה הפסיכומטרית. בין היתר מועלות תהיות בנוגע לכושר החיזוי של הבחינה הפסיכומטרית את הישגי המועמדים בלימודים הגבוהים בכלל, ובתחומי לימוד שונים בפרט, ולתרומתה לחיזוי מעבר לתעודת הבגרות. כמו כן, עולות לעיתים טענות שמערכת המיון, והבחינה הפסיכומטרית בפרט, מוטה לרעת קבוצות מועמדים שונות (שמאופיינות על פי מגדר, שפה, או משתנים אחרים).

המחקר הנוכחי מבקש לענות לטענות אלה על ידי הצגת ממצאים עדכניים על תוקף והוגנות מערכת המיון לאוניברסיטאות בישראל.

תוקף

על פי ה"סטנדרטים למבחנים חינוכיים ופסיכולוגיים" תוקף הוא "המידה שבה ראיות ותאוריה תומכות בפירושים שניתנים לציוני המבחן עבור השימושים המוצעים בהם" (AERA, APA, & NCME, 2014, p. 11). הסטנדרטים מתארים מקורות אחדים שבהם אפשר להשתמש בתהליך של תיקוף, כמו: תוכן המבחן, תהליכי התגובה של נבחנים, המבנה הפנימי של המבחן, וקשרים עם משתנים אחרים. בתיקוף מבחן המשמש לברירה, ראיות המבוססות על קשרים עם משתנים אחרים, ובפרט על הקשר שבין המבחן ובין מה שהוא נועד לנבא (הקריטריון), הן הרלוונטיות ביותר. כך, תוקפו של מבחן המשמש לברירה להשכלה גבוהה נשפט בעיקר על פי מידת הצלחתו בניבוי ההצלחה בלימודים ("תוקף ניבוי").

המדד המקובל ביותר להצלחה בלימודים במחקרי תוקף ניבוי הוא ציון שנה א' לתואר בוגר (Zwick, 2019a). ציון הגמר לתואר בוגר עשוי להיחשב כקריטריון מתאים יותר, אך השימוש בו שכיח פחות, בעיקר כיוון שהוא זמין רק מאוחר יותר, וגם משום שיש נשירת תצפיות בדרך. בהינתן שציון שנה א' הוא לכאורה רק תחליף לקריטריון ה"אולטימטיבי", חשוב לציין שנתונים שנאספו בארץ ובעולם

מעידים באופן עקבי על כך שערכי תוקף הניבוי שמתקבלים בתיקוף מול ציון הגמר אינם נופלים מאלו שמתקבלים בתיקוף מול ציון שנה א' (Zwick, 2006), ולרוב אף גבוהים מהם (קלפר, טורוול ואורן, 2014; Kennet-Cohen, Bronner, & Oren, 1999). עוד נציין שהמתאם בין ציון שנה א' לציון הגמר על פי נתוני מאל"ו הוא 0.85. (ערך גבוה זה, שמתבסס על כ-100,000 תצפיות ממחזורים אחדים, מתקבל למרות קיצוץ התחום הניכר). ניתן להסיק, אם כן, שציון שנה א' הוא תחליף טוב לקריטריון האולטימטיבי – ציון הגמר לתואר בוגר, ובפרט, ששימוש בו אינו כרוך בהערכת יתר של תוקף הניבוי.

המדד המקובל ביותר במחקר על תוקף הניבוי של כלי מיון הוא מקדם המתאם של פירסון בין ההישג בכלי המיון לבין ההישג בלימודים. שני גורמים טיפוסיים במחקר כזה מחלישים את מקדם המתאם הנצפה: טעות מדידה וקיצוץ תחום.

טעות מדידה מגבילה את המתאם הנצפה בין שני מדדים (הוא אינו יכול לעלות על השורש הריבועי של מכפלת המהימנויות של שני המדדים). טעות מדידה קיימת ברוב המשתנים במדעי החברה: המדדים שבהם אנו משתמשים הם רק קירובים למבנים (קונסטרוקטים) שאנחנו רוצים למדוד. ההחלטה אם לתקן עבור טעויות המדידה ("תיקון עבור החלשה") תלויה בשאלות שאנו מבקשים לשאול: במחקר תאורטי אנו עשויים להתעניין בקשרים בין המבנים, ואז יש לתקן עבור טעות המדידה. ברם, המחקר הנוכחי הוא מעשי-יישומי ולא תאורטי, וכיוון שבפועל המוסדות משתמשים בכלים הלא מושלמים הקיימים, אין מקום לתקן עבור טעות המדידה בכלי המיון. אשר לתיקון עבור טעות המדידה בקריטריון, תיקון כזה מאפשר לאמוד את תוקף החזאי לו הקריטריון היה נמדד באופן מושלם. כך, למשל, מהימנות ציון שנה א' לתואר בוגר נאמדה כ-0.75-0.87. במחקרים שנעשו לאחרונה בעולם (Beatty, Walmsley, Sackett, Kuncel, & Koch, 2015; Westrick, 2017), וכ-0.68. במחקר שנעשה בישראל (קנת-כהן, ברונר ואורן, 1999). בהינתן ערכים כאלה, אזי אם, לדוגמה, המתאם הנצפה בין כלי מיון ובין הקריטריון הוא 0.38, ואומדן מהימנות הקריטריון הוא 0.80, המתאם יעלה ל-0.42. בעקבות תיקונו לטעות מדידה בקריטריון. בפרקטיקה לא מקובל לתקן במחקרי תוקף ניבוי בגין טעות המדידה בקריטריון, כיוון שמתייחסים להישג בפועל בלימודים כאל קריטריון שמהימנותו ותוקפו מושלמים. התייחסות זו מוצדקת גם לאור העובדה שהחלטות שונות שנוגעות לקבלה ללימודים מתקדמים, למלגות, לפרסים וכו' מבוססות על ההישג בפועל (Cahan & Gamliel, 2001).

אשר לקיצוץ תחום, מגבלה בסיסית מאוד של מחקר התוקף היא שלרוב הוא יכול להתבסס רק על נתוני חלק מהמועמדים. בהקשר הנוכחי, למשל, מועמדים שציון הסכם שלהם היה נמוך מסף הקבלה לא התקבלו ללימודים, ולכן אין להם ציון בקריטריון, והם לא יכולים להיכלל במחקר. בשל קיצוץ התחום הזה (של ציון הסכם, וכתוצאה מזה, גם של כלי המיון האחרים שמתואמים איתו, וגם של הקריטריון), מקדמי המתאם שיתקבלו במחקר יהיו נמוכים ממה שהם באוכלוסיית המועמדים, ובכך הם יספקו הערכת חסר בנוגע לשאלה הרלוונטית: יעילות כלי המיון באוכלוסיית המועמדים כולה. כדי להמחיש את האפקט של הערכת החסר, נניח, למשל, כלי מיון וקריטריון המתפלגים נורמלית,

שהמתאם ביניהם הוא 50. בקרב המועמדים. נניח שמתקבלים ללימודים מועמדים שציונם בכלי המיון הוא לפחות סטיית תקן אחת מעל למוצע (16% מהמועמדים). בנסיבות אלה, המתאם בין כלי המיון לקריטריון בקרב המתקבלים יהיה 24. וזאת לעומת הערך של 50. באוכלוסיית המועמדים כולה. כדי לאמוד את המתאמים באוכלוסיית המועמדים כולה, משתמשים בתיקון סטטיסטי – "תיקון לקיצוץ תחום" (Gulliksen, 1987; Sackett & Yang, 2000). תיקון כזה מומלץ על ידי הסטנדרטים למבחנים חינוכיים ופסיכולוגיים (AERA et al., 2014) ומהווה נוהל שגרתי במחקרי תוקף ניבוי. כרקע לנתוני תוקף הניבוי שייבדקו במחקר זה, יוצגו להלן ממצאים על תוקף הניבוי של כלי המיון בארה"ב. בהתאם למה שהוסבר לעיל, מקדמי המתאם שיוצגו לא תוקנו עבור טעויות מדידה במשתנים וכן תוקנו לקיצוץ תחום (תחת הנחה פחות שמרנית מזו הנהוגה במחקרי התוקף במאל"ו בנוגע למידת ההטרוגניות של אוכלוסיית המועמדים, נקודה שנתייחס אליה בהמשך). המיון ללימודים לתואר ראשון בארה"ב מתבסס על הציונים במבחן כניסה סטנדרטי (SAT או ACT) ועל ציוני בית הספר התיכון (מוצע הציונים או דירוג בבית הספר). מחקרים עדכניים (ראו, למשל, Westrick et al., 2015; al., 2019; Westrick, Le, Robbins, Radunzel, & Schmidt, 2015) מראים ששני כלי המיון הם חזאים תקפים של ביצוע אקדמי עתידי (כשתוקף הניבוי של ממוצע ציוני בית הספר התיכון הוא 53-58, ותוקף הניבוי של מבחני הכניסה הסטנדרטיים ACT או SAT הוא 51). בנוסף, כאשר משלבים את הציונים משני כלי המיון משתפרת רמת החיזוי של ההישגים האקדמיים (למשל, מתאם מרובה של 61. לצירוף של SAT ושל ממוצע ציוני בית הספר התיכון).

הוגנות

בדיקת ההוגנות של כלי מיון עוסקת בשתי שאלות: האחת, האם שיעור המתקבלים מן הקבוצות השונות, בהסתמך על ביצועיהן בכלי המיון, הוא הוגן? והשנייה, האם עוצמת הקשר בין כלי המיון לקריטריון דומה בקבוצות השונות? ניתן לומר שהשאלה הראשונה עוסקת באיכות הברירה **בין** הקבוצות (הטיה בברירה), והשאלה השנייה עוסקת באיכות הברירה **בתוך** הקבוצות (תוקף דיפרנציאלי). להלן נרחיב בשאלה של הטיה בברירה, שהיא זו שמקושרת באופן טיפוסי להוגנות, ואח"כ נתייחס בקצרה לשאלה של תוקף דיפרנציאלי.

תוצאותיהם של מבחני כניסה סטנדרטיים מצביעות לעיתים קרובות על הבדלים ניכרים בממוצעי הציונים של קבוצות שנבדלות זו מזו במשתנים כמו מגדר או מוצא אתני. לאור זאת, מושמעת לא אחת בציבור או בעיתונות הפופולרית הטענה שהמבחנים מפלים לרעה את הקבוצה שמבצעת במבחן בממוצע פחות טוב מקבוצה אחרת. מפרספקטיבה פסיכומטרית יש הסכמה (AERA et al., 2014) שהבדלי ביצוע בין קבוצות אינם יכולים, כשלעצמם, להעיד על חוסר הוגנות, ושיש לקשר את בדיקת ההוגנות של מבחן או תהליך מיון עם הקריטריון שלמענו הוא נוצר. בהתאם לכך, הוצעו בספרות מודלים שונים, או הגדרות שונות, להטיה בברירה (לסקירה והשוואה בין המודלים ראו, למשל, Darlington, 1971; Petersen & Novick, 1976). המודלים השונים נבדלים ביניהם באופן שבו הם ממשיגים את הזיקה שמתקיימת בין הביצוע במבחן לביצוע בקריטריון במצב הוגן. כפי שמציינים

הכותבים שלעיל, השוני בין ההגדרות נובע משוני בשיפוטם ערכיים (לרוב סמויים) בנוגע לחשיבות היחסית של שיקולים מריטוקרטיים (קבלת המועמדים שסיכויי ההצלחה שלהם בלימודים הם הגבוהים ביותר) ושיקולים חברתיים (ההשלכות של תהליך הברירה על שיעור המתקבלים מקבוצות חברתיות שונות). במחקר זה תיבדק הוגנות מערכת המיון ביחס לשני משתני רקע דיכוטומיים: שפת היבחנות בבחינה הפסיכומטרית (נבחני ערבית – נבחני עברית) ומגדר (נשים – גברים). בדיקת ההוגנות תבסס על יישום שתי הגדרות להוגנות – מודל הרגרסיה של קלירי (Cleary, 1968) והשוואת d-ים של כהן, ברוח מודל היחס הקבוע של תורנדייק (Thorndike, 1971).

ההגדרה המומלצת (AERA et al., 2014) והרווחת ביותר בספרות המקצועית (ראו סקירה אצל Zwick, 2006) לבדיקת הוגנות בברירה מבוססת על מודל הרגרסיה של קלירי. מודל זה מגדיר הטיה בברירה כמצב שבו קו רגרסיה משותף לניבוי הקריטריון על ידי החזאי (כלי המיון) עושה ניבויים שהם באופן שיטתי נמוכים מדי או גבוהים מדי עבור חברי קבוצה מסוימת. כאשר הניבויים נמוכים מדי, נאמר שהחזאי מוטה לרעת חברי הקבוצה, וכאשר הניבויים גבוהים מדי, נאמר שהחזאי מוטה לטובת הקבוצה. הטיה בברירה על פי הגדרה זו נבדקת על ידי חישוב הסטייה הממוצעת של ציון הקריטריון בפועל של חברי קבוצה מסוימת מציון הקריטריון המנובא מקו רגרסיה שמחושב מעבר לקבוצות או על ידי השוואת קווי הרגרסיה של הקריטריון על החזאי בקבוצות שונות.

הגישה המתבססת על השוואת d-ים של כהן נגזרת, תחת כמה הנחות מקובלות (ראו, למשל, Darlington, 1971), ממודל היחס הקבוע של תורנדייק, שמהווה דוגמה בולטת למודלים שמתייחסים לסוגיית הייצוג הקבוצתי. בגישה זו משווים את הפער בין ממוצעי שתי הקבוצות בחזאי לפער ביניהן בקריטריון. כך, למשל, כאשר הקבוצה הנדונה נמוכה יותר הן בחזאי והן בקריטריון (המצב הטיפוסי), אזי אם הפער בין הקבוצות בקריטריון קטן יותר מאשר בחזאי, מסיקים שהחזאי מוטה לרעת הקבוצה, ואם הפער בקריטריון גדול יותר מאשר בחזאי, מסיקים שהחזאי מוטה לטובת הקבוצה. במרבית המקרים יגיעו שתי הגישות למסקנות שונות לגבי מידת ההוגנות של תהליך ברירה מסוים: עם פערים נתונים בין הקבוצות בחזאי ובקריטריון, הגישה של השוואת d-ים, שנותנת משקל גדול יותר לשיקולים חברתיים, תיטה יותר להסקה על הטיה לרעת הקבוצה החלשה ופחות להסקה על הטיה לטובתה בהשוואה למודל הרגרסיה, שמדגיש את העיקרון המריטוקרטי. רק כאשר תוקף הניבוי הוא מושלם או שאין הבדל בין ממוצעי הקבוצות בקריטריון יתלכדו שתי הגישות (בלר ובן-שחר, 1983; Darlington, 1971; Petersen & Novick, 1976).

