



מבחן כניסה ארצי  
לפסיכולוגיה  
(מכא"ל)  
כדגם לבנית מבחן  
הישגים ארצי

כרמל אורן

## תוכן העניינים

### ה ק ד מ ה

1. רקע כללי והסטורי
2. ועדת מכא"ל
3. בניית המבחן
4. תכונות, תוכן ומבנה המבחן
5. ניתוח המבחן
6. הטיית פריטים ופסילתם
7. הבדלים בין קבוצות
8. מהימנות, נבחנים חוזרים, יציבות האוכלוסיות, תקפות
9. השוואה עם פרויקט מקביל - G.R.E. - פסיכולוגיה
10. סיכום ודיון
11. נספח - התפלגויות משתני הרקע של אוכלוסיות נבחני מכא"ל

### מ ק ו ר ו ת

פרק זה מתאר פרוייקט של מיון כלל-ארצי ללימודים אקדמיים מתקדמים בפסיכולוגיה, באמצעות מבחן הישגים (ידע) בתחומי הפסיכולוגיה. המבחן והפרוייקט עשויים לשמש מודלים עבור תחומים אחרים, כגון מיון ללימודים מתקדמים או הסמכה במקצועות שונים כגון רפואה, משפטים וכד'.

הפרוייקט מואר בפרק זה מזוויות רבות: עקרוניות, פסיכומטריות, טכניות, מתודולוגיות, השוואתיות, ואפילו מוסריות-ערכיות.

הפרק כתוב ברמות שונות. חלק מהקוראים עשוי למצוא עניין רק בחלק מן הכתוב. הכוונה בעיקר לאנשים שאינם קרובים אל מערכת הלימודים המתקדמים בפסיכולוגיה. אולם קוראים כאלה עשויים להפיק הנאה ותועלת מקריאה ברנית, תוך התמקדות בנושאים העקרוניים יותר - הכלל לפי תחומי העניין האישיים.

תודת מחבר הפרק נתונה לגב' רות ריזל, על עזרתה הרבה בהפקת המיצאים, ולגב' ג'קי בטל, על תרומתה בנושא מעקב התוקף.

הפרק פותח בסקירת הרקע הכללי וההסטורי להקמת מכא"ל, תוך תאור העקרונות אשר מנחים את הפרוייקט עד היום, ודיון ביתרונות ובחסרונות של מכא"ל, לאור עקרונות אלו.

בהמשך מתוארים: ועדת מכא"ל וסמכויותיה, תהליך בניית המבחן, וכן התכונות, התוכן, והמבנה של המבחן.

לאחר מכן, מתוארים הכלים המשמשים לניתוח המבחן. פיסקה נרחבת ומפורטת מוקדשת לנושא הטיית פריטים, ותאור הנושא הבעייתי והייחודי של פסילת הפריטים במכא"ל.

בהמשך מוצגות מספר השוואות בין קבוצות שונות של נבחנים, וכן הקשר בין ההישגים במכא"ל לבין מספר משתני רקע. הנתונים הסטטיסטיים והטכניים של המבחן מתוארים בפיסקה אודות המהימנות, התקפות, נושא הנבחנים החוזרים, ויציבות האוכלוסיות מעבר לשנים. לאחר מכן מוצגת השוואה של מכא"ל עם פרוייקט מקביל לו בארה"ב, על פני מספר מימדים מרכזיים.

בסיום הפרק, מובא דיון מקיף בבעייתיות שמקורה בעקרונות הפרוייקט בכלל, בפרט בדרך בה מוגדרים תחומי התוכן במבחן. בדיון זה מוצגים גם יתרונותיו ומעלותיו של מכא"ל. אוכלוסיית נבחני מכא"ל מתוארת בפיסקה האחרונה, בנספח, ובה פירוט התפלגויות משתני הרקע הדמוגרפיים של הנבחנים.

הדעות והעמדות המובעות בפרק זה, אינן מייצגות אלא את כותבן.

מחבר הפרק משמש כראש מדור מחקר במרכז הארצי לבחינות ולהערכה, ומאז 1983 - כרכז מכא"ל.

## 1. רקע כללי והסטורי

ב-1981 היו בארץ ארבע אוניברסיטאות אשר קיימו מסלול לימודים לתואר שני בפסיכולוגיה: אונ' בר-אילן, אונ' חיפה, האונ' העברית בירושלים ואונ' תל-אביב. באותה השנה, הגיעו המחלקות לפסיכולוגיה באותן אוניברסיטאות להסכמה בדבר הקמת פרויקט מיון משותף למסלולי התואר השני שהן הציעו. הפרוייקט כונה "מכא"ל" שהן ראשי התיבות של "מבחן כניסה ארצי לפסיכולוגיה".

זהו מבחן הישגים הבודק ידע בתחומי הפסיכולוגיה הנלמדים במהלך התואר הראשון באוניברסיטאות המשתתפות בפרוייקט.

ב-1983 הצטרפה לפרוייקט אוניברסיטת בן-גוריון. ב-1986 הצטרף הטכניון כשותף "סביל" (שאינו תורם שאלות למבחן, ואינו משפיע על תוכנו) - לצורך מיון למסלול הפסיכולוגיה התעשייתית בפקולטה לתעשייה וניהול. הסיבה ל"סבילותו" של הטכניון, הינה העובדה שלא מתקיימים בו לימודי תואר ראשון בפסיכולוגיה, ולכן אין הוא משפיע על הרכב הבחינה.

מאז 1981 מועברת הבחינה פעם אחת בשנה, במועד ראשי (בד"כ באפריל), וכן במועד מיוחד, כשבועיים עד חודש לאחר מכן, עבור מקרים חריגים ספורים. מספר הנבחנים השנתי הוא כ-270 (ראה התפלגות מפורטת בלוח 11.1).

הצורך בקיום המבחן נובע באופן בסיסי מעודף ביקוש על היצע של מקומות לימוד בתוכניות ה"מוסמך", בעיקר במגמות הקליניות. עודף ביקוש זה מחייב את המחלקות לפסיכולוגיה לבצע ברירה של מועמדים.

הסלקטיביות הרבה, וכן השונות הקיימת בין המחלקות השונות מבחינת הדגשים האקדמיים שלהן ומסלולי ההתמחות שהן מציעות - מעודדות בקרב בוגרי התואר הראשון בפסיכולוגיה ניידות אקדמית. הניידות מתבטאת בהגשת מועמדות ללימודי תואר שני במחלקות אחרות או נוספות על זו שבה הם למדו: אם מתוך רצון להגדיל את סיכויי הקבלה, אם מתוך בחירה במסלולים/מורים עדיפים בעיניהם. תופעה זו הינה אולי חיובית כשלעצמה, אולם היא יוצרת את הצורך בקריטריון קבלה אחיד, סטנדרטי ואובייקטיבי.

בתהליך הברירה משתמשות המחלקות במדדים שונים נוסף על מכא"ל: ציוני ה-B.A., המלצות, ראיונות אישיים וכדומה. אולם מכא"ל הינו המכשיר היחיד הנהנה מתכונות האובייקטיביות והסטנדרטיות, והוא מספק סולם אחיד לדרוג מועמדים מרקע אקדמי שונה.

הקמת הפרוייקט לוותה בתשבות מצד המחלקות מפני פירסומם של נתונים השוואתיים בין המחלקות. אלו נתפסו כהשוואות העלולות לפגוע במוניטין, ביוקרה או במידת המשיכה של חלק מהמחלקות. כמו כן, הובעו תשבות מפני פגיעה אפשרית בבוגרי מחלקה כלשהי, כתוצאה מאי-הכללת חומר לימודים כלשהו בתכנית הלימודים שלהם, בעוד הוא כלול בחומר הלימודים של מחלקות אחרות. על כן, תקנון הקמת מכא"ל כלל בין היתר שני תנאי יסוד להקמת הפרוייקט:

- א. לא יפורסמו מתקרים המשווים בין המחלקות.
- ב. שני תהליכים ילוו את בניית המבחן וחישוב הציונים: הראשון - ביקורת מוקדמת של מורי המחלקות על השאלות המוצעות. ביקורת זו עשויה לפסול שאלות ממכא"ל במקרים בהם החומר הנשאל אינו כלול בתכנית לימודי החובה, ולו של מחלקה אחת בלבד.
- התהליך השני, הינו בדיקה פוסט-הוק, להשוואת הביצוע בכל פריט מאלה שאושרו בבדיקה הראשונה, בין בוגרי האוניברסיטאות השונות החברות במכא"ל. יוגדרו מדדים להטייה של פריטים על-פי מוסדות, ולפי מדדים אלה יפסלו הפריטים המוטים לאחר מעשה ולא יכללו בחישוב הציון במכא"ל.

## 2. ועדת מכא"ל

האחריות לניהול המבחן ולאישור הרכבו הסופי מוטלת על ועדת מכא"ל, הכוללת נציג אחד מכל מחלקה לפסיכולוגיה המשתתפת בפרוייקט.

כל נציג נבחר (או מתמנה) לכהונה של שנה, ואין הגבלה על מספר השנים שבהן הוא עשוי לכהן בוועדה. מקרב חברי הוועדה נבחר יושב ראש (שוב - לכהונה של שנה, ללא הגבלה על מספר השנים), שתפקידו לכנס את הוועדה לישיבות, לנהלן, להפיץ פרוטוקול ולדאוג לביצוע החלטות הוועדה. הוועדה מממנת את כל פעולותיה, כולל בניית והעברת המבחן, דיווח ציונים ופעילות מתקרית - מתשלום דמי בחינה הנגבים מהנבחנים. דמי הבחינה הם מקור המימון היחיד של הפרוייקט ולכן גובהם נקבע כפועל יוצא מעלות הפעלת הפרוייקט. ב-1987 היו דמי הבחינה שווי ערך ל-\$30.

ועדת מכא"ל מעסיקה רכז המוציא לפועל את החלטותיה ורוכשת את כל שירותיה (רכזות, מזכירות, שירותי משרד ושירותים מקצועיים אחרים) מהמרכז הארצי לבחינות ולהערכה, אשר מבצע גם את גביית דמי הבחינה וניהול כספי מכא"ל.

## 3. בניית המבחן

שאלות המבחן נכתבות ע"י מורים המלמדים את מקצועות הפסיכולוגיה השונים באוניברסיטאות בארץ. בתחילת כל שנת לימודים פונה ועדת מכא"ל, באמצעות נציגיה, אל מורה אחד בכל תחום בכל אחת מחמש המחלקות לפסיכולוגיה, ומבקשת ממנו לחבר כחמישה פריטים על החומר שלימד.

השאלות עוברות עריכה ראשונית (לשונית, תרגום מונחים, וכו') והדפסה במרכז הארצי לבחינות. הן נשלחות לביקורת מוצלבת למורים מומחים בכל תחום בכל אחת מהמחלקות. הביקורת כוללת: פתירת השאלות, הגהה לשונית ועובדתית, הצעת שינויים ואישור כללי להכללת השאלה במבחן. סיבות אפשריות לפסילת שאלות בשלב זה, ע"י המורים המומחים: רמת קושי לא מתאימה, העדרות הנושא מחומר הלימוד של מחלקה כלשהי, רמת ספציפיות גבוהה מדי, היעדר רלוונטיות, וכדומה.

מתוך הפריטים שאינם נפסלים בשלב זה, ובעזרת עד 20 אחוז פריטים ישנים משנים קודמות, בונים במרכז הארצי הצעת מבחן המונה 160 שאלות, המוגשות לאישור ועדת מכא"ל. הועדה מכניסה שינויים מעטים בהצעת המבחן, ובעקבותיהם נוצרת גירסה נוספת, "לפני אחרונה". גירסה זו עוברת הגהה וביקורת סופית בידי מספר פסיכולוגים ובוני מבחנים מנוסים במרכז הארצי אשר בעקבותיה נוצרת הגירסה הסופית של המבחן.

#### 4. תכונות, תוכן ומבנה המבחן

כאמור, המבחן מכיל באופן ראשוני 160 שאלות, מסוג "בחירה": לכל שאלה ארבע או חמש אפשרויות תשובה. הנבחן מתבקש לבחור את התשובה האחת המתאימה ביותר לכל שאלה, ולסמן את מספרה בדף התשובות.

אין "ענישה" על תשובות שגויות, ולכן הנבחנים מעודדים בעצם לענות בכל מקרה, גם במצב של אי-ידיעה. הזמן המוקצב למבחן הוא שלוש שעות, והוא כולל את קריאת ההוראות, ומספר דקות נוספות מוקצבות למילוי שאלון קצר של פרטים אישיים. אין כוונה שמבחן זה יהיה מואץ (speeded), דהיינו, גורם הזמן לא אמור להוות מכשול בפני הנבחן, ואף אינו כזה בפועל.

הנושאים הכלולים במבחן חולקו מאז תחילת הפרוייקט לשישה תחומים:

- א. אישיות ומבחני אישיות
- ב. סטטיסטיקה, שיטות מחקר ומדידה
- ג. פסיכולוגיה קוגניטיבית, למידה, תפיסה
- ד. פסיכולוגיה חברתית
- ה. פסיכולוגיה התפתחותית
- ו. פסיכולוגיה פיזיולוגית

ההיקף ומידת הספציפיות הנדרשים בכל נושא הינם ברמה של המבואות וקורסי החובה ל-B.A. באוניברסיטאות בארץ. כל תחום מקבל בציון הסופי משקל ביחס ישר לכמות השאלות שבו. גם הרכב התחומים וגם משקלם היחסי עברו במשך שבע שנות הפעלת הפרוייקט מספר שינויים, אשר שיקפו תזוזות בהלכי הרוח הדומיננטיים במחלקות לפסיכולוגיה. כך למשל, נוצר ב-1986 לחץ לכלול את נושא תורת המבחנים ברשימת תחומי מכא"ל. מאחר שקיימת עדיין לפחות מחלקה אחת אשר אינה מעבירה קורס חובה ב-B.A. בתורת המבחנים, לא נכלל נושא המדידה כתחום נפרד, אך מ-1987 כולל תחום ב': "סטטיסטיקה ושיטות מחקר" גם נושאי מדידה. תחום ג', "פסיכולוגיה קוגניטיבית, למידה ותפיסה", נתפס, מפאת ריבוי הנושאים הכלולים בו, כבעל חשיבות רבה יותר מאשר בעבר. לפיכך, ב-1986, הוכפל כמעט משקלו היחסי בהשוואה לחמשת התחומים האחרים.

בלוח מס' 4.1 מוצגות התפלגויות השאלות לפי תחומי מכא"ל, ולפי שנים, באחוזים מכלל כ-160 שאלות המבחן.

לוח 4.1 - משקלם היחסי של תחומי מכא"ל, לפי שנים, באחוזים (מתוך 160 שאלות)

תחום	שנה						
	87	86	85	84	83	82	81
א. אישיות ומבחני אישיות	15.6	15.6	21.3	21.9	22.1	26.1	24.4
ב. סטטיסטיקה, שיטות מחקר*	15.0	15.0	16.3	15.6	16.4	15.3	15.6
ג. קוג', למידה, תפיסה	25.6	26.3	19.4	18.8	20.2	16.6	18.8
ד. חברתית	15.0	14.4	15.6	15.6	12.7	13.4	16.2
ה. התפתחותית	14.4	13.4	15.6	15.6	15.8	15.9	15.6
ו. פיזיולוגית	14.4	15.0	11.9	12.5	12.7	12.7	9.4

\* ומדידה - החל מ-1987

ב-1984 גויסו מספר עוזרי מחקר מכל האוניברסיטאות החברות, במטרה לערוך סקר מעמיק במחלקות לפסיכולוגיה, באשר להרכב הנושאים הנלמדים בכל אחד מתחומי המבחן. סקר זה הניב שני מוצרים עיקריים: משפט מיפוי שיטחתי להגדרת "מבנה הפסיכולוגיה" וכן אוסף מפורט של רשימות נושאים לפי תחומים, המהווים מכנה משותף לתכניות הלימודים במחלקות החברות. משפט המיפוי מבוסס על עבודתו של א. גריפל, הרכז הראשון של מכא"ל. עוזרי המחקר שהועסקו בסקר, נתבקשו גם לחבר שאלות לתחומי מכא"ל השונים בהתאמה עם משפט המיפוי הכללי. הם נתקלו בקשיים ביישום המשפט לשפת המעשה, קרי, לחבר על פיו פריטים למבחן. המשפט אף לא הוכיח את יעילותו בניתוח תוכן פוסט-הוק, בשיטת סילום רב-מימדי SSA, שבוצע על מספר מבחני מכא"ל. הניתוח לא הצליח לספק שיחזור אמפירי להגדרה התיאורטית המשתמעת ממשפט המיפוי. לעומת זאת, תוצאות הסקר סיפקו תפריט שימושי ביותר להגדרת הנושאים השונים בכל אחד מתחומי המבחן. מאז עריכתו, מהווה תפריט זה שלד בסיסי לבניית המבחן, והוא מבטא את ההסכמה שהיתה קיימת במחלקות לפסיכולוגיה בשנת 1984, באשר לחומר הבסיסי הראוי להילמד במהלך התואר הראשון בפסיכולוגיה. להלן משפט המיפוי השיטחתי להגדרת תוכן שאלות מכא"ל ולאחר מכן - תאור תפריט מכא"ל.

**משפט מיפוי לבדיקת ידע במקצועות הפסיכולוגיה (משפט אב)**

פריט בודק ידע במקצוע הפסיכולוגיה אם ורק אם הוא שואל על:

<p><b>א. סוג הטיפוּל</b></p> <p>הפעלת  </p> <p>כלל- &lt;   אמפירי: ממצא, עובדה, תופעה   על פי   התנהגותית</p> <p>הסקת     הגיוני: סטטיסטי, לוגי   גישה- &lt;   קוגניטיבית</p> <p>משולב: ניסוי, תאוריה, גישה     דינמית</p> <p>פנומנולוגית  </p> <p>מדעית בכלל  </p> <p>ספציפית אחרת  </p>	<p><b>ב. סוג הכלל</b></p> <p>סמנטי: מושג/שם, קביעה  </p> <p>אמפירי: ממצא, עובדה, תופעה   על פי   התנהגותית</p> <p>הגיוני: סטטיסטי, לוגי   גישה- &lt;   קוגניטיבית</p> <p>משולב: ניסוי, תאוריה, גישה     דינמית</p> <p>פנומנולוגית  </p> <p>מדעית בכלל  </p> <p>ספציפית אחרת  </p>	<p><b>ג. סוג הגישה</b></p> <p>נוירופיזיולוגית  </p> <p>התנהגותית  </p> <p>קוגניטיבית  </p> <p>דינמית  </p> <p>פנומנולוגית  </p> <p>מדעית בכלל  </p> <p>ספציפית אחרת  </p>
---	---	---

<p><b>ה. רפרנט</b></p> <p>עצמו  </p> <p>בקבוצה   בגיל- &lt;</p> <p>בכלל  </p>	<p><b>ד. השלב המדעי</b></p> <p>תכנון ניסויים  </p> <p>עריכת תצפיות  </p> <p>עיבוד נתונים   של התנהגות אדם- &lt;</p> <p>קביעה  </p> <p>תיאוריה וגישה  </p>	<p><b>ו. גיל</b></p> <p>ינקות  </p> <p>ילדות  </p> <p>חביון  </p> <p>נערות   והטווח סדור מ- &lt;   ועד   , על פי המקובל במקצועות הפסיכולוגיה.</p> <p>בגרות  </p> <p>זיקנה  </p> <p>בכלל  </p>
---	---	---

<p><b>ז. טווח</b></p> <p>נכון מאוד  </p> <p>ועד   , על פי המקובל במקצועות הפסיכולוגיה.</p> <p>לא נכון כלל  </p>	<p><b>ח. גיל</b></p> <p>ינקות  </p> <p>ילדות  </p> <p>חביון  </p> <p>נערות   והטווח סדור מ- &lt;   ועד   , על פי המקובל במקצועות הפסיכולוגיה.</p> <p>בגרות  </p> <p>זיקנה  </p> <p>בכלל  </p>
---	---

כאמור, ערכו של המשפט היה עד כה תאורטי בלבד. הוא מוצג לעיל כהדגמה לצורת חשיבה וניתוח, שעשויה בהקשרים רבים להיות שימושית ומועילה.