ממצאים על הוגנות כלי המיון בארה"ב מתבססים רק על מודל הרגרסיה. ממצאי מחקרים עדכניים בנוגע להוגנות של מבחן ה-SAT (Marini et al., 2019) וה-ACT (Sanchez, 2013) מעידים על הטיה לטובת קבוצות מיעוט אתניות (סטודנטים שחורים והיספנים) הן על ידי מבחן הכניסה והן על ידי ממוצע ציוני בית הספר התיכון. תמונה הפוכה (הטיה לרעה) מתקבלת ביחס לנשים.

הממצאים שתוארו לעיל משקפים תמונה טיפוסית של תוצאות מחקרי הוגנות ובפרט את הממצא שמבחני כניסה סטנדרטיים מוטים לטובת קבוצות מיעוט שהישגיהן במבחנים היו נמוכים יותר, כמו שחורים בארה"ב. בהינתן ממצא לא אינטואיטיבי זה ועל רקע ביקורות שונות על מודל הרגרסיה

(Zwick, 2019b), בחרנו להציג גם את היישום של הגישה של השוואת d-ים. גם גישה זו עוררה ביקורת שנוגעת, בין היתר, לסתירה לוגית פנימית במודל (Petersen & Novick, 1976). עבודות שנעשו בהמשך (למשל, Cahan & Gamliel, 2006) הציגו פיתוחים שנועדו להתמודד עם ביקורת זו, אך, ככלל, גישה זו פחות מקובלת בחקירת הוגנות בביררה. הבחירה להציג את שתי הגישות תורמת בשני מובנים: ראשית, התכנסות של שתי ההגדרות מבחינת כיוון ההטיה (כאמור, המסקנות על ממדי ההטיה יהיו בדרך כלל שונות) יכולה לספק אינדיקציה על מידת האיתנות של הממצאים, ושנית, כפי שנרמז לעיל, ההגדרה של הוגנות בביררה טרם הבשילה לכדי קונצנזוס (Cole & Zieky, 2001;) (Zwick, 2019b), בין היתר כי במהותה היא סוגיה ערכית, והדרך הנאותה לשקף זאת היא על ידי הצגת זוויות מבט שונות.

ההיבט השני של הוגנות כלי מיון, תוקף דיפרנציאלי, נבדק על ידי השוואת תוקף הניבוי של כלי המיון בקבוצות השונות. הממצאים בנוגע ל-SAT (Marini et al., 2019) מעידים על תוקף ניבוי נמוך יותר, הן של מבחן הכניסה והן של ממוצע ציוני בית הספר התיכון, בקרב קבוצות מיעוט אתניות (בהשוואה לתוקף בקרב סטודנטים לבנים) ועל תוקף גבוה יותר בקרב נשים (בהשוואה לתוקף בקרב גברים). הממצאים בנוגע ל-ACT (Sanchez, 2013), שמדווחים במונחי סטטיסטיים מבוססי-החלטה (ולא במונחי מקדמי מתאם), אינם מצביעים על הבדלים עקביים בין הקבוצות בעניין זה.

סיכום של מטרות המחקר וייחודו

מטרת המחקר היא הצגת ממצאים עדכניים על תוקף והוגנות מערכת המיון לאוניברסיטאות בישראל. ממצאים אלה מבוססים על ניתוח הישגים האקדמיים של סטודנטים שהחלו את לימודיהם באוניברסיטאות בישראל בשנים 2013/14-2015/16 (המחזורים האחרונים אשר עמדו לרשותנו בעת ביצוע המחקר).

בשאלת התוקף תיבדק מידת הצלחתם של כלי המיון (הבחינה הפסיכומטרית, ממוצע הבגרות וציון הסכם), בניבוי הישגים האקדמיים בתום שנה א' בלימודים באוניברסיטה. שאלת ההוגנות בביררה תיבדק ביחס לשפת היבחנות בבחינה הפסיכומטרית ומגדר. בדיקת ההוגנות תכלול בדיקה של הטיה בביררה על פי שני מודלים – מודל הרגרסיה של קלירי והשוואת d-ים של כהן, ובדיקה של תוקף דיפרנציאלי.

במאל"ו נעשים באופן שוטף מחקרים על תוקף כלי המיון (למשל, אורן, קנת-כהן וברונר, 2007; קלפר, טורוול ואורן, 2014; Oren, Kennet-Cohen, Turvall, & Allalouf, 2014), וכן מחקרי הוגנות ביחס לקבוצות שונות, ובפרט ביחס לקבוצות שפה ומגדר (למשל, קנת-כהן, טורוול ואורן, 2011; Gafni, Beller, & Bronner, 2000; Kennet-Cohen, Turvall, & Oren, 2014; Turvall, 2011; Bronner, Kennet-Cohen, & Oren, 2008). מחזורי הלומדים במחקר הנוכחי הם הראשונים שכוללים סטודנטים שנבחנו בבחינה הפסיכומטרית אחרי השינויים שהוכנסו בה ב-2012 (גפני, 2014) – 38%, 70% ו-81% מן הסטודנטים בשלושת המחזורים שלעיל בהתאמה נבחנו במתכונת החדשה ובמובן זה הם מספקים מידע עדכני וייחודי.

המדגם

המחקר נעשה על נתוני סטודנטים בשש אוניברסיטאות בישראל אשר למדו בשנה א' במחזורים תשע"ד (2013/14) עד תשע"ו (2015/16). המחקר התבסס על 53,212 רשומות של סטודנטים אשר למדו ב-613 יחידות עיבוד. יחידת עיבוד מוגדרת כחוג לימודים במוסד ומחזור (בהמשך נשתמש, לשם פשטות, במונח "סטודנט" לתיאור "רשומה של סטודנט" ובמונח "חוג" לתיאור "יחידת עיבוד"). סטודנט נכלל במחקר אם היו ברשותנו ציונים בקריטריון ובכל החזאים (יפורטו בהמשך). חוג נכלל במחקר אם למדו בו לפחות 20 סטודנטים.

משתני המחקר

קריטריון:

ציון שנה א' בלימודי הבוגר (סולם 100-10)

חזאים:

ממוצע ציוני תעודת הבגרות (בגרות), כפי שחושב באוניברסיטאות, כולל בונוסים (סולם 125-5)

ציונים בתחומי הבחינה הפסיכומטרית¹ (סולם 150-50):

הציון בחשיבה מילולית (מילולי)

הציון בחשיבה כמותית (כמותי)

הציון באנגלית (אנגלית)

ציונים כלליים בבחינה הפסיכומטרית (סולם 800-200):

ציון פסיכומטרי רב-תחומי, שבו משקל הציון בתחום החשיבה המילולית ומשקל הציון בתחום

החשיבה הכמותית כפולים ממשקלו של הציון באנגלית

ציון פסיכומטרי בדגש מילולי, שבו משקלו של הציון בתחום החשיבה המילולית הוא פי שלושה

ממשקל כל אחד מהציונים האחרים

ציון פסיכומטרי בדגש כמותי, שבו משקלו של הציון בתחום החשיבה הכמותית הוא פי שלושה

ממשקל כל אחד מהציונים האחרים

¹ ציוני הבחינה הפסיכומטרית וכן ציוני הסכם המבוססים עליהם נקבעו על פי הבחינה שבה התקבל הציון הפסיכומטרי הרב-תחומי הגבוה ביותר, ובתנאי שנעשתה לפני תחילת הלימודים.

ציוני סכמ, המחושבים כצירופים במשקלות שווים של ממוצע הבגרות והציון הכללי בבחינה הפסיכומטרית בקרב המועמדים למוסד ומחזור, על פי נתוני המועמדים למחזורים 1991/92-1992/93² (עם ממוצע 50 וסטיית תקן 10 בקרב המועמדים למוסד ומחזור):
ציון סכמ רב-תחומי, שהוא צירוף של ממוצע הבגרות והציון הפסיכומטרי הרב-תחומי
ציון סכמ בדגש מילולי, שהוא צירוף של ממוצע הבגרות והציון הפסיכומטרי בדגש מילולי
ציון סכמ בדגש כמותי, שהוא צירוף של ממוצע הבגרות והציון הפסיכומטרי בדגש כמותי

עיבוד הנתונים

בכל אחד מהחוגים חושבו:

1. ממוצעים וסטיות תקן של משתני המחקר
 2. מקדמי מתאם פירסון, מתוקנים לקיצוץ תחום, בין הקריטריון לכל אחד מהחזאים. התיקון לקיצוץ תחום (Gulliksen, 1987) התבסס על ההנחה שהברירה נעשתה על פי ציון הסכמ הרב-תחומי. כאומד לסטיית התקן של משתנה זה בקרב המועמדים שימש שורש הממוצע המשוקלל של שונות ציון הסכמ הרב-תחומי בקרב המועמדים לחוג במוסד ומחזור במחזורים 1991/92-1992/93 (8.5).
- בפרק התוצאות יוצגו ממוצעים משוקללים (במספר הסטודנטים בחוג) של הסטטיסטיים השונים לפי תחום לימודים ומעבר לכל תחומי הלימוד.

² שהם מחזורי המועמדים האחרונים שעומדים לרשותנו

ממוצעים וסטיות תקן של הקריטריון והחזאים

בלוח 1 מוצגים ממוצעים וסטיות תקן של משתני המחקר.

לוח 1: ממוצעים וסטיות תקן של משתני המחקר

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכמ			ציון שנה א'	מספר		תחום לימודים ³
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				בדגש כמותי	בדגש מילולי	ר-ב תחומי		סטוד'	חוגים	
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי								
122.9	112.2	116.8	587.0	597.5	592.3	98.6	55.1	55.9	55.5	83.6	5232	120	רוח ואמנויות
16.8	16.3	15.6	77.4	77.3	75.4	8.3	8.9	9.1	9.0	9.3			
115.6	113.9	115.6	583.3	587.1	585.2	98.6	56.2	56.6	56.4	83.7	13406	132	חברה מילולי
17.4	14.1	14.1	67.4	69.2	66.3	7.4	7.4	7.6	7.4	9.0			
132.7	129.8	132.1	675.4	680.6	677.9	105.5	65.2	65.8	65.6	82.9	2178	12	משפטים
12.3	10.1	9.3	42.1	41.4	39.3	5.8	4.8	4.8	4.7	8.1			
124.8	127.5	123.2	648.3	638.6	643.4	102.3	62.0	61.6	61.9	79.4	5610	40	חברה כמותי
14.8	10.9	11.5	50.7	53.8	50.3	6.6	5.9	6.2	6.0	10.2			
127.8	125.9	124.9	648.5	646.3	647.4	104.0	62.0	62.1	62.1	79.7	4269	51	מדעים ביולוגיים
15.1	11.9	12.1	55.9	57.7	54.8	6.1	5.8	6.0	5.8	11.7			
128.9	130.1	123.7	662.9	648.3	655.6	105.3	63.7	63.0	63.4	77.1	1724	37	מדעים פיסיקליים
15.4	11.1	13.7	55.5	63.6	57.5	6.3	6.2	6.8	6.4	13.6			
131.1	135.9	127.6	689.3	670.3	679.8	108.1	67.4	66.6	67.1	74.1	4756	43	מדעים מדויקים
13.4	8.5	11.1	41.3	49.3	43.2	5.2	4.5	5.0	4.7	14.3			
129.0	135.2	125.7	682.2	660.7	671.4	106.5	65.1	64.4	64.8	77.5	9111	91	הנדסה ואדריכלות
14.2	8.1	11.1	39.6	48.8	42.1	5.7	4.8	5.3	5.0	9.9			
140.8	141.1	139.4	732.6	728.3	730.3	111.1	71.7	72.0	72.0	87.0	1827	20	רפואה
8.0	6.4	6.3	28.0	28.3	26.3	4.4	3.7	3.7	3.7	5.6			
117.2	120.5	118.8	611.1	607.3	609.2	101.2	59.2	59.2	59.2	82.4	5099	67	עזר רפואה
16.3	12.1	11.5	53.8	53.9	51.4	7.3	6.6	6.7	6.6	7.7			
124.1	124.5	122.1	636.3	630.9	633.6	102.8	61.1	61.1	61.1	80.7	53212	613	סך הכול
15.4	11.9	12.5	56.1	59.1	55.6	6.7	6.4	6.6	6.4	10.1			

³ הרכב החוגים בתחומי הלימוד השונים מוצג בנספח 1.

אשר לממוצעי המשתנים, יש להזכיר שציוני שנה א' אינם מכילים בין חוגים, ולכן אינם בני השוואה בין תחומי לימוד שונים⁴. מסיבה זו יש לחשב את מקדמי המתאם בין הקריטריון לחזאים בתוך חוג. בהתייחס לסטיות התקן של המשתנים, ניתן להתרשם מממדי קיצוץ התחום של משתנה הברירה, ציון הסכם הרב-תחומי, אם משווים את סטיות התקן שלו בקרב הלומדים (3.7 ברפואה או 4.7 במשפטים או במדעים מדויקים) לסטיית התקן שלו מעבר לכל המועמדים במוסד ומחזור (שהוגדרה 10) או לסטיית התקן הממוצעת שלו בקרב המועמדים לחוג במוסד ומחזור (8.5)⁵. ממדי קיצוץ התחום ניכרים אף יותר כשמשווים את סטיות התקן של הציון הפסיכומטרי הרב-תחומי בקרב הלומדים (בין 75 בתחום לימודים פחות סלקטיבי כמו רוח ואמנויות ל-26 בתחום סלקטיבי כמו רפואה) לסטיית התקן שלו (מעל 100) מעבר לכל הנבחרים בבחינה הפסיכומטרית (מאל"ו, 2020). ממצאים אלו ממחישים את הצורך לתקן לקיצוץ תחום את המתאמים המחושבים בקרב הלומדים.

תוקף הניבוי של החזאים

בלוח 2 מוצגים מקדמי המתאם בין החזאים לקריטריון (מתוקנים לקיצוץ תחום). בנספח 2 מוצגים מקדמי המתאם הנצפים בין החזאים לקריטריון.