בלוח 4.2 מוצגות רשימות הנושאים, או תפריט מכא"ל, כפי שגובש בסקר שנערך ב-1984.

לוח 4.2 - תפריט מכא"ל

ג. קוגניטיבית, תפיסה, למידה	ב. סטטיסטיקה, שיטות מחקר ומדידה*	א. אישיות ומבחני אישיות
<p><u>קוגניטיבית</u> זכרון: תהליכים, LTM, STM, שיטות הצפנה, מבנה ידע: סמנטי, ויזואלי, אפיזודי. שפה - חומסקי, פתרון בעיות -חשיבה-מודלים.</p> <p><u>תפיסה</u> פסיכופיזיקה, עיבוד ויזואלי (פריפריאלי, מרכזי), עיבוד אודיטורי (פריפריאלי, מרכזי), קשב, קביעויות תפיסיות, מולד-נלמד.</p> <p><u>למידה</u> קלאסית, אופרנטית, מערכי חיזוק, בנדורה, חוסר ישע נלמד, טולמן, האל, למידה לטנטית, השוואה, ביקורות.</p>	<p><u>סטטיסטיקה תאורית</u> מדדי מרכז, פיזור ומיקום יחסי, התפלגויות (כולל התפלגויות דגימה), קשר בין משתנים, מתאמים, רגרסיות.</p> <p><u>סטטיסטיקה הסקתית</u> תאוריה של בדיקת השערות, מבחנים פרמטריים (חישובים, אפקטים), מבחנים אי-פרמטריים (חישובים, אפקטים).</p> <p><u>שיטות מחקר</u> מערכי מחקר, התאמת שיטות סטטיסטיות למערכי מחקר, הערכת מחקר - מהימנות ותוקף, איסוף נתונים.</p>	<p><u>גישה פסיכואנליטית:</u> פרויד, אדלר, סליבן, יונג, אריקסון, אנה פרויד, כללי + השוואתי.</p> <p><u>גישה חברתית:</u> בנדורה, קלי, כללי.</p> <p><u>גישה ביהביוריסטית:</u> סקינר, כללי.</p> <p><u>גישה קוגניטיבית:</u> לוינ, גשטלט, כללי.</p> <p><u>גישה הומניסטית:</u> רוג'רס, מאטלן, אולפורט, כללי-השוואתי</p> <p><u>גישת התכונות:</u> שלדון, קטל, כללי + השוואתי.</p> <p><u>גישה אקטיסטנציאלית:</u> פרום.</p> <p><u>מבחני אישיות:</u> שאלוני אישיות, מבחנים השלכתיים, שיטות מחקר, כללי + השוואתי.</p> <p><u>שאלות כלליות:</u> השוואה בין תאוריות.</p>

\* מדידה - החל מ-1987

ו. פיזיולוגית	ה. התפתחותית	ד. חברתית
<p>מערכת העצבים - מבוא, מבנה נוירון, פוטנציאל פעולה, מנוחה סינפסה, העברה סינפטית, רפלקס, מוח-כללי, מבנה, חלקים, קורטקס - אזורים, מוטוריקה, שפה, מוח חצוי, פגיעות מוחיות, ראייה, שמיעה, המיספרות מערכת אנדוקרינית (הורמונלית), מערכת אוטונומית, שיטות מחקר, סמים - תרופות, שינה, כאב.</p>	<p><u>גישה פסיכואנליטית</u> מיניות הילד, הזדהות/הפנמה, זהות מינית, שלבים פסיכו-סקטואלים, אריקסון, תסביך אדיפוס <u>התפתחות קוגניטיבית ופיזיולוגית</u> כשרים מוקדמים, התפתחות סנסומוטורית, רכישת שפה, מושגי יסוד בגישת פיאז'ה, קוהלברג, התפתחות מוסרית, מושג מרחב/זמן, הבדלים בינאישיים בילדות. <u>התפתחות רגשית/ חברתית</u> התנסויות מוקדמות (הרלו), תפקיד בני הגיל בסוציאליזציה, התקשרות attachment, השפעות חינוכיות, משחקים, אינטראקציה חברתית.  <u>גיל ההתבגרות</u> התפתחות גופנית והשלכותיה, התפתחות חברתית, משבר זהות, התפתחות מנטלית רגשית ומינית, אריקסון.</p>	<p>כללית, שיטות מחקר, תפיסה בינאישית, תורת הייחוס, עמדות-מבנה ושינוי עמדות, עמדות והתנהגות, תקשורת בינאישית וערוצי תקשורת, מנהיגות - השפעה ועוצמה, התחברות, משיכה ואהבה, הושטת עזרה, תוקפנות ואלימות, קבוצות-מבנה ודינמיקה. השוואה חברתית, משחקים חברתיים.</p>

5. ניתוח המבחן

תכנית מחשב בשם RITANA משמשת לניתוח פריטים סטנדרטי במרכז הארצי לבחינות. התכנית מספקת גם נתונים סטטיסטיים לניתוח המבחן כולו, וכן נתונים ומדדים לגבי ההבדלים בין המוסדות השונים בביצוע כל פריט ופריט. ההחלטה לגבי פסילתם או הכללתם של פריטי המבחן בחישוב הציון במכא"ל, מתבססת על ניתוח פריטים זה. להלן פירוט המידע המדווח בניתוח הפריטים ברמה של פריטים בודדים. בהמשך - פירוט המידע המדווח ברמה של כל המבחן; ולבסוף - פירוט השיקולים המנחים את ההחלטה בדבר פסילת הפריטים.

מידע על כל פריט:

- א. מתאם point-biserial בין הערך הדיכוטומי: תשובה נכונה/בלתי נכונה (או שכלל לא ענו עליה), לבין ציון הגלם במבחן כולו. ציון הגלם הוא מספר התשובות הנכונות, למעט התשובה על הפריט עצמו.
- ב. מתאם point-biserial בין פריט (ערך דיכוטומי) לבין סך התשובות הנכונות בתחום שאליו מסווג הפריט (למעט הפריט).
- ג. תמישה מתאמים כנ"ל, אך עם יתר התחומים.
- ד. אחוז העונים נכון, אחוז השוגים, ואחוז המשמיטים (שאינם עונים).
- ה. אחוז העונים נכון מתוך סך העונים (ללא המשמיטים).
- ו. אחוז העונים (סעיף ה') מתורגם לציון תקן המכונה דלתא, שממוצעו 13 וסטית התקן שלו הינה 4. התיקנון מתבצע ע"י התאמה לכל פרופורציה של עונים את ציון התקן אשר מתאים לאותה פרופורציה של שטח מתחת לעקומה הנורמלית.
- ז. לכל אחת מאלטרנטיבות התשובה - מסיחים ותשובה נכונה - מדווחים שלושת המדדים הבאים: אחוז הבחורים באלטרנטיבה, ציון הגלם הממוצע (במבחן כולו) של קבוצת הבחורים בכל אלטרנטיבה, ואחרון - המתאם point-biserial בין הערך הדיכוטומי: בחירה/אי בחירה באלטרנטיבה, לבין ציון סה"כ המבחן כולו, כולל הפריט עצמו.

לכל פריט מחושבים אחוזי העונים נכונה בקרב כל אחת מקבוצות בוגרי המוסדות השונים. עד סמך פרופורציות אלו, מחושבים שני מדדים להטייה על-פי מוסדות. "מקדמי יעילות" אלה משמשים כמבחני הבדל למוצעים בין קבוצות, במקרה של דיכוטומיה (התשובה האפשרית הינה נכון/לא נכון):

1. יחס המתאם  $\eta$  (אטה) - שהוא היחס בין השונות בין הקבוצות, לבין השונות הכללית. מקדם זה מבטא את עוצמת הקשר בין ההשתייכות הקבוצתית, לבין הביצוע בפריט. נוסחת העבודה לחישוב המקדם הינה:

$$5.1 \quad (ETA) = \eta = \sqrt{1 - \frac{\sum n_i p_i q_i}{n_t P_t Q_t}}$$

כאשר:

- $n_i$  = גודל קבוצה  $i$
- $p_i$  = פרופורצית העונים נכון על הפריט בקרב נבחני קבוצה  $i$
- $1 - p_i = q_i$
- $t$  = כל הנבחנים מכל הקבוצות

2. מקדם היעילות (efficacy) של הבחנה בין קבוצות -  $\mu$  (מיו, פותח ע"י לואי גוטמן).

נוסחת העבודה לחישובו:

$$5.2 \quad (MU) = \mu = \frac{2n_t (n_t P_t Q_t - \sum_{i=1}^m P_i Q_i)}{\sum_{a=1}^m \sum_{b=1}^m (X_a (n_b - x_b) + X_b (n_a - x_a)) \left| \frac{X_a}{n_a} - \frac{X_b}{n_b} \right|}$$

כאשר:

- t - סך כל הקבוצות
- i - קבוצה i, מתוך m קבוצות
- a, b - זוג כלשהו של קבוצות, מתוך m
- X - מספר העונים נכון על פריט
- n - גודל הקבוצה
- P - פרופורצית העונים נכון  $\left(\frac{X}{n}\right)$
- Q - (1-P)

ערכו של  $\mu$  נע בין אפס ל-1, כאשר אפס מעיד על כך שפריט מסויים אינו "יעיל" כלל בהבחנה בין קבוצות נבחנים שונים.  $\mu$  רגיש מאוד לפערים בין הקבוצות בקצוות של דרגות הקושי: באיזורי ה-100% וה-0% עונים; ורגיש פחות באיזור ה-50% עונים. השימוש במקדמים  $\eta$  (אטה) ו- $\mu$  (מיו) נעשה באופן השרירותי הבא: פריט מוגדר כ"מוטה" ונפסל מלהיות כלול בחישוב הציון במבחן, אם ערך ה- $\mu$  (מיו) שלו גדול או שווה ל-0.5, וערך ה- $\eta$  (אטה) שלו גדול או שווה ל-0.3. הגדרה זו, בצירוף שיקול דעת של ועדת מכ"ל, המביא בחשבון את ערך הפער המקסימלי בין הקבוצות הקיצוניות - פוסלים מדי שנה כ-30 מתוך 160 פריטי מכ"ל. זהו ממצא אמפירי המגלה יציבות רבה למדי מעבר לשנים.