⁴ גם יכולת ההשוואה של ממוצעי ציוני הסכם בתחומי לימוד שונים היא מוגבלת: כיוון שציוני הסכם מחושבים בתוך מוסד (ומחזור) לימודים, הם אינם בני השוואה בין תחומי לימוד שנלמדים בהרכב שונה של מוסדות.

⁵ שהיא הערך (השמרני) שאליו תוקנו המתאמים במחקר זה

לוח 2: מתאמי פירסון בין החזאים לקריטריון

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכם			תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				ר-ב תחומי	בדגש מילולי	בדגש כמותי	
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי					
.31	.31	.37	.38	.41	.40	.37	.45	.46	.46	רוח ואמנויות
.28	.30	.37	.37	.40	.39	.31	.40	.41	.41	חברה מילולי
.28	.24	.36	.32	.38	.36	.30	.34	.37	.35	משפטים
.24	.34	.32	.39	.36	.39	.39	.44	.43	.44	חברה כמותי
.32	.42	.36	.46	.42	.45	.38	.49	.47	.48	מדעים ביולוגיים
.33	.44	.42	.50	.47	.49	.45	.54	.53	.54	מדעים פיסיקליים
.35	.47	.42	.52	.47	.50	.46	.55	.53	.54	מדעים מדויקים
.28	.34	.33	.40	.38	.40	.39	.46	.44	.45	הנדסה ואדריכלות
.23	.29	.25	.34	.32	.33	.31	.34	.33	.33	רפואה
.22	.17	.22	.24	.26	.26	.21	.25	.27	.26	עזר רפואה
.28	.33	.34	.39	.39	.40	.35	.42	.43	.43	סך הכול

מעבר לכל תחומי הלימוד, תוקף הניבוי של ציון הסכם הרב-תחומי (.43) גבוה מתוקף הניבוי של כל אחד משני מרכיביו בנפרד – הציון הפסיכומטרי הרב-תחומי (.40) וממוצע הבגרות (.35). העובדה שתוקף הניבוי של ציון הסכם גבוה מתוקף הניבוי של ממוצע הבגרות מעידה שלציון הפסיכומטרי יש תרומה לניבוי הקריטריון מעבר לממוצע הבגרות. ממצא זה מתקיים גם בחלוקה לתחומי לימוד: בכל תחומי הלימוד גבוה תוקף הניבוי של ציון הסכם מתוקף הניבוי של ממוצע הבגרות, כאשר הפער, שמעיד, כאמור, על תרומתו של הציון הפסיכומטרי לניבוי, נע בין 10. (בחברה מילולי ובמדעים ביולוגיים) ל-02. (ברפואה). במקביל, במרבית תחומי הלימוד, אך לא בכלם, גבוה תוקף הניבוי של ציון הסכם מתוקף הניבוי של הציון הפסיכומטרי, כשהפער בכיוון זה נע בין 06. (ברוח) ל-02. (בחברה מילולי). בשני תחומי לימוד (רפואה ועזר רפואה) זהה תוקף הניבוי של ציון הסכם לתוקף הניבוי של הציון הפסיכומטרי, ובתחום לימודים נוסף (משפטים) הוא נמוך ממנו. תרומתו של הציון הפסיכומטרי מעבר לממוצע הבגרות לניבוי הקריטריון (שיש לה חשיבות פרקטית) היא, אם כן, ניכרת יותר מתרומתו של ממוצע הבגרות מעבר לציון הפסיכומטרי, הן מעבר לכל תחומי הלימוד, והן

בחלוקה לתחומי לימוד, והיא השתקפות של העובדה שתוקף הניבוי של הציון הפסיכומטרי גבוה מתוקף הניבוי של ממוצע הבגרות, הן מעבר לכל תחומי הלימוד, והן בחלוקה לתחומי לימוד (להוציא את חברה כמותי, שם התקפים של שני החזאים זהים).

אשר לציונים בשלושת תחומי הבחינה הפסיכומטרית נמצא שמעבר לכל תחומי הלימוד, תוקף הניבוי של הציון בחשיבה מילולית ושל הציון בחשיבה כמותית דומים (34. ו-33, בהתאמה) ותוקף הניבוי של הציון באנגלית נמוך יותר (28). עם זאת, בחלוקה לתחומי לימוד, בתחומי לימוד מילוליים (רוח ואמנויות, חברה מילולי ומשפטים) נמצא יתרון לציון בחשיבה מילולית על פני הציון בחשיבה כמותית, ובתחומי לימוד כמותיים (למעט בעזר רפואה) נמצא יתרון לציון בחשיבה כמותית על פני הציון בחשיבה מילולית.

כתוצאה מכך, בתחומי לימוד מילוליים גבוה תוקף הציון הפסיכומטרי בדגש מילולי מתוקף הציון הפסיכומטרי בדגש כמותי (וגם מתוקף הציון הפסיכומטרי הרב-תחומי), ובתחומי לימוד כמותיים (למעט בעזר רפואה) גבוה תוקף הציון הפסיכומטרי בדגש כמותי מתוקף הציון הפסיכומטרי בדגש מילולי (וגם מתוקף הציון הפסיכומטרי הרב-תחומי, להוציא את חברה כמותי והנדסה ואדריכלות, שבהם הם זהים).

דפוס דומה של יחסי מתאמים מתקיים גם ביחס לשלושת ציוני הסכם: בתחומי לימוד מילוליים התוקף של ציון הסכם בדגש מילולי הוא הגבוה ביותר (לעיתים זהה לתוקף ציון הסכם הרב-תחומי) ובתחומי לימוד כמותיים (למעט בעזר רפואה) התוקף של ציון הסכם בדגש כמותי הוא הגבוה ביותר (לעיתים זהה לתוקף ציון הסכם הרב-תחומי). ההבדלים בין המתאמים של שלושת ציוני הסכם הם, באופן כללי, קטנים מההבדלים בין המתאמים של שלושת הציונים הפסיכומטריים הכלליים, אך, כאמור, הדפוס הוא דומה. ממצאים אלו מתקפים את ההמלצה להשתמש בציון הסכם בדגש מילולי בהחלטות הקבלה לתחומי לימוד מילוליים ובציון הסכם בדגש כמותי בהחלטות הקבלה לתחומי לימוד כמותיים.

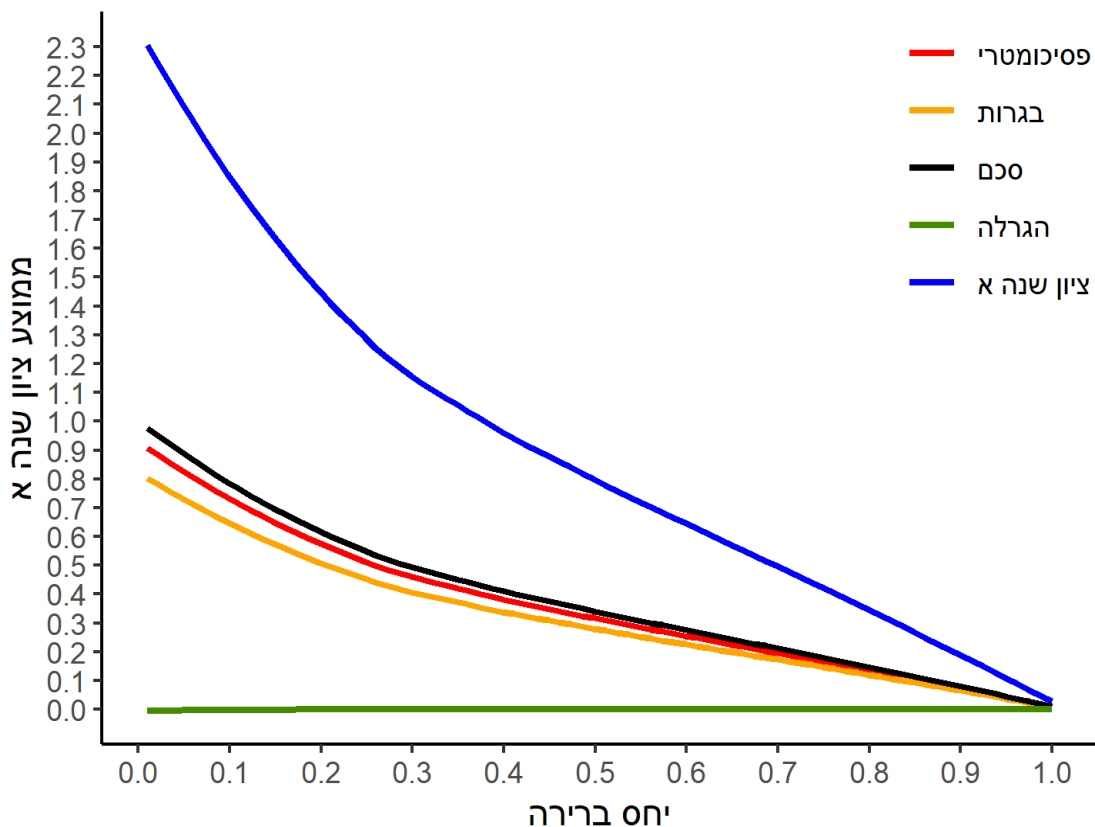
התועלת הנובעת מהשימוש בחזאים

חוקרים לא מעטים ציינו שמקדם מתאם, כמדד לתוקף הניבוי של מבחן, הוא מושג שקשה לתת לו פירוש (Rosenthal & Rubin, 1982; Sackett, Borneman, & Connelly, 2008; Schrader, 1965). הוצעו דרכים שונות לתת פשר למקדם המתאם בין חזאי לקריטריון (כהן, 2012). אחת הדרכים המוצעות היא להציג את רמת הביצוע הצפויה בקריטריון של המתקבלים על סמך החזאי, כפונקציה של תוקף הניבוי של החזאי ויחס הברירה (שיעור המתקבלים מבין המועמדים), כפי שיודגם להלן. כדי לעשות זאת נדרשים ציוני הקריטריון של כל המועמדים, שלא קיימים, כמובן, כיוון שלמועמדים שלא התקבלו אין ציוני קריטריון. הפתרון לכך הוא לעשות הדמיה של אוכלוסיית המועמדים על סמך המתאם, שהתקבל במחקר, בין החזאי לקריטריון.

להלן יתוארו השיטה והתוצאות של הדמיה כזו. עבור כל אחד מ-5 חזאים הוגדרה אוכלוסייה היפותטית של מועמדים שמתפלגת דו-נורמלית בשני משתנים: החזאי הנדון וקריטריון ("ציון שנה א"). חמשת החזאים היו: תואם הציון הפסיכומטרי הרב-תחומי ("פסיכומטרי"), תואם ממוצע הבגרות

("בגרות"), תואם ציון הסכם הרב-תחומי ("סכם"), חזאי עם תוקף אפס ("הגרלה"), וחזאי עם תוקף מושלם (הקריטריון עצמו – "ציון שנה א"). לכל המשתנים נקבעו ממוצע 0 וסטיית תקן 1. המתאמים בין פסיכומטרי, בגרות וסכם ובין הקריטריון נקבעו, בהתבסס על לוח 2, כ-0.40, 0.35 ו-0.43, בהתאמה, והמתאמים בין הגרלה וציון שנה א ובין הקריטריון נקבעו כ-0 ו-1, בהתאמה. אחרי שהוגדרה אוכלוסיית המועמדים נעשו בה תהליכי ברירה, בהתבסס על החזאי, ביחסי ברירה שונים. בכל צירוף של חזאי ושל יחס ברירה נבדק ממוצע הציונים בקריטריון בקרב המתקבלים. כך, למשל, כאשר החזאי היה פסיכומטרי ויחס הברירה היה 0.3, נבחרו 30% מהמועמדים הגבוהים ביותר בפסיכומטרי, וחושב ממוצע ציון שנה א שלהם. תרשים 1 מציג את תוצאות ההדמיה.

תרשים 1: ממוצע ציון שנה א בקרב המתקבלים (בהינתן ממוצע 0 וסטיית תקן 1 בקרב המועמדים) כפונקציה של החזאי הממין ויחס הברירה⁶



התוצאות ממחישות מספר עקרונות חשובים בנוגע לתועלת הכלים המשמשים למיון המועמדים לאוניברסיטאות בישראל. ראשית, ככל שיחס הברירה נמוך יותר, כך תועלת המיון (שנמדדת כאן על ידי ממוצע ציון שנה א) גדולה יותר. במילים אחרות, החוגים (והמוסדות) שמפיקים את התועלת

⁶ ההדמיה נעשתה עבור יחסי ברירה שבין 0.01 ל-1 בצעדים של 0.01. בתרשים זה, וכן בתרשימים הבאים, הנקודות חוברו באמצעות החלקה (smoothing).

המרבית ממיון מועמדים באמצעות חזאי עם תוקף ניבוי נתון הם החוגים (והמוסדות) הסלקטיביים ביותר. שנית, עבור כל יחס ברירה נתון, תועלת המיון על סמך ממוצע הבגרות קטנה מזו של מיון על סמך הציון הפסיכומטרי ושימוש בציון הסכם מניב את התועלת הגבוהה ביותר. כך, למשל, נניח חוג שמקבל 10% בלבד מהמועמדים. אם המיון מתבסס על ממוצע הבגרות בלבד יהיה ממוצע ציון שנה א' של המתקבלים 0.62, אם המיון מתבסס על הציון הפסיכומטרי בלבד יהיה ממוצע ציון שנה א' של המתקבלים 0.70, ואם המיון מתבסס על ציון הסכם יהיה ממוצע ציון שנה א' של המתקבלים 0.76. ניתן להשוות תוצאות אלו לשתי נקודות התייחסות: אם המיון מתבסס על הגרלה, מה ששקול למיון על סמך חזאי בעל תוקף אפסי, הרי שממוצע ציון שנה א' של המתקבלים יהיה 0. לעומת זאת, אם המיון מתבסס על ציון שנה א' (לו היה ידוע לנו), מה ששקול למיון על סמך חזאי בעל תוקף מושלם, הרי שממוצע ציון שנה א' של המתקבלים יהיה 1.76.

הממצאים שתוארו לעיל מדגימים את העיקרון שהוכיח ברוגדן (Brogden, 1946) שתוקף הניבוי של חזאי נמצא ביחס ישר לגודל התועלת המופקת משימוש בו: בדוגמה שתוארה, הפער בין ממוצע ציון שנה א' של המתקבלים על סמך חזאי בעל תוקף מושלם לבין ממוצע המתקבלים על סמך הגרלה עומד על 1.76 סטיות תקן. במצב כזה, שימוש בציון הסכם, שתוקף הניבוי שלו 0.43, מניב קבוצת מתקבלים שממוצע ציוניה גבוה ב-0.76 סטיות תקן (0.43×1.76) מזה של המתקבלים על סמך הגרלה. קשר זה מתקיים ומודגם בתרשים 1 עבור כל אחד מהחזאים: ביחס ברירה נתון, היחס בין גובה העקומה של החזאי ובין גובה העקומה של ציון שנה א' זהה לתוקף הניבוי של החזאי.

הוגנות מערכת המיון

שיטה

המדגם

המדגם הראשוני שממנו נגזרו המדגמים הסופיים לשתי בדיקות ההוגנות הוא המדגם שתואר במחקר התוקף, ללא הדרישה של לפחות 20 סטודנטים בחוג.

בבדיקת ההוגנות ביחס לשפת היבחנות נוספו התנאים שהסטודנט נבחן בבחינה הפסיכומטרית בעברית או בערבית, ושהוא למד בחוג שבו לפחות 5 סטודנטים מכל שפת היבחנות. המדגם הסופי כלל 34,613 סטודנטים מ-340 חוגים.

בבדיקת ההוגנות ביחס למגדר נוספו התנאים שהמגדר של הסטודנט היה ידוע, ושהוא למד בחוג שבו לפחות 5 סטודנטים מכל מגדר. המדגם הסופי כלל 50,413 סטודנטים מ-588 חוגים.

משתני המחקר

משתני המחקר היו זהים למשתנים אשר תוארו לעיל במחקר התוקף.

עיבוד הנתונים

בכל בדיקת הוגנות הוגדרו שתי קבוצות: קבוצת מוקד (נבחני ערבית בבדיקת ההוגנות ביחס לשפת היבחנות; נשים בבדיקת ההוגנות ביחס למגדר) וקבוצת התייחסות (נבחני עברית בבדיקת ההוגנות ביחס לשפת היבחנות; גברים בבדיקת ההוגנות ביחס למגדר).

בכל בדיקת הוגנות, חושבו בכל אחד מהחוגים:

1. ממוצעים וסטיות תקן של משתני המחקר בכל קבוצה (מוקד והתייחסות)
2. תוקף דיפרנציאלי: מקדמי מתאם פירסון, מתוקנים לקיצוץ תחום, בין הקריטריון לחזאים בכל קבוצה (מוקד והתייחסות)
3. הטיה בברירה לפי שני המודלים האלה:
לפי מודל הרגרסיה של קלירי ("קלירי")

ראשית, ציון הקריטריון תוקנן על פי הממוצע וסטיית התקן מעבר לשתי הקבוצות (מוקד והתייחסות) בחוג. התקנון בוצע על מנת להביא את ציוני הקריטריון לסולם בר השוואה בין חוגים, כיוון שהתוצאות המוצגות הן ממוצעים מעבר לחוגים שונים. שנית, עבור כל אחד מן החזאים חושבה משוואת רגרסיה לניבוי הקריטריון מעבר לשתי הקבוצות. שלישית, עבור קבוצת המוקד חושבה הסטייה הממוצעת בין ציון הקריטריון המנובא על פי משוואת הרגרסיה המשותפת לבין הציון בפועל (ציון מנובא פחות ציון בפועל). ערך חיובי מבטא הטיה לטובת קבוצת המוקד וערך שלילי הטיה לרעתה.

לפי השוואת d-ים של כהן ("הפרש d-ים")

חושב פער מתוקנן (מדד Cohen's d לגודל אפקט) בין ממוצעי הקבוצות (ממוצע קבוצת התייחסות פחות ממוצע קבוצת המוקד) בחזאים ובקריטריון. המדד להטיה בברירה (הפרש d-ים) הוגדר כ-d בקריטריון פחות ה-d בחזאי. ערך חיובי של הפרש זה מבטא הטיה לטובת קבוצת המוקד וערך שלילי הטיה לרעתה.

חישוב ה-d-ים התבסס על הנוסחה $d = r / \sqrt{p * q(1 - r^2)}$ שבה: r הוא מקדם המתאם פירסון, מתוקן לקיצוץ תחום, בין משתנה השיוך הקבוצתי (0 = קבוצת המוקד, 1 = קבוצת התייחסות) לבין המשתנה הרציף הנדון (הקריטריון או אחד החזאים) ו-p ו-q הם הפרופורציות של הקבוצות בקרב הלומדים בחוג.

בפרק התוצאות יוצגו ממוצעים משוקללים של הסטטיסטים השונים בתחומי לימוד מילוליים, תחומי לימוד כמותיים⁷ ומעבר לכל תחומי הלימוד.

תוצאות: הוגנות ביחס לשפת היבחנות

ממוצעים וסטיות תקן של הקריטריון והחזאים

בלוח 3 מוצגים ממוצעים וסטיות תקן של משתני המחקר, לפי שפת היבחנות.

לוח 3: ממוצעים וסטיות תקן של משתני המחקר, לפי שפת היבחנות^a

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכום			ציון שנה א'	מספר		קבוצה	תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				בדגש כמותי	בדגש מילולי	רב-תחומי		סטוד'	חוגים		
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	רב-תחומי									
120.8	116.5	119.4	602.2	608.8	605.5	98.6	57.1	57.6	57.4	85.7	8768	134	נבחני עברית	מילוליים
16.3	14.0	13.6	66.7	67.2	65.1	7.4	7.5	7.7	7.5	7.4				
99.8	107.0	103.4	528.1	519.8	523.9	99.4	54.4	53.9	54.1	74.2	3421	134	נבחני ערבית	
15.3	13.8	13.5	62.1	62.6	60.0	7.4	7.2	7.3	7.2	10.7				
130.7	131.5	127.5	673.9	664.8	669.3	105.2	63.9	63.8	63.9	80.6	17238	206	נבחני עברית	כמותיים
13.0	9.6	10.6	44.6	48.4	44.4	5.6	5.1	5.4	5.2	9.8				
116.8	128.9	121.0	641.7	623.8	632.7	108.6	64.6	63.9	64.3	72.5	5186	206	נבחני ערבית	
14.9	10.3	10.8	45.6	47.9	44.2	5.0	4.9	5.1	4.9	11.5				
127.4	126.4	124.8	649.8	645.9	647.8	103.0	61.6	61.7	61.7	82.3	26006	340	נבחני עברית	סך הכול
14.2	11.3	11.7	53.1	55.5	52.3	6.3	6.0	6.2	6.1	9.1				
110.0	120.2	114.0	596.5	582.5	589.5	104.9	60.6	59.9	60.2	73.2	8607	340	נבחני ערבית	
15.1	11.8	11.9	52.8	54.3	51.1	6.1	5.9	6.1	5.9	11.2				

^a ממוצעים משוקללים במספר הסטודנטים בקבוצה בחוג

עיון בממוצעי הקבוצות מעבר לכל החוגים מראה שבציונים הכלליים בבחינה הפסיכומטרית נמוך ממוצע נבחני ערבית מממוצע נבחני עברית בכסטיית תקן (במונחי סטיית התקן הממוצעת בקבוצה בחוג), ובממוצע הבגרות גבוה ממוצע נבחני ערבית מממוצע נבחני עברית בכשליש סטיית תקן. אין משמעות להשוואה בין הקבוצות בציון שנה א' ובציוני הסכום כיוון שהם אינם מכילים בין חוגים (ציון שנה א' אינו מכיל בין חוגים בתוך מוסד ומחזור, וציוני הסכום אינם מכילים בין חוגים ממוסדות ומחזורים שונים).

⁷ בשל האילוצים הנוספים על המדגם בבדיקות ההוגנות, בחרנו להציג את התוצאות לפי סיווג גס יותר של תחומי הלימוד: תחומי לימוד מילוליים (שכוללים את התחומים רוח ואמנויות, חברה מילולי, ומשפטים) ותחומי לימוד כמותיים (כל היתר).

בלוח 4 מוצגים מקדמי המתאם (מתוקנים לקיצוץ תחום) בין החזאים לקריטריון, לפי שפת היבחנות.

לוח 4: מתאמי פירסון בין החזאים לקריטריון, לפי שפת היבחנות^a

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכם			קבוצה	תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				ר-ב תחומי	בדגש מילולי	בדגש כמותי		
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי						
.23	.32	.36	.37	.38	.38	.45	.48	.49	.48	נבחני עברית	מילוליים
.16	.18	.19	.23	.22	.23	.33	.33	.32	.33	נבחני ערבית	
.27	.42	.37	.46	.42	.45	.51	.55	.53	.54	נבחני עברית	כמותיים
.29	.34	.28	.40	.36	.39	.47	.48	.46	.47	נבחני ערבית	
.26	.39	.36	.43	.40	.42	.49	.52	.51	.52	נבחני עברית	סך הכול
.24	.28	.25	.33	.30	.33	.41	.42	.41	.41	נבחני ערבית	

^a ממוצעים משוקללים במספר הסטודנטים בקבוצה בחוג

הממצאים בלוח 4 מעידים שמעבר לכל תחומי הלימוד תוקף הניבוי של חזאי מערכת המיון גבוה יותר בקרב נבחני עברית. ההפרש בין המתאמים בשתי הקבוצות נע בין 0.11. (בציון הסכם הרב-תחומי ובציונים בתחומים חשיבה מילולית וחשיבה כמותית בבחינה הפסיכומטרית) ל-0.02. (בציון בתחום האנגלית).

הפערים בתוקף לרעת נבחני ערבית ניכרים יותר בתחומי הלימוד המילוליים.

חשוב להבהיר, עם זאת, שחלק מהאפקטים שתוארו לעיל מקורם בכך שהתפלגות שתי הקבוצות לחוגים היא שונה (בחוגים שונים יש פרופורציות שונות של נבחני ערבית ונבחני עברית), וחוגים נבדלים זה מזה ביכולת החזאים לנבא את הקריטריון.

בלוח 5 מוצגים שני המדדים להטיה בברירה ביחס לשפת היבחנות.

לוח 5: מדדי קלירי והפרש d-ים להטיה בברירה כלפי נבחני ערבית^a

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכמ			מדד	תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				ר-ב תחומי	בדגש מילולי	בדגש כמותי		
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי						
0.52	0.60	0.53	0.54	0.50	0.51	0.69	0.63	0.61	0.62	קלירי ^b	מילוליים
1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	d קריטריון ^c	
0.77	0.05	0.57	0.37	0.60	0.50	-0.81	-0.29	-0.15	-0.22	d חזאי ^c	
0.62	1.34	0.83	1.02	0.79	0.89	2.20	1.68	1.55	1.61	הפרש d-ים ^d	
0.45	0.53	0.49	0.48	0.46	0.47	0.59	0.58	0.56	0.57	קלירי ^b	כמותיים
0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	d קריטריון ^c	
0.59	-0.29	0.12	-0.02	0.16	0.07	-1.09	-0.69	-0.58	-0.64	d חזאי ^c	
0.17	1.05	0.63	0.78	0.60	0.69	1.85	1.45	1.34	1.39	הפרש d-ים ^d	
0.48	0.56	0.51	0.50	0.48	0.49	0.63	0.60	0.58	0.59	קלירי ^b	סך הכול
0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	d קריטריון ^c	
0.65	-0.17	0.28	0.12	0.31	0.22	-0.99	-0.55	-0.43	-0.49	d חזאי ^c	
0.33	1.15	0.70	0.86	0.67	0.76	1.97	1.53	1.41	1.47	הפרש d-ים ^d	

^a הממד של קלירי משוקלל במספר הסטודנטים בקבוצת המוקד בחוג, כיוון שהוא אפיון של הקבוצה (שקלול כזה הוא אנלוגי לסכימה מעבר לחוגים של הסטייה של הציון בפועל של חברי קבוצת המוקד מקו הניבוי, ולחלוקת הסכום במספר חברי קבוצת המוקד). הממד של הפרש d-ים וה-d-ים עצמם משוקללים במספר הסטודנטים הכולל בחוג, כיוון שה-d-ים, בכך שהם מבטאים השוואה בין הקבוצות, הם אפיון מעבר לקבוצות.

^b הממד של קלירי הוא במונחי סטיות תקן של הקריטריון (ציון מנובא פחות ציון בפועל).

^c כיוון ה-d-ים הוא ממוצע נבחני עברית פחות ממוצע נבחני ערבית.

^d הפרש ה-d-ים הוא d בקריטריון פחות d בחזאי.

נזכיר שבהינתן האופן שבו הוגדרו שני המדדים להטיה בברירה (ראו הערות בתחתית הלוח) ערך חיובי במדד מבטא הטיה לטובת נבחני ערבית וערך שלילי – הטיה לרעתם. על פי המדד של קלירי כל החזאים מוטים לטובת נבחני ערבית. מידת ההטיה של ממוצע הבגרות היא הגדולה ביותר (0.63 סטיות תקן), מידת ההטיה של הציונים הכלליים בבחינה הפסיכומטרית קטנה יותר (0.49 סטיות תקן בממוצע), ומידת ההטיה שנובעת מציוני הסכם היא ביניהן (0.59 סטיות תקן בממוצע). מבין תחומי הבחינה הפסיכומטרית, מידת ההטיה הרבה ביותר לטובת נבחני ערבית היא בחשיבה כמותית (0.56 סטיות תקן), והקטנה ביותר באנגלית (0.48 סטיות תקן).

גם על פי המדד של הפרש d-ים מוטים כל החזאים לטובת נבחני ערבית. דירוג החזאים על פי מידת ההטיה שלהם דומה למה שתואר ביחס למדד של קלירי: מידת ההטיה של ממוצע הבגרות היא הגדולה ביותר (1.97), והיא תוצר של העובדה שבהסתכלות תוך-חוגית ממוצע הבגרות של נבחני ערבית גבוה מממוצע הבגרות של נבחני עברית בכסטיית תקן (d בחזאי: -0.99), אך ציוניהם בקריטריון נמוכים מציוניהם של נבחני עברית בכסטיית תקן (d בקריטריון: 0.98). המינוח "הסתכלות תוך-חוגית" נועד להדגיש שבחישוב ה-d-ים משווים את ממוצעי שתי הקבוצות **בתוך חוג** (וממצעים מעבר לחוגים). התוצר של חישוב כזה הוא שונה מההתרשמות שעשויה להתקבל מהשוואת ממוצעי שתי הקבוצות מעבר לחוגים (לוח 3) במצב שבו התפלגות שתי הקבוצות לחוגים היא שונה. מידת ההטיה של הציונים הכלליים בבחינה הפסיכומטרית קטנה יותר (0.76 בממוצע), ומידת ההטיה שנובעת מציוני הסכם היא ביניהן (1.47 בממוצע). מבין תחומי הבחינה הפסיכומטרית, מידת ההטיה לטובת נבחני ערבית היא הרבה ביותר בחשיבה כמותית (1.15), והקטנה ביותר באנגלית (0.33). הבדלים אלו במידת ההטיה של הציונים בתחומי הבחינה הפסיכומטרית, באים לידי ביטוי בהשוואה בין שלושת הציונים הכלליים בבחינה הפסיכומטרית ובין שלושת ציוני הסכם: ההטיה לטובת נבחני ערבית גדולה יותר בציון בדגש כמותי.