## 6. הטיית פריטים ופסילות

בספרות העוסקת בתורת המבחנים כמו Crocker & Algina (1986) מוגדר אוסף של פריטים כבלתי מוטים, אם:

1. הפריטים מושפעים מאותם מקורות של שונות בכל תת-הקבוצות המשוות, ו-
2. בקרב נבחנים בעלי רמת יכולת דומה בכושר או בתכונה האמורים להמדד ע"י המבחן, מקורות השונות הבלתי-מוסברים או הבלתי רלוונטיים מתפלגים באופן דומה עבור כל תת-הקבוצות.

Crocker ו- Algina סוקרים שלוש גישות עיקריות בחקר ההטייה של פריטים, מתוך אמונה שהן המבטיחות והכלליות ביותר:

א. גישה המבוססת על "תורת התכונה החבויה", או "התגובה לפריט" (IRT). מורכבות הרבה של המודל אינה מאפשרת לדון בו במסגרת הנוכחית. למעשה, השימוש במודל זה אינו אפשרי במסגרת מבחן כמו מכ"ל, מפאת המספרים הקטנים של נבחנים מדי שנה.

ב. גישה המבוססת על המבחן הסטטיסטי המכונה  $\chi^2$  - "גישת חי בריבוע". טכניקות בגישה זו פותחו ע"י Scheuneman (1979) ו- Camilli (1979), והן יכולות להחשב קירובים טובים ל"תאורית התגובה לפריט". יתרונו בכך שהן קלות יותר לישום מאלו המבוססות על IRT. הטכניקות הללו מגדירות פריט כבלתי מוטה, אם בקרב קבוצת נבחנים בעלי הישג דומה (דהיינו, שציוניהם נופלים בטווח מוגדר של ציוני המבחן), פרופורצית הנבחנים העונים נכון על הפריט דומה עבור כל תת-הקבוצות. "התכונה החבויה" שבבסיס IRT, מוחלפת כאן בהישג הכללי הנצפה

במבחן. הישג כללי זה מחולק שרירותית למספר אינטרוולים, ובתוך כל אחד מהם משווים אחוזי העונים נכונה על כל פריט, בין כל תת-קבוצות. אם קיימת שונות רבה בפרופורציות העונים, הפריט נחשב מוטא:

ג. גישה המבוססת על "קושי הפריט". הטכניקות בגישה זו משתמשות במדדים של קושי הפריט: מדד ישיר, כמו פרופורצית העונים עליו נכון (P), או טרנספורמציה של פרופורציה זו. הטכניקות מתבססות על שתי הגדרות: האחת מגדירה קבוצת פריטים כבלתי מוטים, אם דרגות הקושי של הפריטים עבור תת-קבוצה א', מתואמים באופן מושלם עם דרגות הקושי של הפריטים עבור תת-קבוצה ב'. ההגדרה השנייה של פריטים כבלתי מוטים דורשת שהבדלי הקושי בין תת-קבוצות יהיו זהים עבור כל הפריטים. פריט יחשב מוטא, אם הפער בין תת-קבוצות באחוזי העונים עליו נכון, שונה משמעותית מהפער הטיפוסי ביתר הפריטים. ראוי לציין כי ההגדרה השנייה מקיימת גם את התנאי של הראשונה, אך לא להיפך.

הטכניקה הידועה ביותר המבוססת על ההגדרה הראשונה פותחה ע"י Angoff & Ford (1973). היא משתמשת ב-דלתא, המוגדרת כדלקמן:  $\Delta_i = 4Z_i + 13$ , כאשר  $Z_i$  הוא הערך הסטנדרטי בהתפלגות הנורמלית (0, 1), אשר חותך את הפרופורציה Pi, שהיא פרופורצית העונים נכון על פריט i. (וכן ראה סעיף 5 ו' שלעיל).

בדיאגרמה שבה ציר ה- x מבטא את דרגות הקושי בתת-קבוצה א', ציר ה-y מבטא את דרגות הקושי בתת-קבוצה ב', וכל נקודה מייצגת פריט - אוסף הנקודות עבור קבוצת פריטים בלתי-מוטים יפול על קו אלכסוני ישר.

המתאם בין ערכי הדלתא בשתי הקבוצות יהיה במקרה כזה 1. אולם גם במצב של מתאם קרוב מאוד ל-1, יתכנו מספר פריטים שנופלים במרחק מה מהקו הישר, והם יחשבו כמוטים. זוהי תמצית שיטת אנגוף. לשיטה מספר נקודות תורפה, אך ניתנו להן מספר פתרונות ע"י Angoff עצמו (1982) ו-Jensen (1980).

כזכור ציינו לעיל הגדרה שנייה לפריטים בלתי מוטים, הדורשת שהבדלי הקושי בין קבוצותיה יהיו זהים עבור כל הפריטים. טכניקות הקשורות בה הוצעו בעבודותיהם של Cardall & Coffman (1964), Cleary & Hilton (1986) ו-Plake & Hoover (1977). גם השיטות הנגזרות מההגדרה הראשונה, גם אלו הנגזרות מהשנייה, לוקות בעיקר בכך שעדות להטייה על-פיהן, עשויה להיות תוצאה מלאכותית (artifact) של הבדלים ביכולת בין תת-קבוצות המשוות.

### 6.1 מקרה מכא"ל

נושא פסילת הפריטים בדיעבד במכא"ל זוכה מדי שנה להתייחסות רבה מצד בוני המבחן וחברי ועדת מכא"ל. מדובר בפסילה מסיבות של הטייה בין-מוסדות. מקרה מכא"ל הינו ייחודי על רקע הספרות שנסקרה לעיל בנושא הטייה של פריטים, בכך שמדובר בהשוואה בו-זמנית של יותר מ-2 תת-קבוצות. לפיכך, כל טכניקה להכרעה בדבר לגיטימיות של פריט, צריכה להציע כלי המגדיר בו-זמנית, עבור כל חמשת המוסדות, האם פריט הינו מוטא או לא.

השיטה הנהוגה מאז תחילת הפרוייקט, והמתוארת בסוף סעיף 5 "ניתוח המבחן", מניחה שוויון ביכולת בין קבוצות הנבחנים השונות. הנחה זו מוצאת ביטוי בעובדה שהחלטה לגבי פסילה מתקבלת עבור כל פריט בנפרד, ללא תלות בביצוע היחסי של הקבוצות ביתר פריטי המבחן.

פריט מוגדר בשיטה זו כ"מוטה", אם מתגלים בו פערים מוחלטים באחוזי העונים עליו נכונה, בין חמש קבוצות הבוגרים. פריט נחשב "לגיטימי" אם מתקיים בו שוויון פחות או יותר, באחוזי העונים עליו, בין חמש קבוצות הבוגרים השונות. אחת הביקורות העיקריות על שיטת פסילת פריטים זו כוונה להנחת היסוד בדבר שוויון היכולת בין בוגרי המוסדות השונים. מנתונים אמפיריים עלה שוויון זה אינו מתקיים: לא ברמה הכללית של סך כל הנושאים הכלולים במכ"ל, ולא בכל תחום בנפרד. תכניות הלימודים העניקו דגשים משתנים לתחומים השונים בכל מוסד ומוסד; מן הסתם, התקיימו ומתקיימים, גם הבדלים בין המחלקות ברמת הסלקטיביות בקבלה ללימודי התואר הראשון, וכן הבדלים בסלקציה העצמית של התלמידים. וזאת מבלי להניח שום הנחות לגבי הבדלים ברמת ההוראה בין המחלקות.

ליקוי נוסף במדדים הקיימים לצורך פסילת פריטים, נעוץ בעובדה שהם אינם מתחשבים במידת הקושי האבסולוטית של כל פריט (מעבר לכל המוסדות), ובהבדלים בין המוסדות בפניזור מידות הקושי של הפריטים. כך, למשל, להבדל נתון בין שני מוסדות באחוז העונים נכון על פריט, יש משמעות שונה אם מדובר בפריט בינוני מבחינת רמת הקושי שלו, או שהוא קיצוני ברמת הקושי (קשה או קל במיוחד). כמו כן, ביטוי רמת הקושי של פריט, יחסית ליתר הפריטים בתחום מסויים, במוסד מסויים, אמור להביא בחשבון את מידת ההומוגניות של רמות הקושי בקרב פריטי אותו תחום באותו מוסד. לפיכך, עלו מספר הצעות לשיטות פסילה חלופיות, אשר המשותף לחלקן היה שההשוואה בין רמות הביצוע של הקבוצות השונות בכל פריט ופריט, מתחשבת ברמה הבסיסית של כל קבוצה וקבוצה, בכל אחד מתחומי המבחן. באורח זה, ההשוואה תיעשה בין רמות הביצוע היחסיות של כל הקבוצות בכל פריט, ולא בין רמות הביצוע המוחלטות שלהן בכל פריט.

מחבר פרק זה, בשיתוף עם פרופ' דוד נבון ודר' דוד בודסקו מאוני' חיפה, הציעו מספר טכניקות חלופיות להגדרת הטייה של פריטים בהקשר הרב-קבוצתי. להלן תאורן:

1. השוואה בין דרוגי הקושי, מתורגמים לציוני תקן, או פרצנטילים של הפריטים בתוך כ"א מאוכלוסיות בוגרי המוסדות השונים. נדגים את הדבר באמצעות פריט, אשר אחוז העונים עליו נכונה מקרב בוגרי מוסד א', גדול בהרבה מאחוז העונים עליו נכונה מקרב בוגרי מוסד ב'. אם לבוגרי מוסד א' רמת יכולת גבוהה מזו של בוגרי מוסד ב' (ואין תשיבות לסיבת הדבר), יתכן שלא יהיה פער גדול, אם בכלל, בין דרוגיו היחסיים של הפריט בכ"א מהאוכלוסיות. פריט כזה היה מוגדר כמוטה על פי השיטה הקיימת, אך לא יוגדר כך לפי השיטה המוצעת.

הפער הנצפה באחוזי העוננים על הפריט נכונה, הינו (לפחות בחלקו) פער "אמיתי" ולא פער מפלה.

מאידך, יתכן פריט, אשר השיטה הנהוגה לא תגדירו כמוטה, משום שאין פער גדול בין פרופורציות העוננים עליו נכונה בין שתי האוכלוסיות. אולם אם לאוכלוסיה א' רמת יכולת גבוהה מאוכלוסיה ב', דרוג הקושי של הפריט באוכלוסיה ב' יהיה גבוה (קשה יותר) מאשר ב-א', וכך יוגדר הפריט כמוטה. ושוב, סיבת הדבר היא הבדלים בין רמות האוכלוסיה המושוות.  
להלן הצעת מדד כזה:

$$D_{ij} = \frac{\text{ההפרש בין ציוני התקן של פריט באוכלוסיות } i \text{ ו-} j}{\text{טעות התקן בשתי אוכלוסיות אלה}}$$

כאשר:

$$\sqrt{1-r_{ij}} = \text{טעות התקן בשתי האוכלוסיות}$$

$$r_{ij} = \text{המתאם בין ציוני הפריטים באוכלוסיות } i \text{ ו-} j.$$

את  $D_{ij}$  ניתן לשקלל בגודלי האוכלוסיות, ולסכם כך שיתקבל מדד כללי  $D$ , מעבר לכל זוגות הקבוצות המושוות. כמו כן, יש לפתח סטטיסטי מבחן למובהקות המדד  $D$ .

2. קיימת דיפרנציאציה בין אוכלוסיות בוגרי המוסדות השונים, לא רק בקושי היחסי של הפריט, אלא גם ביכולת האבחנה שלו, כנמדד, למשל, ע"י המתאם הביסריאלי פריט/סה"כ.

כאשר נערך ניתוח פריטים נפרד לכל אחת מהאוכלוסיות, מתגלה כי פריטים אחדים "מתנהגים" בתוך אוכלוסיה מסוימת (ולא באחרות) באופן שהיה פוסל אותו מהכללה במבחן לו "התנהג" כך בניתוח מעבר לכל הנבחנים. למשל, מצב של מתאם שלילי בין פריט לסה"כ, או מתאם חיובי בין מסיח לבין סה"כ וכו'.

יש לשקול בכובד ראש הפעלת הקריטריונים הפוסלים פריט מסיבות של "ניתוח-פריטים", גם ברמה של תת-קבוצות בודדות!

3. גישה הממזגת שיקולים מסעיפים 1 ו-2 :

שיקולי "הטיה", ושיקולי "הבחנה" של פריטים.

נגדיר שתי התפלגויות, A ו- B.

A - התפלגות מתאמי פריט/סה"כ, (כשכל ערך בהתפלגות הינו המתאם החציוני, מבין חמש האוכלוסיות של פריט מסויים) כאשר המתאם הוא מדד ליכולת ההבחנה של הפריט.

- B - התפלגות ערכי D (ראה סעיף 1) -  
 כש- D הינו המדד למידת ההטיה של פריט.  
 מחשבים לכל פריט, מאון בהתפלגות  $(P_A)A$  ומאון בהתפלגות  $(P_B)B$ .  
 אם  $P_A < P_B$  - הפריט מוטא, והוא מפלה.  
 אם  $P_B < P_A$  - החטייה יכולה להיות מוסברת באמצעות הבחנה.  
 ניתן גם כאן לחשב סטטיסטי מבחן למוקדקות ההפרש.

הועלו גם מספר שיטות פסילה המתבססות על מדדים פשוטים, קלים לחישוב וליישום:

4. ההפרש המקסימלי בין זוג מוסדות כלשהו מבין החמישה במדד הקושי של פריט.  
 5. שונות דרוגי הקושי של פריט מעבר לכל המוסדות, כשכל דירוג מחושב בתוך מוסד לכל תחום.

במדור מחקר של המרכז הארצי לבחירות, נערך מחקר זוטא, לחשוואת מספר שיטות לפסילת פריטים, עם זו הנהוגה בפועל. במחקר זה נתגלתה חפיפה גבוהה בהרכב הפריטים שנפסלו לפי השיטה הנהוגה (עפ"י מדדי "אטה" ו-"מיו"), עם הפריטים שנפסלו עפ"י שיטות 4 ו-5 שלעיל. גם מהימנות המבחן שנותר לאחר הפסילה, וכן רמת הקושי הממוצעת שלו, השתנו אך במעט כתוצאה משינוי השיטה.  
 נתאר בקצרה מחקר זה:

נערכה השוואה בין שיטה 5 (להלן ה"נסיונית") לשיטה הנהוגה בפועל (להלן "הנהוגה") באוכלוסיית נבחני 1984. לפי שיטה 5, כל שאלה קיבלה דירוג עפ"י אחוז עונים עליה נכונה בתוך כל מוסד בנפרד. חושבה שונות הדירוגים הללו בין חמשת המוסדות, והשאלות שנפסלו היו אלו בהן השונות של הדירוגים היתה הגבוהה ביותר. הוחלט לפסול מספר זהה של פריטים כשיטה ה"נהוגה" (אטה + מיו), דהיינו, 25 פריטים, ועוד שלושה שנפסלו מסיבות של ניתוח פריטים, בלא קשר להטייה. נתוחי הפריטים לגירסות המקוצרות לפי שתי השיטות הראו שיפור זעום במהימנות  $0.83: K-R$  ב"נסיונית", לעומת 0.82 ב"נהוגה"; וכן עליה קלה ברמת הקושי: 60.2 אחוז עונים ממוצע ב"נסיונית", לעומת 62.12 אחוז עונים ממוצע ב"נהוגה". מתוך 25 הפריטים שנפסלו בכל אחת מהשיטות, 15 היו משותפים. עשרת הפריטים הייחודיים לכל שיטה, אופיינו כדילקמן:

ממוצע מתאם פריט-תחום	אחוז עונים ממרצע	
0.24	69.14	פריטים שנפסלו ע"י השיטה הנסיונית
0.35	50.86	פריטים שנפסלו ע"י השיטה הנהוגה



השיטה ה"נסיונית", פוסלת פריטים קלים יותר, ובעלי כושר הבחנה נמוך יותר. ביטוי למידת הדמיון בין השיטות ניתן למצוא במתאמי פירסון הבאים:  
 $r=0.68$  בין  $\eta$  (אטה) לבין מדד שונות דרוגי הקושי הנסיוני, ו-  $r=0.53$  בין  $\mu$  (מיו) לבין אותו מדד.

על פני הדברים, נראה שיש עדיפות קלה לשיטה הנסיונית, אך זו חייבת לקבל חיזוק באמצעות רפליקציות, וכן השוואות נוספות עם שיטות אחרות. השיטה ה"נהוגה" לפסילה, מקלה את המבחן, ומקטינה את מהימנותו, בהשאירה בו פחות פריטים מבחינים. השיטה הנסיונית, נראית פחות "מזיקה" מהנהוגה.

מתקר זה היה קטן ומוגבל, וכן לא נבדקו השיטות הנגזרות משלושת הגישות שנסקרו בתחילת ראש פרק זה. אין ספק שראוי להמשיך את המחקר בתחום, ולפתח את החשיבה התיאורטית בנושא זה.

לסיכום הדיון בבעיית הטיית הפריטים במכא"ל והשיקולים לפסילתם, מובעת בזאת עמדת המחבר:

גם אילו נמצאה שיטה או טכניקה להגדרת פריטים כמוקים בין-מוסדית, אשר הינה נעדרת ליקויים ונקודות תורפה, ואשר הינה נוחה ליישום ולחישוב בתנאים ובנתונים של פרויקט מכא"ל, הרי שמבחינה פסיכומטרית טהורה, קשה להצדיק פרוצדורה זו של לאחר-מעשה. במבחן בעל עולמות תוכן מוגדרים היטב, ברמה מקצועית גבוהה ומתוך הסכמה של מומחים לגבי טיב הפריטים מראש, אין צורך בפסילת פריטים בדיעבד, אלא במקרים נדירים של כשלונות סטטיסטיים בלתי צפויים המתגלים בניתוח פריטים קונבנציונלי.

החליך המקובל בפרוייקט כיום, אשר מקיים מנגנון וטכניקה של הגדרת פריטים מוטים בין-מוסדות לצורך פסילתם - משקף עמדה פוליטית של חברי הפרוייקט ולא עמדה מקצועית. אין המחבר פוסל או ממעיט בערכן של החלטות פוליטיות ובדעתו רק להבהיר ולהצהיר על כך בגלוי.

7. הבדלים בין קבוצות

בשנים '81 עד '86 נבחנו כ-1,600 איש במכא"ל. קבוצה גדולה זו מאפשרת בדיקת מספר שאלות מחקריות להשוואת הישגיהן של תת-קבוצות במכא"ל על תחומיו. התפלגויות הנתונים הביורגריים של הנבחנים, מתוארות בנספח, בסוף הפרק, סעיף 11. בלוחות מסויימים מוצגים גם נתונים על קבוצת נבחני 1987.

7.1 הבדלים בין מינים

בלוח 7.1.1 מוצג פרופיל ההישגים לכל מין בנפרד. ההישגים במכא"ל מוצגים בציוני תקן, וממוצע ה-B.A. - בציוני גלס, בסולם 100 - 0. בסוגריים - שכיחות הנבחנים. ציון ה-B.A. דווח ע"י כמחצית מהנבחנים בלבד.