הדפוסים שתוארו לעיל, בנוגע למידת ההטיה של החזאים השונים, מתקיימים הן בתחומי לימוד מילוליים והן בתחומי לימוד כמותיים, כאשר מידת ההטיה של כל החזאים לטובת נבחני ערבית היא רבה יותר בתחומי לימוד מילוליים.

תוצאות: הוגנות ביחס למגדר

ממוצעים וסטיות תקן של הקריטריון והחזאים

בלוח 6 מוצגים ממוצעים וסטיות תקן של משתני המחקר, לפי מגדר.

לוח 6: ממוצעים וסטיות תקן של משתני המחקר, לפי מגדר^a

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכמ			ציון שנה א'	מספר		קבוצה	תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				בדגש כמותי	בדגש מילולי	ר-ב תחומי		סטוד'	חוגים		
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי									
125.2	119.4	120.1	618.0	619.6	618.8	97.9	57.2	57.4	57.3	83.7	6815	256	מילוליים	גברים
16.7	14.5	14.4	69.1	70.3	67.8	8.0	8.0	8.1	8.0	8.9				נשים
116.9	113.5	116.9	584.8	592.5	588.7	100.1	56.8	57.4	57.1	83.8	12755	256	כמותיים	גברים
16.2	13.8	13.6	64.6	66.1	63.4	7.0	7.2	7.4	7.3	8.9				נשים
130.5	133.7	125.7	679.0	660.8	669.8	104.7	64.3	63.6	64.0	78.3	16054	332	סך הכול	גברים
14.1	9.3	11.6	44.6	51.9	46.2	6.0	5.3	5.7	5.5	11.2				נשים
123.8	127.4	124.2	648.3	641.1	644.6	105.6	63.4	63.3	63.4	78.9	14789	332	סך הכול	גברים
14.4	10.4	11.0	47.9	51.1	47.4	5.8	5.4	5.6	5.5	10.3				נשים
128.9	129.4	124.0	660.8	648.5	654.6	102.7	62.2	61.8	62.0	79.9	22869	588	סך הכול	גברים
14.9	11.1	12.5	53.1	58.0	53.6	6.7	6.3	6.5	6.3	10.6				נשים
120.6	120.9	120.8	618.9	618.6	618.7	103.1	60.3	60.6	60.5	81.2	27544	588		נשים
15.3	12.1	12.3	56.2	58.5	55.3	6.4	6.3	6.5	6.4	9.7				

^a ממוצעים משוקללים במספר הסטודנטים בקבוצה בחוג

עיון בממוצעי הקבוצות מעבר לכל החוגים מראה שבציונים הכלליים בבחינה הפסיכומטרית נמוך ממוצע הנשים מממוצע הגברים בכשני שליש סטיות תקן בממוצע (במונחי סטיית התקן הממוצעת בקבוצה בחוג), ובממוצע הבגרות יש יתרון זניח לנשים. כאמור, אין משמעות להשוואה בין הקבוצות בציון שנה א' ובציוני הסכמ כיוון שהם אינם מכילים בין חוגים.

בלוח 7 מוצגים מקדמי המתאם (מתוקנים לקיצוץ תחום) בין החזאים לקריטריון, לפי מגדר.

לוח 7: מתאמי פירסון בין החזאים לקריטריון, לפי מגדר^a

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכם			קבוצה	תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				ר-ב תחומי	בדגש מילולי	בדגש כמותי		
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי						
.24	.29	.34	.35	.37	.37	.35	.41	.42	.42	גברים	מילוליים
.32	.31	.40	.39	.43	.42	.33	.41	.43	.42	נשים	
.27	.36	.32	.41	.37	.40	.40	.46	.44	.45	גברים	כמותיים
.30	.35	.34	.41	.39	.41	.36	.43	.42	.43	נשים	
.26	.34	.33	.39	.37	.39	.39	.44	.44	.44	גברים	סך הכול
.31	.33	.37	.40	.41	.41	.35	.42	.43	.43	נשים	

^a ממוצעים משוקללים במספר הסטודנטים בקבוצה בחוג

הממצאים בלוח 7 מעידים שמעבר לכל תחומי הלימוד ההבדלים בין קבוצות המגדר בתוקף הניבוי של חזאי מערכת המיון הם קטנים יחסית: תוקף הניבוי של ציון הסכם דומה בשתי הקבוצות, תוקף הניבוי של ציוני הבחינה הפסיכומטרית גבוה יותר בדרך כלל בקרב הנשים (הפרש מקסימלי של 0.05. בציון בתחום האנגלית), ותוקף הניבוי של ממוצע הבגרות גבוה יותר בקרב הגברים (הפרש של 0.04). הקשרים שתוארו מתקיימים בדרך כלל בשני תחומי הלימוד, כאשר בתחומי הלימוד המילוליים הפער בין גברים לנשים בתוקף הניבוי של ציוני הבחינה הפסיכומטרית (לטובת הנשים) הוא גדול יותר (הפרש מקסימלי של 0.08. בציון בתחום האנגלית).

כפי שצוין קודם, חלק מהאפקטים שתוארו לעיל מקורם בכך שבחוגים שונים יש פרופורציות שונות של גברים ונשים, וחוגים נבדלים זה מזה ביכולת החזאים לנבא את הקריטריון.

בלוח 8 מוצגים שני המדדים להטיה בברירה ביחס למגדר.

לוח 8: מדדי קלירי והפרש d-ים להטיה בברירה כלפי נשים^a

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכמ			מדד	תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				ר-ב תחומי	בדגש מילולי	בדגש כמותי		
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי						
-0.02	-0.02	-0.01	-0.04	-0.03	-0.03	0.03	0.00	0.01	0.00	קלירי ^b	מילוליים
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	d קריטריון ^c	
0.39	0.31	0.13	0.35	0.26	0.31	-0.37	-0.03	-0.08	-0.05	d חזאי ^c	
-0.40	-0.32	-0.13	-0.36	-0.26	-0.32	0.36	0.02	0.07	0.04	הפרש d-ים ^d	
0.01	0.00	0.03	-0.01	0.01	0.00	0.06	0.04	0.05	0.05	קלירי ^b	כמותיים
-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	-0.03	d קריטריון ^c	
0.20	0.11	-0.15	0.09	-0.03	0.02	-0.52	-0.29	-0.34	-0.31	d חזאי ^c	
-0.22	-0.14	0.12	-0.12	0.01	-0.05	0.50	0.26	0.31	0.29	הפרש d-ים ^d	
-0.01	-0.01	0.01	-0.02	-0.01	-0.02	0.05	0.02	0.03	0.03	קלירי ^b	סך הכול
-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	d קריטריון ^c	
0.27	0.19	-0.04	0.19	0.08	0.14	-0.46	-0.19	-0.24	-0.21	d חזאי ^c	
-0.29	-0.21	0.02	-0.21	-0.10	-0.15	0.44	0.17	0.22	0.19	הפרש d-ים ^d	

^a הממד של קלירי משוקלל במספר הסטודנטים בקבוצת המוקד בחוג, כיוון שהוא אפיון של הקבוצה (שקלול כזה הוא אנלוגי לסכימה מעבר לחוגים של הסטייה של הציון בפועל של חברי קבוצת המוקד מקו הניבוי, ולחלוקת הסכום במספר חברי קבוצת המוקד). הממד של הפרש d-ים וה-d-ים עצמם משוקללים במספר הסטודנטים הכולל בחוג, כיוון שה-d-ים, בכך שהם מבטאים השוואה בין הקבוצות, הם אפיון מעבר לקבוצות.

^b הממד של קלירי הוא במונחי סטיות תקן של הקריטריון (ציון מנובא פחות ציון בפועל).

^c כיוון ה-d-ים הוא ממוצע גברים פחות ממוצע נשים.

^d הפרש ה-d-ים הוא d בקריטריון פחות d בחזאי.

על פי המדד של קלירי ממדי ההטיה זניחים ביחס לכלל החזאים.

על פי המדד של הפרש d-ים ממוצע הבגרות מוטה לטובת הנשים (0.44): בהסתכלות תוך-חוגית ממוצע הבגרות של הנשים גבוה מממוצע הבגרות של הגברים (d בחזאי: -0.46), בעוד שציוני שתי הקבוצות בקריטריון הם דומים (d בקריטריון: -0.02), כשכאמור, ההסתכלות התוך-חוגית עשויה לייצר תמונה שונה מזו המתקבלת מהשוואת ממוצעי הקבוצות מעבר לחוגים (לוח 6). הציונים הכלליים בבחינה הפסיכומטרית מוטים מעט לרעת נשים, כשממדי ההטיה קטנים באופן ניכר מאלה של ממוצע הבגרות (-0.15 בממוצע). כתוצאה מכך, ציוני הסכם מוטים לטובת הנשים (0.19 בממוצע).

בתחומי הבחינה הפסיכומטרית, ההטיה נעה בין הטיה לרעת נשים באנגלית (-0.29) לחוסר הטיה בחשיבה מילולית (0.02). הבדלים אלו במידת ההטיה של הציונים בתחומי הבחינה הפסיכומטרית, באים לידי ביטוי בהשוואה בין שלושת הציונים הכלליים בבחינה הפסיכומטרית ובין שלושת ציוני הסכם: הציון בדגש מילולי מוטה יותר לטובת נשים (בציון הסכם) או מוטה לרעתן (בציון הכללי בבחינה הפסיכומטרית).

הדפוסים שתוארו לעיל, בנוגע למידת ההטיה של החזאים השונים, מתקיימים הן בתחומי לימוד מילוליים והן בתחומי לימוד כמותיים, כאשר בתחומי הלימוד הכמותיים החזאים מוטים יותר לטובת נשים או פחות לרעתן.

משמעות ממצאי ההטיה בברירה במונחי הייצוג של קבוצת המוקד

סוגיית ההטיה בברירה עוסקת, בסופו של דבר, בייצוג של קבוצת המוקד בקרב המתקבלים. על מנת להמחיש אותה, נעשתה הדמיה, שמראה את שיעור בני קבוצת המוקד בקרב המתקבלים (כאשר הברירה מתבססת על משתנה מסוים), בהינתן שיעורה במועמדים והפער המתוקנן (d) בין ממוצעי קבוצת ההתייחסות וקבוצת המוקד במשתנה הנדון.

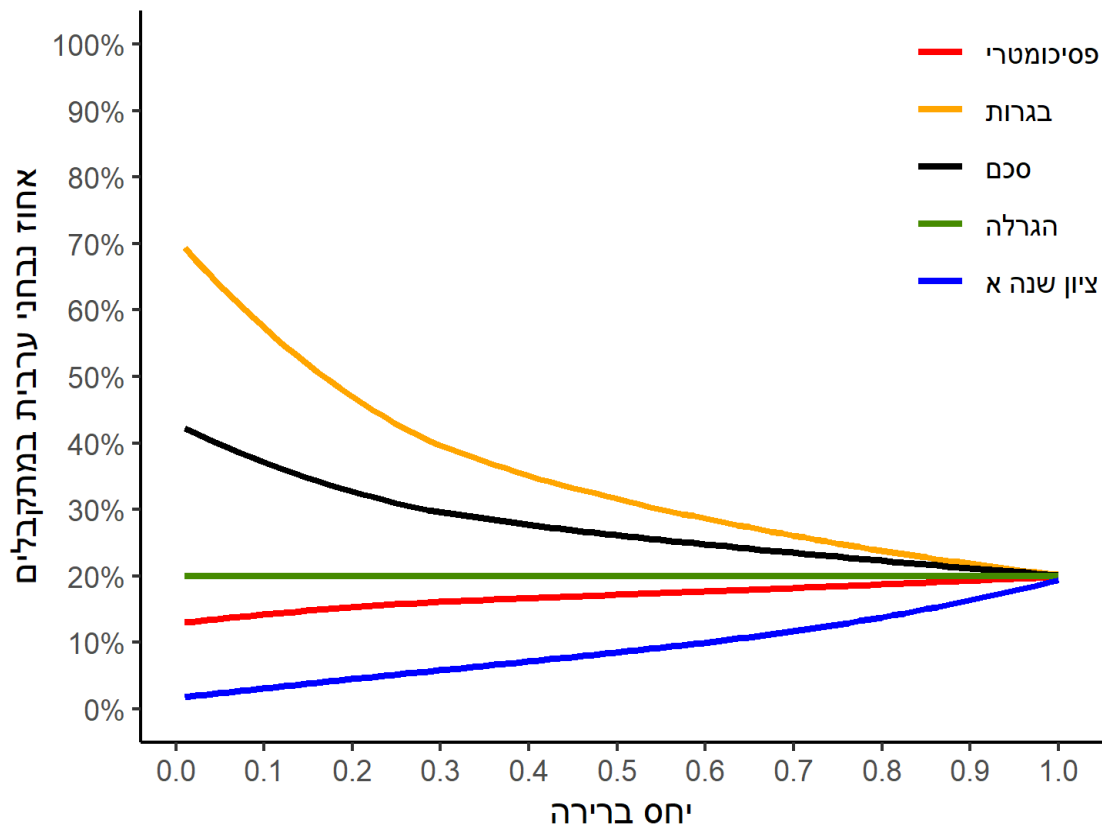
להלן יתוארו השיטה והתוצאות של ההדמיה. עבור כל אחד מ-5 משתנים – 4 חזאים (פסיכומטרי, בגרות, סכם והגרלה) וקריטריון (ציון שנה א) – הוגדרה אוכלוסייה היפותטית של מועמדים שמתפלגת נורמלית, עם שיעור נתון של בני קבוצת המוקד (על פי הלשכה המרכזית לסטטיסטיקה, 2019א), כשהפער המתוקנן (d) בין ממוצעי קבוצת ההתייחסות וקבוצת המוקד נקבע על פי הערך שנמצא עבור המשתנה התואם במחקר זה. אחרי שהוגדרה אוכלוסיית המועמדים נעשו בה תהליכי ברירה, בהתבסס על המשתנה הנדון, ביחסי ברירה שונים, וחושב שיעור בני קבוצת המוקד בקרב המתקבלים.