לוח 7.1.1 - הבדלים בין מינים במכא"ל וב-B.A.

תחום	סהכ כנכי	אישיות		סטטי		קוג'		חברי		התפתי		פיזי'		ציון B.A.	
		ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.
מין	שכיחות	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.
זכר	440	0.123	0.83	0.036	0.83	0.230	0.77	0.155	0.89	0.148	0.79	0.001	0.78	0.230	0.74
נקבה	1144	-0.016	0.88	0.052	0.82	0.025	0.84	-0.010	0.90	0.040	0.79	0.087	0.82	0.040	0.82
ערך t		2.94	-0.35*	4.43	3.27	2.43	-1.94*	4.25	-1.14*						

\* הפרש לא מובהק

בציון מכא"ל הכללי, עולים הגברים על הנשים בפער מובהק של 0.40 סטיות תקן. בתחומים אישיות והתפתחותית, וכן בציון ה-B.A - אין פער מובהק בין המינים. ביתר התחומים, וכן בציון הכללי - עדיפות הגברים מובהקת. סדר הפערים המובהקים, מהגבוה אל הנמוך, הינו: סטטיסטיקה ושיטות מחקר, פיזיולוגיה, קוגניטיבית, חברתית. פרופיל הבדלים זה מצוי בהתאמה עם ממצאים מחקריים המקובלים בתחום מבחני הכושר לגבי המינים (Donlon, 1984): עדיפות גבוהה לגברים במדעים המדוייקים, ונמוכה יותר (או נחיתות) בתחומים ה"רכים" - המילוליים או בלתי-כמותיים. אישיות והתפתחותית הם תחומים כאלה. התכנים הספציפיים בתחום ההתפתחותי, הקשורים עם לידה, ועם ילדים - עשויים להסביר הסבר נוסף את הענין הרב שיש לנשים בתחום, ואת הצטיינותן היחסית (לעצמן) בו. המפתיע הוא, שהפער לטובת הגברים מתקיים במכא"ל בהקשר של הישגים, ולא כושר. הסבר חלקי לפער זה, נוסף על הצפוי מתחום מבחני הכושר, עשוי להמצא בסלקציה העצמית השונה של המינים, המתבטאת גם בפער כמותי (הפוך) ולא רק איכותי.

## 7.2 הבדלים בין מעדיפי מגמות שונות בפסיכולוגיה

לכל נבחן ניתנה אפשרות לסמן בשאלות הפרטים האישיים את עדיפותו הראשונה בכ"א מחמשת המוסדות, כך שניתן היה לסמן מגמות שונות במוסדות שונים. והוגדרו ארבע קבוצות של מגמות, אליהן קובצו אנשים כדילקמן:

1. קליניות - מי שסימן בשאלון, באיזשהו מוסד, אחת או יותר מהאופציות: קלינית, קלינית של הילד, שיקומית.
2. מחקריות - מי שסימן: מחקרית, קוגניטיבית, פסיכוביולוגית.
3. חברתיות - מי שסימן: תעסוקתית-ייעוצית, חברתית, חברתית-תעשייתית.
4. חינוכית - מי שסימן: חינוכית.

1509 איש מנבחי '81-'86 סימנו אחת או יותר מהמגמות, באחד או יותר מהמוסדות. הם יצרו 14 צירופים שונים של קבוצות-מגמות, כשהצירופים מכילים קבוצה אחת, שתיים או שלוש. הצירופים המכילים קבוצה אחת הם ה"נקיים", והמעידים במידה רבה על נחישות בחירתם של המסמנים אותם. צירוף "לא נקי", ובו שתיים או שלוש קבוצות, נוצר כאשר, למשל, אדם בוחר בעדיפות ראשונה מגמה קלינית במוסד א', ובמוסד ב' עדיפותו הראשונה היא חברתית.

בלוח 7.2.1 מוצגות השכיחויות של צירופי הקבוצות השונים, וכן הפרופילים של כל הצירופים במכא"ל על תחומי. ההישגים המוצגים הם ציוני התקן בגירסה המקוצצת של מכא"ל. ההשוואה המשמעותית ביותר הינה בין הקבוצות "הנקיות" (שכל מי שסימן אותן לא סימן אחרת בנוסף), והן אמנם מוצגות בראש הטבלה. אולם גם השוואת ההישגים והשכיחויות של יתר הצירופים הינה מעניינת ביותר. עבור ארבע הקבוצות ה"נקיות", בוצעו ניתוחי שונות לבדיקת ההבדלים בין הקבוצות.

לוח 7.2.1 - השוואת פרופילים של מסמני מגמות בעדיפות ראשונה

תחום		אישיות		ססטי		קוגי		חברי		התפי		פיזיו		סריכ	
ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.
קכינית (745)	0.175	0.80	0.155	0.79	0.117	0.91	0.099	0.81	0.132	0.79	0.173	0.76	0.144	0.85	0.144
מחקרית (90)	-0.154	0.90	0.051	0.83	0.011	0.97	-0.074	0.84	-0.075	1.03	0.081	0.78	-0.140	0.95	-0.140
חברתית (138)	-0.068	0.87	0.021	0.85	0.023	0.84	0.120	0.79	0.047	0.80	0.065	0.81	-0.027	0.85	-0.027
חינוכית (93)	-0.445	1.01	-0.388	0.89	-0.340	0.85	-0.221	0.88	-0.376	0.80	-0.396	1.07	-0.609	0.93	-0.609
ערך F (df=3)	19.05*	12.65*	7.24*	5.27*	11.74*	13.91*	22.53*								
אתוז שונות מסוברת	5.1	3.4	2.0	1.5	3.2	3.8	6.0								
צירופי מגמות															
קכינית (35) ומחקרית	0.136	0.80	0.100	0.88	0.112	0.84	0.183	0.72	-0.064	0.83	0.160	0.58	0.049	0.82	0.049
קכינית (96) וחברתית	0.149	0.71	0.168	0.37	0.105	0.81	0.363	0.66	0.150	0.74	0.123	0.62	0.112	0.76	0.112
קכינית (112) וחינוכית	-0.141	0.82	-0.128	0.80	-0.068	0.87	-0.063	0.84	-0.111	0.92	-0.070	0.86	-0.171	0.79	-0.171
מחקרית (73) וחברתית	0.184	0.60	0.263	0.80	0.127	0.88	0.395	0.47	0.350	0.51	0.318	0.76	0.341	0.72	0.341
מחקרית (22) וחינוכית	-0.153	0.63	-0.023	0.67	-0.041	0.64	0.156	0.47	0.115	0.46	-0.101	0.60	-0.188	0.66	-0.188
קכינית (15) מחקרית וחברתית	0.177	0.63	0.457	0.80	0.113	0.74	0.362	0.59	0.089	0.57	0.445	0.79	0.381	0.81	0.381
קכינית (13) מחקרית וחינוכית	-0.161	0.86	0.121	0.75	-0.099	1.23	0.447	1.04	0.019	0.98	0.321	0.71	0.049	1.12	0.049
קכינית (14) חברתית וחינוכית	-0.101	0.56	-0.202	0.69	-0.276	1.10	-0.049	0.65	-0.055	0.50	-0.242	0.97	-0.399	0.82	-0.399
מחקרית (17) חברתית חינוכית	-0.098	0.54	0.063	0.65	-0.001	0.95	0.059	0.60	0.192	0.66	0.095	0.83	-0.044	0.77	-0.044
חברתית (46) וחינוכית	0.108	0.68	-0.147	0.99	-0.312	1.04	0.170	0.69	-0.091	1.01	-0.272	0.93	-0.243	1.02	-0.243

\* P < 00.5

כל תחומי מכ"ל, כמו גם ציון סה"כ, נמצאו מסבירים באופן מובהק את ההבדלים בין ארבע קבוצות המגמות ה"נקיות". ההשתייכות למגמות מסבירה 6 אחוזים מהשונות בציוני סה"כ. תחום האישיות הינו המובחן ביותר (5.1% מהשונות), ואחריו פיזיולוגיה וסטטיסטיקה. חברתית הינו המובחן החלש ביותר (1.5% מהשונות). 1066 איש, שהם שני שלישי מהנבחרים המסמנים, בחרו בקבוצת מגמות אחת בלבד. 70 אחוז מהם, שהם 745 איש, בחרו במגמות הקליניות. מעדיפי הקלינית הינם בעלי ההישג הממוצע הגבוה ביותר (כ- 0.14 ס.ת. מעל ממוצע כל הנבחרים), ומסמני המגמה החינוכית - בעלי ההישג הממוצע הנמוך ביותר - 0.6 סטיות תקן מתחת לממוצע כל הנבחרים.

הפרופילים של כל קבוצה נותנים חיזוק לתוקף התוכן של מכ"ל, בכך שתחומי ההצטיינות היחסית של כל קבוצה מצויים בהתאמה מסויימת עם תחומי התוכן המרכזיים של מגמת הלימוד שנבחרה.

כך, למשל, בוחרים המגמות הקליניות מצטיינים (יחסית לעצמם) באישיות ומבחני אישיות. בוחרים המגמות המחקריות מצטיינים (יחסית לעצמם) בסטטיסטיקה, שיטות מתקר, פסיכולוגיה קוגניטיבית, למידה ותפיסה, וגרועים יחסית באישיות.

מעדיפי המגמות החברתיות מצטיינים (יחסית לעצמם) בפסיכולוגיה חברתית. הפרופיל של קבוצת המגמות המחקריות הינו היחידי שאינו "שטוח": יש בו קיטוב בין תחומים מדעיים לבין תחומי האישיות, לטובת הראשונים, כמובן.

התמונה המתוארת מצביעה על מודעות מסויימת הקיימת בקרב הקבוצות ה"נקיות" באשר לנקודות החוזק/חולשה שלהם, וכן על הכוונה מקצועית עצמית המצויה בהתאמה עם הכישורים. אין כמובן דרך להבהיר את כיוון הסיביות: ההצטיינות גוררת בחירת כיוון מקצועי, או להיפך.

באשר לקבוצות הצירופים האחרות, אלו אשר בחרו שתים או שלוש קבוצות של מגמות - לא ניתן לדעת מה באמת עדיף בעיניהם, וכן קשה לבודד את הקשר שבין ההעדפה המקצועית לבין רמת הקבוצה. את ההשוואה המפורטת של הפרופילים נשאר לקורא הסקרן, אולם נציין רק מספר עובדות בולטות:

כל הצירופים בהם משתתפת המגמה החינוכית נמוכים מממוצע כלל נבחני מכ"ל.

הצירופים בהם משתתפות המגמות הקלינית והמחקרית - גבוהים מהממוצע.

הצירופים המצטיינים בציון הכולל, ובכל תחומי מכ"ל - מונים 88 נבחרים שסימנו מחקרית וחברתית, או קלינית, מחקרית וחברתית: כ- 0.36 ס.ת. מעל ממוצע נבחני מכ"ל!

### 7.3 קשר בין מכ"ל לבין משתני רקע

בסעיף זה נדווח שני מימצאים:

#### 7.3.1 גיל הנבחן

המתאם בין הציון הכללי במכ"ל לבין גיל הנבחן (בעת ביצוע הבחינה) הינו -0.16, והוא מובהק (חושב על 1554 נבחנים). ההסבר הפשוט ביותר לכך הוא שגיל גבוה מלווה, כנראה, בהתרחקות בזמן מחומר הלימודים, שאינה פועלת לטובת הזכרון והיכולת בבחינה... וזאת מבלי להניח הנחות לגבי ירידה ביכולת השכלית עם הגיל.

### 7.3.2 השכלת האב

המתאם (פירסון) בין ההישג הממוצע במכא"ל לבין השכלת האב הינו  $0.13$  ( $p < 0.005$ ),  $n=1554$ .

הקשר הידוע מהספרות בין משתני יכולת שכלית לבין השכלת ההורים גם הוא חיובי: בקרב כ-20,000 מנבחני המרכז הארצי לבחינות והערכה בשנת 1984, בבחינה הפסיכומטרית, נמצא מתאם של  $0.40$  בין השכלת האב לבין הציון הפסיכומטרי הכללי. אולם מפתיע להוכיח שהמתאם שנמצא בקרב נבחני מכא"ל נשאר חיובי ומובהק, למרות קיצוץ התחום החמור בשונות היכולת של הקבוצה, יחסית לאוכלוסית מועמדי האוניברסיטאות: לא זאת בלבד, שנבחני מכא"ל הם כאלה שהתקבלו לאוניברסיטה וסיימו תואר ראשון, הם גם מהגבוהים שבמתקבלים, עפ"י קריטריונים של יכולת אקדמית. נראה שמהשפעת ההורים קשה להפטר, גם בגיל הממוצע המופלג של  $26.5$ .

## 8. מהימנות, נבחנים חוזרים, תקפות, ויציבות האוכלוסיות מעבר לשנים

### 8.1 מהימנות

נגדיר בקצרה את מושג המהימנות, מתוך התייחסות לשני סוגים: פנימית, ומדידות חוזרות. במהימנות העקביות הפנימית (internal consistency), הכוונה היא למידת העקביות בין פריטי המבחן, כלומר, עד כמה הם מהווים מדידות של אותה תכונה. מהימנות זו עשויה להיות מושפעת משונות טעותית שמקורה במידת ההטרוגניות של התכנים והאופרציות המנטליות, אשר נמדדים ע"י פריטי המבחן. אחת הדרכים היותר מקובלות לאומדן מהימנות פנימית הינה נוסחת קודר-ריצ'רדסון, המכונה  $K-R$  (20). מהימנות של מדידות חוזרות נאמדת ע"י המתאם בין ההישגים שהשיגו אותם נבחנים בשתי העברות שונות של המבחן (test-retest). מהימנות זו מורה על המידה בה ניתן להכליל את ההישגים במבחן מעבר למקרים שונים, וככל שמהימנות זו גבוהה יותר, ניתן לומר על הציונים במבחן שהם נתונים פחות להשפעתם של שינויים אקראיים - בסביבה, במצב, או בנבחן.

בלוח 8.1.1 מושווים נתוני הגלם, אחוזי הקושי והמהימנות הפנימית של המבחן במתכונתו המלאה (עפ"י 160 שאלות), עם מתכונתו המקוצצת, לפי שנים.

לוח 8.1.1 מהימנות עקביות פנימית, וממוצעי גלם, לפי שנים

שנה	מס' פריטים	ציון גלם ממוצע	ס.ט. ציון גלם	ממוצע אחוז העונים נכון	מהימנות K-R(20)	ממוצע אחוז העונים נכון	מהימנות K-R(20)
							גירסה מלאה (160)
81	117	-	-	-	-	52.58	0.82
82	125	-	-	-	-	53.22	0.80
83	119	70.27	10.40	58.85	-	56.66	0.86
84	119	75.85	10.45	63.47	0.81	60.52	0.86
85	131	86.92	13.08	65.97	0.87	64.19	0.89
86	132	88.30	11.76	66.48	0.83	62.84	0.85
87	127	89.11	11.64	69.80	0.84	67.88	0.85

ככלל, קיצוץ המבחן מתבטא בירידה מסוימת במידת הקושי שלו ובהקטנת המהימנות שלו. הירידה במהימנות צפויה סטטיסטית מעצם קיצורו. לשאלה, 'האם שיטת הקיצוץ הספציפית פוגעת במבחן יותר מעצם קיצורו' - הרי שלפחות בשלוש השנים האחרונות, נראה שהתשובה שלילית.

מהימנות העקביות הפנימית (K-R 20) של המבחן בשנים האחרונות נעה סביב ערך של 0.85, המורה על מהימנות גבוהה בהחלט, כשמדובר במבחן ידע שאינו הומוגני מבחינת התכנים הכלולים בו.

מהימנות מבחן חוזר נבדקה, בקרב נבחנים חוזרים בשנים שונות. בקרב כל נבחני מכא"ל בשנים 81-86, נבחנו 77 איש יותר מפעם אחת. מתאם (פירסון) בין הישגי הפעם הראשונה לבין הישגי הפעם השניה הוא 0.54 בגירסה המקוצצת, ו-0.60 בגירסה המלאה. אוכלוסית הנבחנים החוזרים הינה חלשה ומקוצצת מלכתחילה: ההישג הממוצע שלהם (בבחינה הראשונה) נמוך ב-0.47 סטית תקן ממוצע נבחני מכא"ל, וסטיות התקן של הישגם קטנה ב-37% מזו של כלל הנבחנים. הפעלת תיקון לקיצוץ תחום עפ"י נוסחת פירסון (1903), מעלה את המתאם בין הבחינות בגירסה המלאה ל-0.77. זהו, אם כן, אומדן למהימנות מבחן חוזר של מכא"ל, וגם הוא אומדן חסר, הן מפאת מגבלות המהימנות הפנימית של מכא"ל, הן משום שלא נעשים מאמצים לבנות את גירסות מכא"ל השונות כמקבילות, לפי ההגדרה המחמירה של נוסחים מקבילים: זהות בדרגות הקושי, כושר ההבחנה, וכיסוי תוכן בין פריטי הגירסאות. סיבה נוספת לפגיעה אפשרית ברמת העקביות הנצפית היא העובדה, שמדובר במבחן הישגים ובנבחנים שניגשים לבחינה פעם נוספת לאחר השקעת מאמץ מצידם להשתנות.

ברמת הקושי הממוצעת של מכא"ל חלה ירידה מסוימת עם השנים, דהיינו, הוא הופך קל יותר. האפשרות שהאוכלוסיה "משתפרת", ולא שהמבחן הופך קל יותר, תיבדק בסעיף 8.3.1, תוך שימוש בשאלות עוגן משותפות בין השנים.

## 8.2 נבחנים חוזרים

ציינו ש-77 איש נבחנו במכ"ל (עד 1986, כולל) יותר מפעם אחת: 72 - פעמיים, 4 - שלוש פעמים, ונבחנת אחת - ארבע פעמים! בספרות הפסיכומטרית מוקדש מקום רב לתופעת השינוי בהישגים כתוצאה מהיבחנות חוזרת. הממצאים המוכרים, ואלה אשר מתייחסים למבחני כושר, (Donlon, 1984), מעידים על שיפור ממוצע של כעשירית עד שליש סטיית תקן בין הבחינה השניה לראשונה. השיפור מוסבר בחלקו בצורה סטטיסטית: מאחר וקבוצת הנבחנים החוזרים הינה מלכתחילה קבוצה שהישגיה נמוכים מהממוצע, הרי שבמצב של מבחן שמהימנותו אינה מושלמת, צפוי מראש שתוצאת המדידה החוזרת תהיה גבוהה מהראשונה, כתוצאה מרגרסיה אל הממוצע. זהו שיפור שאינו תקף, ואינו "אמיתי". סיבה אפשרית נוספת לשיפור, הינה הקטנת אלמנט ה"רעש" במדידה, כתוצאה מהקטנת החרדה, הביטול ואי-הבהירות שמהם סבלו הנבחנים בפעם הראשונה. זהו שיפור תקף במובן זה, שהמבחן החוזר מודד טוב יותר את פוטנציאל ההצלחה העתידית, אך לא "אמיתי" במובן זה שלא חלה עליה אמיתית ביכולת הנמדדת ע"י המבחן. רק חלק מהשיפור מוסבר הן כשיפור "אמיתי", הן כתקף: כאשר הוא תוצאה של תכנית הכשרה המלווה בעליה אמיתית ברמת היכולת או הידע של הקבוצה הנבחנת בשנית, שנערכה בתקופה שבין שתי הבחינות. בהקשר של מכ"ל, שאותו סיווגנו בתחילת הפרק כקרוב יותר למשפחת מבחני ההישג או הידע, היינו מצפים לרמת שיפור גבוהה בין שתי מדידות, משום שמדובר בידיע בר-רכישה, ובנבחנים בעלי מוטיבציה הישג גבוהה, אשר עליה מעיד עצם רצונם לגשת לבחינה בשנית.