מידת ההטיה בברירה של החזאי משתקפת בפער שבין שיעור בני קבוצת המוקד כאשר הברירה מתבססת על פיו ובין שיעורם כאשר הברירה מתבססת על הקריטריון. זוהי, אם כן, המחשה של מדד אחד מבין שני המדדים שנבדקו במחקר זה – המדד של הפרש d-ים, שלמעשה, נוסח במקור

(Thorndike, 1971) במונחים של ייצוג קבוצתי, ושההצגה הגרפית שלו פשוטה יותר. עקרונית, ניתן להציג המחשה מקבילה גם עבור הממד של קלירי.

ההדמיה הראשונה שתוצג להלן מתייחסת לבדיקת ההוגנות ביחס לשפת היבחנות. עבור הדמיה זו נקבע אחוז נבחני ערבית בקרב המועמדים כ-20%. הפערים בין ממוצע נבחני עברית לממוצע נבחני ערבית נקבעו, בהתבסס על לוח 5, כך: פער של 0.22 סטיות תקן לטובת נבחני עברית בפסיכומטרי, פער של 0.99 סטיות תקן לטובת נבחני ערבית בבגרות, פער של 0.49 סטיות תקן לטובת נבחני ערבית בסכמ, פער של 0 בהגרלה, ופער של 0.98 סטיות תקן לטובת נבחני עברית בקריטריון. תרשים 2 מציג את תוצאות ההדמיה.

תרשים 2: אחוז נבחני ערבית בקרב המתקבלים כפונקציה של המשתנה הממין יחס הברירה



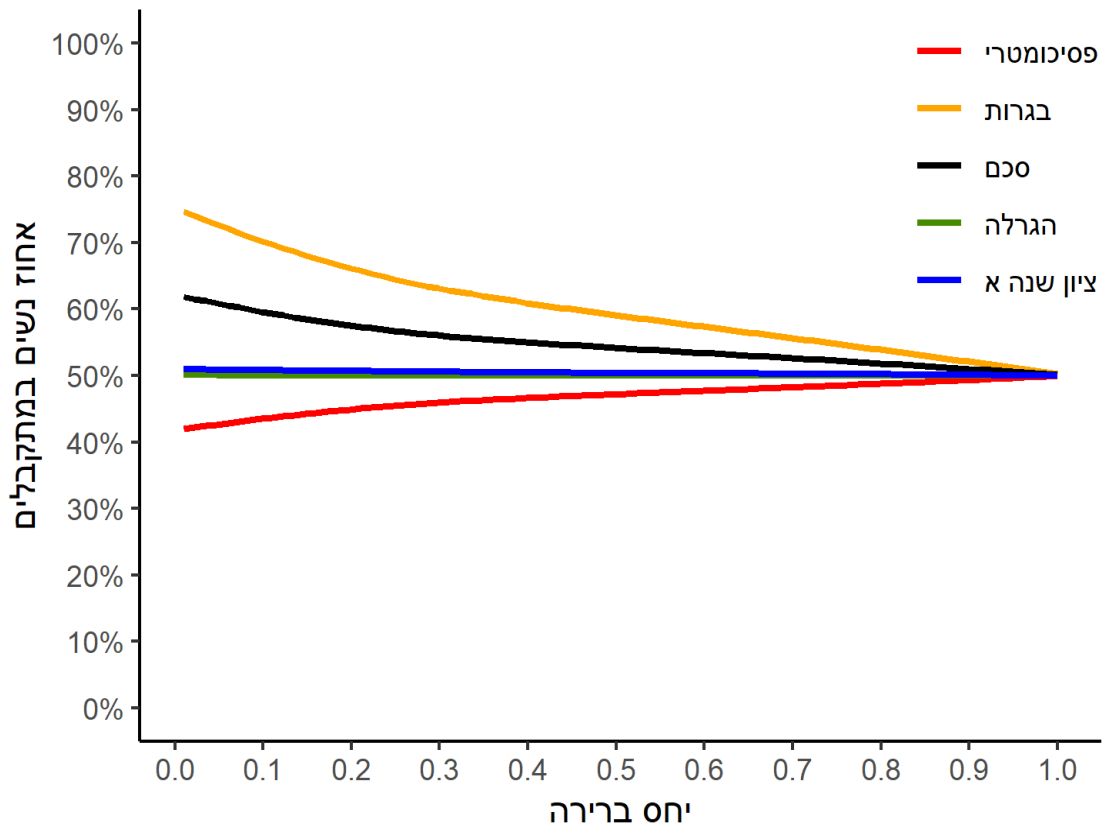
ההדמיה ממחישה שכאשר המיון מתבסס על הגרלה, אחוז נבחני ערבית בקרב המתקבלים שווה תמיד לשיעורם במועמדים. כאשר המיון מתבסס על ממוצע הבגרות או על ציון הסכמ (שבהם הפער בין הקבוצות בהסתכלות התוך-חוגית הוא לטובת נבחני ערבית), אחוז נבחני ערבית בקרב המתקבלים גבוה משיעורם במועמדים, וכאשר המיון מתבסס על הציון הפסיכומטרי או על הקריטריון (שבהם הפער בין הקבוצות בהסתכלות התוך-חוגית הוא לטובת נבחני עברית), אחוז נבחני ערבית בקרב המתקבלים נמוך משיעורם במועמדים.

הסקה על הטיה בברירה מתקבלת מהשוואה בין שיעור נבחני ערבית במתקבלים על סמך החזאי ובין שיעורם במתקבלים על סמך הקריטריון. ניתן לראות בתרשים שהעקומות שמייצגות את שיעור נבחני ערבית במתקבלים על סמך הציון הפסיכומטרי, ממוצע הבגרות או ציון הסכם נמצאות כולן מעל לעקומה שמייצגת את שיעור נבחני ערבית במתקבלים על סמך ציון שנה א'. במילים אחרות, אחוז נבחני ערבית בקבלה על סמך כל אחד מהחזאים גבוה משיעורם לו הקבלה הייתה מתבססת על הקריטריון – מצב שמשקף הטיה לטובת נבחני ערבית על ידי כל החזאים. עוד נציין שהעקומה שמייצגת קבלה על סמך הציון הפסיכומטרי היא הקרובה ביותר לעקומה שמייצגת קבלה על סמך ציון שנה א'. במילים אחרות, הציון הפסיכומטרי הוא החזאי עם ממדי ההטיה הקטנים ביותר מבין שלושת החזאים.

לבסוף, התרשים ממחיש שככל שיחס הברירה גבוה יותר, ממדי ההטיה של החזאים השונים קטנים יותר: כשיחס הברירה עולה, כל העקומות הולכות ומתכנסות לאותו ערך – אחוז נבחני ערבית במועמדים (20%), והמרחק בין העקומות שמייצגות את החזאים לעקומה שמייצגת את הקריטריון קטן.

ההדמיה השנייה מתייחסת לבדיקת ההוגנות ביחס למגדר. עבור הדמיה זו נקבע אחוז הנשים בקרב המועמדים כ-50%. הפערים בין ממוצע הגברים לממוצע הנשים נקבעו, בהתבסס על לוח 8, כך: פער של 0.14 סטיות תקן לרעת נשים בפסיכומטרי, פער של 0.46 סטיות תקן לטובת נשים בבגרות, פער של 0.21 סטיות תקן לטובת נשים בסכם, פער של 0 בהגרלה, ופער של 0.02 סטיות תקן לטובת נשים בקריטריון. תרשים 3 מציג את תוצאות ההדמיה.

תרשים 3: אחוז הנשים בקרב המתקבלים כפונקציה של המשתנה הממייין ויחס הברירה



ההטיה של ממוצע הבגרות לטובת נשים ושל הציון הפסיכומטרי לרעתן באה לידי ביטוי בתרשים 3 בעובדה שהעקומה שמייצגת את ממוצע הבגרות נמצאת מעל לעקומה של ציון שנה א', ואילו העקומה שמייצגת את הציון הפסיכומטרי נמצאת מתחתיו. התרשים ממחיש גם את ההטיה לטובה של ציון הסכמ. בהשוואה לתרשים 2 ניתן להיווכח שכלל, ממדי ההטיה ביחס לנשים קטנים יותר מאשר ביחס לנבחי ערבית (המרחקים האנכיים בין העקומות שמייצגות את החזאים לעקומה שמייצגת את ציון שנה א' קטנים יותר).

מטרתו של מחקר זה הייתה להציג ממצאים עדכניים על תוקף והוגנות מערכת המיון לאוניברסיטאות בישראל, וזאת כדי לספק ביסוס אמפירי למדיניות הקבלה ללימודים של האוניברסיטאות. המדיניות הנהוגה מבוססת, במצב של עודף ביקוש על היצע למקומות לימוד, על קבלת מועמדים שהישגיהם בלימודים צפויים להיות הגבוהים ביותר, כשניבוי הישגים הללו מתבסס על הישגים בחזאים, ללא התחשבות בשייכות קבוצתית. תוצאות המחקר תומכות במדיניות זו, כפי שיפורט להלן.

תוקף

בשאלת התוקף נמצא שתוקף הניבוי של ציון הסכם הרב-תחומי (43). גבוה מתוקף הניבוי של כל אחד משני מרכיביו בנפרד – הציון הפסיכומטרי הרב-תחומי (40) וממוצע הבגרות (35). העובדה שתוקף ציון הסכם גבוה מזה של ממוצע הבגרות מעידה שלציון הפסיכומטרי יש תרומה לניבוי הקריטריון מעבר לממוצע הבגרות. תרומה כזאת נמצאה בכל תחומי הלימוד. היתרון, במונחי תוקף ניבוי, של הצירוף של שני החזאים – הציון הפסיכומטרי וממוצע הבגרות – מתיישב עם ההמלצה של הסטנדרטים, שבמצב שבו מתקבלות החלטות חורצות גורלות, יש להסתמך על מדדים מרובים (AERA et al., 2014). יתרונו של צירוף משוקלל של שני החזאים תועד באופן עקבי בכל מחקרי התוקף במאל"ו (למשל, Oren et al., 2014), והוא מדווח גם עבור מערכות מקבילות בעולם. כך, למשל, במחקר העדכני על תוקף ה-SAT מדווח שתוקף הניבוי של ה-SAT הוא 51, של ממוצע ציוני בית הספר התיכון 53, ושל הצירוף של השניים (מתאם מרובה) – 61 (Westrick et al., 2019). בהשוואה בין ערכי התוקף המדווחים במחקר הנוכחי לבין אלו המדווחים במחקרי תוקף בארה"ב, חשוב לחדד שתי הבחנות בין המחקרים:

ראשית, התיקון לקיצוץ תחום שבו משתמשים במחקרי התוקף בארה"ב מגדיר את סטיית התקן בקרב המועמדים כסטיית התקן בקרב **אוכלוסיית הנבחנים השנתית** במבחן הכניסה (SAT או ACT). לעומת זאת בתיקון שבו משתמשים במאל"ו מוגדרת סטיית התקן בקרב המועמדים כסטיית התקן הממוצעת בקרב **הנרשמים ללימודים בחוג, בתוך מוסד ומחזור לימודים**. מטבע הדברים, אוכלוסיית הנרשמים הזאת היא תוצר של ברירה עצמית שמופעלת על ידי המועמדים שמודעים לתנאי הקבלה בחוגים ובמוסדות השונים. לפיכך, ההגדרה של אוכלוסיית המועמדים במחקרי התוקף של מאל"ו מספקת אומדן חסר לסטיית התקן בקרב המועמדים ה"אמיתיים" – אלו שהיו רוצים ללמוד בחוג ובמוסד מסוים, ללא תלות בסיכויי הקבלה שלהם אליו. אומדן החסר הזה מייצר תיקון לקיצוץ תחום שהוא שמרני מדי, ובפרט שמרני מזה שמשמש במחקרי התוקף של המערכות המקבילות בארה"ב. כדי להעריך את האפקט של השמרנות בתיקון חישבנו את תוקף הניבוי של הציון הפסיכומטרי כאשר סטיית התקן של משתנה הברירה בקרב המועמדים מחושבת כסטיית התקן בקרב **כל הנרשמים למוסד** (10), וזאת בניגוד לחישובה במחקר זה כסטיית התקן הממוצעת בקרב **הנרשמים לחוג במוסד** (8.5). תיקון זה הניב תוקפי ניבוי של 48, 45, ו-41. לציון הסכם, הציון

הפסיכומטרי וממוצע הבגרות, בהתאמה (לעומת הערכים המדווחים במחקר של 43, 40 ו-35. לשלושת החזאים, בהתאמה). זהו, כאמור, רק אומדן חלקי לאפקט של השמרנות בתיקון, כיוון שגם התיקון הליברלי יותר שהצגנו לעיל מבוסס על אוכלוסיית הנרשמים ללימודים ולא על אוכלוסיית הנבחנים במבחן הכניסה, ולפיכך נותר אפקט של ברירה עצמית (בהחלטה אם ולאיזה מוסד להירשם) שלא הובא בחשבון בתיקון. שמרנות זו בהגדרת אוכלוסיית המועמדים מקטינה את האומדנים לתוקף הניבוי בהשוואה למחקרים שנעשים בארה"ב.

הבחנה נוספת בין ממצאי המחקר הנוכחי לממצאים שמדווחים במחקרי התוקף בארה"ב נוגעת לתוקף הצירוף של מבחן הכניסה וציוני בית הספר התיכון: הממצאים המדווחים בארה"ב מתייחסים למתאם המרובה של שני מרכיבי ציון הקבלה, דהיינו, לצירוף שמבוסס על שקלול אופטימלי של מרכיביו, וזאת להבדיל מציון הסכם שמבוסס על צירוף של מרכיביו במשקלות קבועים (שווים). מאחר שאין הצדקה משכנעת להצגת מתאם שמבוסס על שקלול אופטימלי, לא דווח מתאם כזה במחקר הנוכחי, אך חשוב להביא בחשבון את העובדה שחוסר האחידות בין המחקרים פוגם בהשוואות לממצאי התוקף המדווחים בארה"ב בנוגע לצירוף של החזאים, ובפרט, שהממצאים שמדווחים בארה"ב מבוססים על אומדן הרבה פחות שמרני של התוקף.

במחקר זה נבדק גם תוקף הניבוי של שני ציונים כלליים בבחינה הפסיכומטרית שהוכנסו לשימוש באוקטובר 2011 – ציון בדגש מילולי וציון בדגש כמותי – ושל ציוני הסכם המבוססים עליהם. נמצא שבתחומי לימוד מילוליים תוקף הציון הפסיכומטרי בדגש מילולי גבוה מתוקף הציון הפסיכומטרי בדגש כמותי (וגם מתוקף הציון הפסיכומטרי הרב-תחומי), ובתחומי לימוד כמותיים (למעט בעזר רפואה) גבוה תוקף הציון הפסיכומטרי בדגש כמותי מתוקף הציון הפסיכומטרי בדגש מילולי (והוא גבוה או שווה לתוקף הציון הפסיכומטרי הרב-תחומי). דפוס דומה של יחסי מתאמים מתקיים גם ביחס לשלושת ציוני הסכם. ממצאים אלו, שהתקבלו גם במחקר קודם (Oren et al., 2014), מתקפים את ההמלצה להשתמש בציון הסכם בדגש מילולי בהחלטות הקבלה לתחומי לימוד מילוליים ובציון הסכם בדגש כמותי בהחלטות הקבלה לתחומי לימוד כמותיים. מעבר להשלכותיהם המעשיות, מספקים הממצאים הללו ראיות מתכנסות וראיות מבחינות לכך שהציונים בתחומים חשיבה מילולית וחשיבה כמותית בבחינה הפסיכומטרית, ששקלולים שונים שלהם מניבים את הציון הכללי בדגש מילולי ואת הציון הכללי בדגש כמותי, הם אמנם מדדים של היכולות המיועדות – חשיבה מילולית וחשיבה כמותית: העובדה שבתחומי לימוד מילוליים יש יתרון לציון בדגש מילולי מהווה ראיה מתכנסת לתוקף הציון בחשיבה מילולית (מתאם גבוה עם מדד אחר שמיועד למדוד אותה תכונה או תכונה דומה) וראיה מבחינה לתוקף הציון בחשיבה כמותית (מתאם נמוך יחסית עם מדד שמיועד למדוד תכונה שונה), ובאותו אופן מתקבלות ראיות מתכנסות וראיות מבחינות מן המתאמים של שני הציונים בתחומי לימוד כמותיים.

לקורא שאינו מצוי במחקרים על תוקף הניבוי של כלי מיון, יש קושי להעריך את גודלם של הערכים שדווחו במחקר זה – האם מתאם של 40. הוא גבוה? נמוך? לעניין זה אפשר להציג כלל אצבע מקובל לפירוש מקדמי מתאם (בערכים מוחלטים): מתאם נמוך מ-10. נחשב זניח, מתאם בטווח

של 10-29. נחשב נמוך, מתאם בטווח של 30-49. נחשב בינוני, ומתאם של 50. ומעלה נחשב גבוה (Cohen, 1988). לפיכך, על פי כלל האצבע הזה, הערכים המדווחים במחקר זה נחשבים לבינוניים. לחילופין, אפשר לבחון קשרים בין משתנים בעולם האמיתי כאמות מידה ריאליסטיות (Rosenthal, 1995, 1990). כך, למשל, המתאם הטיפוסי בין לקיחת אספירין ובין סכנה מופחתת למוות מהתקף לב הוא 0.02. המתאם הטיפוסי בין ציון בראיון קבלה לעבודה להצלחה בעבודה הוא 0.20, והמתאם בין מגדר למשקל הוא 0.26 (Meyer et al., 2001). לבסוף, כדי להמחיש את משמעותם המעשית של הערכים שהתקבלו במחקר הצגנו, באמצעות הדמיה, את תרומתו של חזאי עם תוקף ניבוי נתון לממוצע הציונים בקריטריון בקרב המתקבלים על פי החזאי, תחת יחסי ברירה שונים. הממצאים שהוצגו ממחישים איך, ככל שתוקף הניבוי של החזאי גבוה יותר, כך גדולה יותר תועלת המיון. בנוסף, הרווח העולה משיפור בתוקף עולה ככל שיחס הברירה – שיעור המתקבלים מבין המועמדים – נמוך יותר, תופעה שהומחשה בסדרת טבלאות שפרסמו טיילור וראסל (Taylor & Russel, 1939). בהינתן הערכים הטיפוסיים של תוקף ניבוי שנמצאו במחקר זה, אזי גם ביחסי הברירה הרווחים (55%-88%, ראו: הלשכה המרכזית לסטטיסטיקה, 2019ב), שימוש בציון הסכם מעלה את ממוצע ציוני הקריטריון של המתקבלים ב-0.10-0.31 סטיות תקן לעומת קבלה על פי חזאי שתוקפו 0 (הגרלה). לשם השוואה, מיון על סמך חזאי שתוקפו מושלם היה מעלה את ממוצע ציוני הקריטריון של המתקבלים ב-0.23-0.72 סטיות תקן. ציון הסכם, שתוקף הניבוי שלו 0.43, מאפשר, אם כן, להגיע ל-43% ממה שניתן היה להשיג עם כלי מיון בעל תוקף מושלם. לבסוף, מכיוון שתועלת השימוש בכלי מיון גבוהה יותר ככול שיחס הברירה נמוך יותר (ראו תרשים 1), אזי החוגים והמוסדות אשר מרוויחים יותר משימוש בכלי המיון הינם הסלקטיביים ביותר.

הוגנות

בשאלת ההוגנות נמצא שהחזאי המרכזי של מערכת המיון, ציון הסכם, מוטה מאוד לטובת נבחני ערבית ומוטה מעט לטובת נשים. ניתן לסכם, לפיכך, וזאת בהתייחס לשתי קבוצות המוקד שנבדקו במחקר זה, שתהליך המיון אינו פוגע (בהתייחס לנשים) ובמידה רבה אף מעלה (בהתייחס לנבחני ערבית) את הנגישות של קבוצות אלה להשכלה הגבוהה. ממצא זה התקבל על פי שני המדדים להטיה בברירה, מה שמעיד על איתנותו. אשר לחזאים הנוספים של מערכת המיון, נמצא, בהתייחס לנבחני ערבית, שכלל החזאים מוטים לטובתם, וזאת על פי שני המדדים להטיה: ממוצע הבגרות מוטה לטובתם במידה רבה מאוד והציון הכללי בבחינה הפסיכומטרית מוטה לטובתם במידה פחותה. בהתייחס לנשים, התמונה שמציגים שני המדדים להטיה בברירה אינה זהה: על פי המדד של קלירי ממדי ההטיה זניחים ביחס לכל החזאים. על פי המדד של הפרש d-ים, ממוצע הבגרות מוטה לטובת נשים במידה בינונית, ואילו הציון הכללי בבחינה הפסיכומטרית מוטה במידה מועטה לרעתן. כתוצאה, החזאי הממייין, ציון הסכם, נמצא מוטה מעט לטובת הנשים. משמעותם של ממצאי ההטיה על פי המדד של הפרש d-ים הומחשה, באמצעות הדמיה, כייצוג-יתר (הטיה לטובה) או ייצוג-חסר (הטיה לרעה) של חברי קבוצת המוקד כאשר הקבלה מתבססת על החזאי הנדון בהשוואה לייצוגם כאשר הקבלה מתבססת על הקריטריון. גם את המדד של קלירי ניתן לתרגם לייצוג-יתר או ייצוג-חסר

בהשוואה לייצוג ה"נכון", אך במקרה זה הייצוג ה"נכון" משתנה מחזאי לחזאי, ולכן ההצגה הגרפית מורכבת יותר.

חקירת ההוגנות כללה גם בדיקה של תוקף דיפרנציאלי. בעניין זה נמצא, בהתייחס לשפת היבחנות, שתוקף הניבוי של כל החזאים גבוה יותר בקרב נבחני עברית. בהתייחס למגדר נמצא שההבדלים בתקפים בין גברים לנשים היו קטנים ולא עקביים בכיוונם.

הממצאים שתוארו לעיל מתיישבים עם ממצאי המחקרים שנעשו בעבר במאל"ו (לפירוט על המחקר האחרון וסקירת ממצאי מחקרים קודמים, ראו: קנת-כהן ואחרים, 2011), ובמידה רבה גם עם המחקרים על הוגנות כלי המיון בארה"ב. בהתייחס לאחרונים (שבהם, כאמור, בדיקת ההטיה בברירה מתבססת רק על המדד של קלירי) הממצא הטיפוסי המתקבל הוא הטיה לטובת קבוצות מיעוט אתניות ולרעת נשים. בהתייחס לשפת היבחנות, תוצאות המחקר הנוכחי מתיישבות, אם כן, במלואן עם ממצאים מקבילים בארה"ב. בהתייחס למגדר, תוצאות המחקר הנוכחי מתיישבות באופן חלקי עם הממצאים בארה"ב: על פי המדד של קלירי ממדי ההטיה במחקר הנוכחי היו זניחים, ועל פי המדד של הפרש d-ים, ההטיה של ממוצע הבגרות הייתה לטובת נשים. ראוי לציין שגם בארה"ב מוצאים באופן עקבי (למשל, Sanchez, 2013; Marini et al., 2019) שמבין שני מרכיבי מערכת המיון – מבחן הכניסה הסטנדרטי וציוני בית הספר התיכון – ממדי ניבוי היתר כלפי קבוצות מיעוט אתניות הם גדולים יותר וממדי ניבוי החסר כלפי נשים הם קטנים יותר כאשר החזאי הוא ציוני בית הספר התיכון מאשר כאשר הוא מבחן הכניסה. תמונה זו תואמת את ממצאי המחקר הנוכחי.

מה מקורם של הממצאים שהתקבלו בשני מחקרי ההוגנות? בהתייחס לשפת היבחנות, חשוב לזכור שבעוד שכלי המיון בודקים את היכולת הלימודית בשפת האם של נבחני ערבית, ההישגים בלימודים עצמם נמדדים בעברית, שבה רמת השליטה שלהם נמוכה יותר. אין זה מפתיע, לפיכך, למצוא שכלי המיון מוטים לטובת נבחני ערבית. תמיכה להסבר זה ניתן לקבל מן הממצא, שהתקבל במחקר זה, שבתחומי הלימוד המילוליים מידת ההטיה לטובת נבחני ערבית גדולה יותר מאשר בתחומי לימוד כמותיים. במילים אחרות, ככל שהקריטריון טעון יותר בשפה, ההטיה של כלי המיון לטובה חזקה יותר. בנוסף לגורם של שפת היבחנות, שרלוונטי למחקר הנוכחי, נציין שרבים הציעו וחקרו סיבות לממצא הרווח של הטיה לטובת קבוצות מיעוט אתניות בארה"ב, שעשויות להיות רלוונטיות גם לנבחני ערבית במחקר זה. המסקנה הכוללת מהעבודה בתחום היא שיש צירוף של גורמים שעשויים להביא לממצא של הטיה לטובה ושקשה לבודד את המקורות המדויקים. בין הגורמים שהוצעו: הבדלים בין הקבוצות במשתנים נוספים (איכות בתי הספר, תמיכה של הסביבה הביתית) שעשויים להשפיע על ההצלחה בלימודים, הבדלים בחוויית הלימודים באוניברסיטה (יותר קשיי הסתגלות ותחושה של זרות לקבוצות מיעוט אתניות), הבדלים בקורסים שנלקחים, ועוד גורמים רבים שנעים מטעות מדידה ועד למושגים מתחום הפסיכולוגיה החברתית, כמו איום הסטריאוטיפ (Zwick, 2002, 2019a).

בהתייחס למגדר, חשוב לחזור ולהדגיש שהאפקטים שנמצאו הם קטנים מאוד. עם זאת, כאמור, מגמת הממצאים שהתקבלו ביחס לחזאים השונים, מתיישבת עם הידע שנצבר על אודות הבדלי

מגדר במדדי יכולת לימודית: בסקירתם המקיפה על הבדלי ביצוע בין גברים לנשים, דיווחו ווילינגהם וקול (Willingham & Cole, 1997) על הישגים ממוצעים גבוהים יותר של נשים בציוני בית ספר ועל הישגים ממוצעים גבוהים יותר של גברים במבחנים מתוקננים. את הממצאים על הטיה ברירה (על פי המדד של הפרש d-ים; כאמור, על פי המדד של קלירי ממדי ההטיה זניחים) ניתן להסביר בהינתן אופיו של הקריטריון. בהשוואה למבחנים מתוקננים, הקריטריון מבוסס גם על מבחנים פתוחים ולא רק על מטלות של ברירה, ובשונה מציוני בית ספר הוא מתבסס בעיקר על מבחנים (או עבודות) ואינו כולל מרכיבי הערכה סובייקטיביים, כדוגמת מוטיבציה והשתתפות פעילה. הממצאים שהתקבלו ביחס לקריטריון זה משקפים את היותו תוצר של מדידה משולבת, הן זו המאפיינת ציוני בית ספר והן זו המאפיינת ציונים במבחנים מתוקננים, ומספקים בכך תמיכה ליתרון הטמון בשימוש בצירוף של הבחינה הפסיכומטרית וממוצע הבגרות.

לבסוף, מן הראוי להתייחס לדמיון ולשוני בין שני המדדים להטיה ברירה. כפי שנאמר קודם, הגישה של הפרש d-ים נוטה יותר להסקה על הטיה לרעת הקבוצה החלשה בחזאי ופחות להסקה על הטיה לטובתה בהשוואה למדד של קלירי, ולכן במצב השכיח שבו קבוצת המוקד היא הקבוצה החלשה יותר, הגישה של הפרש d-ים נוטה יותר להסקה על הטיה לרעתה ופחות להסקה על הטיה לטובתה. במובן זה, הגישה של הפרש d-ים נתפסת כמזוהה יותר עם האינטרס החברתי של מתן ייצוג לקבוצות חברתיות חלשות בקרב המתקבלים, כיוון שמצב נתון נתפס על פיה כיותר פוגעני כלפי קבוצות חלשות בהשוואה לגישה של קלירי. הביטוי הטכני להבדל בין הגישות הוא בכך ששתי הגישות בודקות את הפער בין ההישג בפועל להישג המנובא, אך בעוד שהמדד של קלירי מביא בחשבון את תופעת הרגרסיה לממוצע (כתוצאה מהמתאם הלא-מושלם בין החזאי לקריטריון) בחישוב ההישג המנובא, המדד של הפרש d-ים – לא.

התוצאות שהתקבלו במחקר זה מתיישבות באופן כללי עם הציפייה שתוארה לעיל בנוגע למסקנות של שני המדדים. הסבר אפשרי למקרים שבהם הן אינן מתיישבות עם ציפייה זו נעוץ בעובדה שיש הבדלים באופרציונליזציה של שני המדדים במחקר זה: המדד של קלירי לא תוקן לקיצוץ תחום בעוד שהמדד של הפרש d-ים – כן. חוסר התיקון של המדד של קלירי קשור לעובדה שבאף אחד מאין ספור היישומים שלו לא נעשה תיקון כזה. עקרונית, התוצאות הלא מתוקנות אינן מוטות כאשר החזאי שנבדק הוא משתנה הברירה (ציון הסכם הרב-תחומי, במקרה שלנו) אך הן מוטות כאשר החזאי הנבדק הוא משתנה שעבר ברירה עקיפה (Linn, 1983). יש מקום למחקר המשך שיבדוק את תוצאות היישום של המדד של קלירי עם תיקון לקיצוץ תחום. אשר לתיקון לקיצוץ תחום של המדד של הפרש d-ים, נציין בעיה חוזרת ונשנית בתיקון של מדד זה, והיא העובדה שנוסחת התיקון מחייבת שימוש בפרופורציות של שתי הקבוצות – קבוצת המוקד וקבוצת ההתייחסות – בקרב המועמדים (Bobko, Roth, & Bobko, 2001). מאחר שערכים אלה לא היו זמינים לנו, השתמשנו בנוסחת התיקון בפרופורציות של הקבוצות בקרב הלומדים. עם זאת, בדיקה שעשינו בנתוני הלשכה המרכזית לסטטיסטיקה מלמדת שהפרופורציות של הקבוצות במועמדים ובמתקבלים דומות למדי (הלשכה המרכזית לסטטיסטיקה, 2019א). לפיכך, ההחלטה להשתמש בנתוני הלומדים לצורך חישוב הפרופורציות של הקבוצות לא אמורה ליצור עיוות משמעותי בתוצאות.

על רקע הדיון שלעיל בהבדלים – המהותיים והטכניים – בין שני המדדים להטיה בברירה, מתקבל מידע רב ערך בנוגע לאיתנות של הממצאים במצבים שבהם מוצאים עקביות בין תוצאות שני המדדים ובמצבים שלא. בחינה משולבת של מסקנות שני המדדים ביחס לשתי קבוצות המוקד מלמדת שבהקשר של שפת היבחנות יש הטיה ניכרת של כל החזאים לטובת נבחני ערבית, ובהקשר של מגדר ממדי ההטיה חלשים יותר. בשורה התחתונה, בהתייחס למערכת המיון כולה, שמתבססת על הצירוף של ממוצע הבגרות והציון הפסיכומטרי, ניתן לסכם שהיא מוטה – במידה כזו או אחרת – לטובת נבחני ערבית ולטובת נשים. הממצאים בעניין זה, שכיוונם וממדיהם הם תוצר של האימפקט המשולב של שני החזאים, חוזרים ומחזקים את חשיבות השימוש בחזאים מרובים בתהליך הברירה להשכלה הגבוהה.

מקורות

- אורן, כ', קנת-כהן, ת' וברונר, ש' (2007). **נתונים מקובצים על תוקף מערכת המיון לאוניברסיטאות בחיזוי הצלחה בלימודי שנה א' (מחזורים תשס"ג – תשס"ה)** (דוח מס' 342). ירושלים: מרכז ארצי לבחינות ולהערכה.
- בלר, מ' ובן-שחר, ג' (1983). על הוגנות השימוש במבחנים פסיכולוגיים. **מגמות, כח**, 42-56.
- גולדצויג, א' (2015). **סקירת מערכות מיון להשכלה הגבוהה ברחבי העולם** (דוח מס' 417). ירושלים: מרכז ארצי לבחינות ולהערכה.
- גפני, נ' (2014). **סיכום של שנה לרפורמה בבחינה הפסיכומטרית – הוספת מטלת כתיבה ועוד**. מאמר שהוצג בכינוס אפ"י – אגודה ישראלית לפסיכומטריקה, וואהל סנטר, רמת גן.
- הלשכה המרכזית לסטטיסטיקה (2019א). **מועמדים ללימודים במוסדות להשכלה גבוהה, תשס"ח-תשע"ט** (2018/19-2007/8). נדלה מתוך <https://bit.ly/3rU1aPE>
- הלשכה המרכזית לסטטיסטיקה (2019ב). **שנתון סטטיסטי לישראל 2019 – מספר 70**. נדלה מתוך <https://bit.ly/3pMK7gE>
- טשנר, נ' (2013). **תנאי קבלה למוסדות להשכלה גבוהה - סקירה משווה**. ירושלים: מרכז המחקר והמידע, הכנסת.
- כהן, י' (2012). **מעבר ל"אחוז שונות מוסברת" – מתן פשר למקדם התוקף**. מאמר שהוצג בכינוס אפ"י – אגודה ישראלית לפסיכומטריקה, מרכז "נע לגעת", נמל יפו.
- מאל"ו (2020). **הבחינה הפסיכומטרית – דוח רב שנתי 2019-2005**. נדלה מתוך <https://www.nite.org.il/research-and-publications/statistical-data/>
- קלפר, ד', טורוול, א' ואורן, כ' (2014). **תוקף הניבוי של כלי המיון לאוניברסיטאות בישראל מול ממוצע תואר בוגר** (דוח מס' 403). ירושלים: מרכז ארצי לבחינות ולהערכה.
- קנת-כהן, ת', ברונר, ש' ואורן, כ' (1999). ניתוח-על של תוקף הניבוי של מרכיבי מערכת המיון לאוניברסיטאות בישראל כלפי מידת ההצלחה בלימודים. **מגמות, מ**, 54-71.
- קנת-כהן, ת', טורוול, א' ואורן, כ' (2011). **הוגנות מערכת המיון לאוניברסיטאות לפי שפת היבחנות במכפ"ל (ערבית או עברית) ומגדר** (דוח מס' 376). ירושלים: מרכז ארצי לבחינות ולהערכה.
- AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education) (2014). *Standards for educational and psychological testing*. Washington, DC: Author.
- Allalouf, A., Cohen, Y. & Gafni, N. (2020). Higher education admissions practices in Israel. In M. E. Oliveri & C. Wendler (Eds.), *Higher education admissions practices: An international perspective* (pp. 174-189). Cambridge: Cambridge University Press.
- Beatty, A. S., Walmsley, P. T., Sackett, P. R., Kuncel, N. R., & Koch, A. J. (2015). The reliability of college grades. *Educational Measurement: Issues and Practice*, 34, 31–40.
- Bobko, P., Roth, P. L., & Bobko, C. (2001). Correcting the effect size of d for range restriction and unreliability. *Organizational Research Methods*, 4, 46-61.

- Brogden, H. E. (1946). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, 37, 65-76.
- Cahan, S., & Gamliel, E. (2001). Prediction bias and selection bias: An empirical analysis. *Applied Measurement in Education*, 14, 109-123.
- Cahan, S., & Gamliel, E. (2006). Definition and measurement of selection bias: From constant ratio to constant difference. *Journal of Educational Measurement*, 43, 131-144.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38, 369–382.
- Darlington, R. B. (1971). Another look at "culture fairness". *Journal of Educational Measurement*, 8, 71-81.
- Gafni, N., Beller, M., & Bronner, S. (2000). *A cross-cultural perspective on gender differences in higher education: Admissions and scholastic achievement* (Report No. 274). Jerusalem: National Institute for Testing and Evaluation.
- Gulliksen, H. (1987). *Theory of mental tests*. Hillsdale, NJ: Erlbaum. (Original work published 1950)
- Kennet-Cohen, T., Bronner, S., & Oren, C. (1999). *The predictive validity of the components of the process of selection of candidates for higher education in Israel* (Report No. 264). Jerusalem: National Institute for Testing and Evaluation.
- Kennet-Cohen, T., Turvall, E., & Oren, C. (2014). Detecting bias in selection for higher education: Three different methods. *Assessment in Education: Principles, Policy & Practice*, 21, 193-204.
- Linn, R. L. (1983). Pearson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement*, 20, 1-15.
- Marini, J. P., Westrick, P. A., Young, L., NG, H., Shmueli, D., & Shaw, E. J. (2019). *Differential validity and prediction of the SAT*. (College Board Statistical Report). New York: The College Board.
- Meyer, J. M., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56, 128-165.
- Oren, C., Kennet-Cohen, T., Turvall, E., & Allalouf, A. (2014). Demonstrating the validity of three general scores of PET in predicting higher education achievement in Israel. *Psicothema*, 26, 117-126.

- Petersen, N. S., & Novick, M. R. (1976). An evaluation of some models for culture-fair selection. *Journal of Educational Measurement, 13*, 3-29.
- Rosenthal, R. (1990). How are we doing in soft psychology? *American Psychologist, 45*, 775-777.
- Rosenthal, R. (1995). Progress in clinical psychology: Is there any? *Clinical Psychology: Science and Practice, 2*, 133-150.
- Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology, 74*, 166-169.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63*, 215-227.
- Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112-118.
- Sanchez, E. (2013). *Differential effects of using ACT® College Readiness Assessment scores and high school GPA to predict first-year college GPA among racial/ethnic, gender, and income groups*. (ACT Research Report No. 2013-4). Iowa City, IA: ACT.
- Schrader, W. B. (1965). A taxonomy of expectancy tables. *Journal of Educational Measurement, 2*, 29-35.
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection. *Journal of Applied Psychology, 23*, 565-578.
- Thorndike, R. L. (1971). Concepts of culture-fairness. *Journal of Educational Measurement, 8*, 63-70.
- Turvall, E., Bronner, S., Kennet-Cohen, T., & Oren, C. (2008). *Fairness in the higher education admissions procedure: The Psychometric Entrance Test in Arabic* (Report No. 349). Jerusalem: National Institute for Testing and Evaluation.
- Westrick, P. A. (2017). Reliability estimates for undergraduate grade point average. *Educational Assessment, 22*, 231–252.
- Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M. R., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT scores, high school grades, and SES. *Educational Assessment, 20*, 23-45.
- Westrick, P. A., Marini, J. P., Young, L., NG, H., Shmueli, D., & Shaw, E. J. (2019). *Validity of the SAT for predicting first-year grades and retention to the second year*. (College Board Statistical Report). New York: The College Board.
- Willingham, W. W., & Cole, N. S. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

Wikström, C., & Wikström, M. (2020). Merit-based admissions in higher education. In M. E. Oliveri & C. Wendler (Eds.), *Higher education admissions practices: An international perspective* (pp. 34-50). Cambridge: Cambridge University Press.

Zwick, R. (2002). *Fair game? The use of standardized admissions tests in higher education*. New York: RoutledgeFalmer.

Zwick, R. (2006). Higher education admissions testing. In R. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 647-679). Westport, CT: American Council on Education, Praeger.

Zwick, R. (2019a). Assessment in American higher education: The role of admission tests. *The ANNALS of the American Academy of Political and Social Science*, 683, 130–148.

Zwick, R. (2019b). Fairness in measurement and selection: Statistical, philosophical, and public perspectives. *Educational Measurement: Issues and Practice*, 38, 31–40.

נספחים

נספח 1: הרכב החוגים בתחומי הלימוד

להלן דוגמאות לחוגי הלימוד בתחומי הלימוד השונים:

רוח ואמנויות – ארכיאולוגיה, היסטוריה, לימודי האסלאם, לימודי מזרח אסיה, מקרא, ספרות, אנגלית, פילוסופיה, תולדות האמנות, קולנוע וטלוויזיה

משפטים – משפטים

חברה מילולי – מדע המדינה, יחסים בין-לאומיים, חינוך, עבודה סוציאלית, סוציולוגיה, פסיכולוגיה, תקשורת, גאוגרפיה

חברה כמותי – כלכלה, חשבונאות, מנהל עסקים

מדעים ביולוגיים – ביולוגיה, לימודי הסביבה, תזונה

מדעים פיסיקליים – פיסיקה, כימיה, מדעי כדור הארץ

מדעים מדויקים – מתמטיקה, סטטיסטיקה, מדעי המחשב

הנדסה ואדריכלות – הנדסת חשמל, הנדסת מכונות, הנדסת מחשבים, ארכיטקטורה

רפואה – רפואה, רפואת שיניים

עזר רפואה – רוקחות, סיעוד, ריפוי בעיסוק, הפרעות בתקשורת, פיזיותרפיה

נספח 2: מקדמי מתאם נצפים

בלוח 9 מוצגים מקדמי המתאם הנצפים בין החזאים לקריטריון.

לוח 9: מתאמי פירסון נצפים בין החזאים לקריטריון

ציוני הבחינה הפסיכומטרית						בגרות	ציוני סכמ			תחום לימודים
ציונים בתחומי הבחינה			ציון פסיכומטרי כללי				ר-ב תחומי	בדגש מילולי	בדגש כמותי	
אנגלית	כמותי	מילולי	בדגש כמותי	בדגש מילולי	ר-ב תחומי					
.31	.30	.36	.38	.40	.40	.37	.44	.45	.45	רוח ואמנויות
.21	.21	.30	.29	.32	.31	.22	.32	.34	.33	חברה מילולי
.15	.07	.23	.16	.24	.21	.11	.18	.23	.21	משפטים
.11	.20	.16	.24	.21	.24	.20	.29	.28	.29	חברה כמותי
.20	.30	.21	.33	.28	.31	.22	.36	.33	.35	מדעים ביולוגיים
.22	.33	.29	.38	.34	.37	.32	.44	.41	.43	מדעים פיסיקליים
.14	.29	.16	.31	.23	.28	.20	.35	.30	.33	מדעים מדויקים
.11	.18	.12	.21	.16	.19	.19	.28	.25	.27	הנדסה ואדריכלות
.09	.13	.08	.17	.13	.16	.14	.19	.17	.18	רפואה
.17	.10	.16	.17	.20	.19	.13	.19	.20	.20	עזר רפואה
.17	.21	.22	.27	.26	.27	.21	.31	.30	.31	סך הכול

RR-21-05
מאי 2021