לוח 8.2.1 מוצג פרופיל השיפור הממוצע של 77 הנבחנים החוזרים, לפי תחומים, בגירסת "העבודה" של מכ"ל (המקוצצת). השיפור מוגדר כפער הבא: ציון בחינה שניה פחות ציון בחינה ראשונה, בציוני תקן. עבור הציון הכללי, מבוטא הפער גם באחוזונים.

לוח 8.2.1 פרופיל הישגים של נבחנים חוזרים

הפער		ציון תקן פעם II		ציון תקן פעם I		תחום
ס.ת.	ממוצע	ס.ת.	ממוצע	ס.ת.	ממוצע	
0.93	0.092	0.80	-0.115	0.70	-0.233	אישיות
0.86	-0.111	0.71	-0.174	0.73	-0.204	סטטיסטיקה
0.11	0.285	0.97	-0.098	0.63	-0.199	קוגניטיבית
0.91	0.087	0.87	-0.078	0.61	-0.087	חברתית
0.91	0.098	0.91	-0.129	0.66	-0.161	התפתחותית
0.94	0.186	0.79	-0.071	0.63	-0.208	פיזיולוגית
0.72	0.246	0.82	-0.204	0.63	-0.465	ציון כללי
2.45	10.0	2.73	40.2	1.91	30.3	ציון אחוזון



34% מהנבחנים החוזרים ביצעו פחות טוב בבחינה השניה מאשר בראשונה, דהיינו, השיגו "שיפור" שלילי שהגיע עד ירידה ב-2.02 סטיות תקן. רק 66% שיפרו שיפור חיובי, שהגיע עד 1.58 סטיות תקן. קבוצת הנבחנים פעמיים עולה מ-0.47 ס.ת. מתחת למוצע הנבחנים, ל-0.20 ס.ת. מתחת למוצע, דהיינו - שיפור ממוצע של רבע ס.ת.. באחוזונים - הם עולים מאחוזון ממוצע 30, לאחוזון 40. למעשה, זהו אפקט שעשוי לאכזב מעט את המנסים כוחם בשנית: לא רק שסיכויך להשתפר לעומת סיכויך לרדת בביצוע הינם "רק" 2:1, הרי שהשיפור הממוצע הצפוי מעלה אותך 10 אחוזונים בלבד. כאמור, ניתן היה לצפות שבמבחן ידע, השיפור בנסיון שני יהיה משמעותי יותר. מעניין לציין כי מירב השיפור הממוצע הושג בתחומים קוגניטיבית ופיזיולוגית, ובסטטיסטיקה השינוי היה שלילי.

### 8.3 יציבות האוכלוסיות

מדי שנה מכיל המבחן אחוז מסויים של שאלות אשר הופיעו בשנים קודמות. שאלות אלו מכונות "עוגן", והן מאפשרות השוואה אובייקטיבית של רמות האוכלוסיות בין השנים. בלוח 8.3.1 מצויות מספר השוואות כאלה.

לוח 8.3.1 השוואת ממוצעי קושי (אחוזי עונים נכונה) בין שנתונים סמוכים לפי

#### עוגנים משותפים

87-86	86-85	85-84	84-83	83-82	82-81	זוג השנים המשוות
13 59.0	14 61.6	36 65.1	38 59.0	22 52.3	36 58.0	מספר פריטי העוגן ממוצע קושי השנה הראשונה
20.7	18.5	18.2	18.7	18.2	21.2	ס.ת. קושי השנה הראשונה
64.8	65.2	66.4	62.8	51.0	59.2	ממוצע קושי השנה שאחריה
19.5 5.8	17.8 3.6	16.9 1.3	17.6 3.8	16.9 -1.3	22.2 1.2	ס.ת. קושי השנה שאחריה פער בקושי

קיימות טכניקות מורכבות להשוואת סולמות ציונים של מבחנים מקבילים. הכוונה להשוואת ציונים של נבחנים שונים, אשר נבחנו במבחנים שונים, האמורים למדוד את אותם כשרים. מאמצים מתודולוגיים וסטטיסטיים אלו אינם מבוצעים עבור מבחן מכא"ל, משום שבבסיסו הוא משמש מדי שנה לקבלת החלטות אודות נבחני אותה השנה, והצורך להשוואה עם הישגיהם של נבחני שנים אחרות הינו נדיר ביותר. הדמיון בין השנתונים הינו רב, וההמלצה של ועדת מכא"ל לקבוע את ציון מכא"ל כתקף במחלקות למשך שנתיים, מתבססת במידה רבה על דמיון זה, ועשויה להיעזר במימצא של מהימנות מבחן-חוזר (ראה סעיף 8.1). עם זאת, הנתונים המוצגים בלוח 8.3.1 שלעיל, מצביעים על כך שבהשוואה בין כל זוג שנתונים סמוכים, על סמך פריטי העוגן המשותפים (למעט האנומליה של שנת 82) - חל שיפור מה באוכלוסיה הצעירה יותר, שנע בין 1.2 ל-5.8

אחוזי עונים. ללא טכניקות כיוול מורכבות, וללא שימוש במשתני רקע נוספים, לא נוכל להכריע בין שתי האפשרויות הבאות, או להעריך את תרומתן היחסית: א. האוכלוסיה באמת משתפרת משנה לשנה. ב. קיימת העברת אינפורמציה משנה לשנה, והפריטים המשותפים הופכים בעקבות כך למוכרים יותר, ולכן קלים יותר. מחד גיסא, המבחן כולו, כולל פריטים שאינם פריטי עוגן, הופך לקל יותר משנה לשנה, בשיעור דומה לשיעור השיפור בפריטי העוגן. (ראה לוח 8.1.1). עובדה זו מחזקת את השערת השיפור ה"אמיתי" באוכלוסיה. מאידך גיסא - השוואת שונויות אחוזי העונים נכונה על פריטי העוגן בין זוגות השנים, מראה ירידה בפיזור עם ההתקדמות בשנה, ועובדה זו תומכת בהשערת העברת האינפורמציה מדור לדור. ירידה דומה בפיזור העונים אינה מתקיימת בפריטים הרגילים שאינם פריטי עוגן (ראה לוח 8.1.1).

#### 8.4 נתוני תוקף: תוקף בו-זמני, תוקף ניבוי, ותוקף נראה

נגדיר מספר הגדרות של מושג התוקף. כללית, הכוונה היא "עד כמה הכלי מודד את מה שהוא אמור למדוד". מידת התאמת כלי המדידה למטרותיו המוצהרות, ניתנת לשיפוט במספר דרכים, והעיקריות בהן הן:

##### תוקף בו-זמני (concurrent validity)

מתקפים את תוצאות המבחן כנגד קריטריון המהווה מדד ישיר ובלתי תלוי של מה שהמבחן אמור למדוד. הקריטריון קיים באותו זמן בו מועבר המבחן או סמוך מאוד לזמן זה. תוקף בו-זמני משמש בעיקר לעריכת דיאגנוזה של מצב נתון.

##### תוקף חיזוי (predictive validity)

מתקפים את תוצאות המבחן כנגד קריטריון אשר ערכיו מתקבלים כעבור פרק זמן מסוים מאז זמן המבחן. תוקף זה משמש לצרכי מיון וקבלה לעבודה או ללימודים.

##### תוקף נראה (face validity)

תוקף זה אינו מהווה תוקף במובן הטכני; הוא אינו מתייחס למה שהמבחן אכן בודק, אלא למה שנראה על פני השטח, שהוא אמור למדוד. הכוונה ב"תוקף נראה" היא לבחון עד כמה המבחן "נראה" תקף בעיני הנבחנים, מקבלי החלטות על סמך המבחן, ושאר צופים בלתי מקצועיים. מושג זה קשור ליחסי ציבור.

נתוני תוקף בו-זמני, ניתנו לחישוב מאז תחילת הפרוייקט. מדובר בקשר בין ההישגים במכא"ל לבין קריטריון הצלחה בלימודי התואר הראשון - ממוצע הציונים המצטבר בחוג לפסיכולוגיה, עד למועד בחינת מכא"ל. נתון זה מדווח ע"י הנבחנים בשאלון הפרטים האישיים, הממלא על ידם בסמוך לבחינה.

בלוח 8.4.1 מוצגים המתאמים בין מכא"ל על מרכיביו, לבין ציון ה-B.A. הממוצע, לפי שנים. בתחתיתו, מוצגים הממוצעים וסטיות התקן של ציוני ה-B.A.

לוח 8.4.1 - מתאמי פירסון בין מכא"ל לבין ממוצע B.A.

שנה חזאי	83 (N=172)	84 (N=205)	85 (N=231)	86 (N=147)	87 (N=224)
ציון כללי (גירסה מלאה)	0.42	0.56	0.64	0.53	0.58
ציון כללי (גירסה מקוצצת)	0.43	0.53	0.61	0.51	0.59
פירוט (גירסה מקוצצת)					
אישיות	0.39	0.39	0.45	0.28	0.39
סטטיסטיקה	0.26	0.41	0.45	0.28	0.38
קוג', למ', תפ'	0.26	0.37	0.46	0.40	0.43
תברתית	0.27	0.29	0.42	0.28	0.39
התפתחותית	0.34	0.30	0.38	0.37	0.29
פיזיולוגית	0.21	0.44	0.44	0.50	0.52
ממוצע ציוני B.A.	85.1	84.6	85.1	85.6	86.1
ס.ת. ציוני B.A.	3.7	3.7	4.2	3.8	3.7

המתאמים המשמשים לאומדן תוקף בו-זמני של מכא"ל, נעים בין 0.42 ל-0.61, והם בהחלט גבוהים. מובן שאלו הם חסמים תחתונים לקשר האמיתי בין המשתנים, בשל בעיות קיצוץ התחום של שני המדדים, ובשל בעיות המהימנות של ממוצע ציוני ה-B.A., אשר מחושב, באופן שונה בכל מוסד, מדווח וולונטרית ע"י הנבחנים, ואינו סטנדרטי ותקני לא רק בקרב מורים של אותו מוסד, אלא גם אצל מורה בודד, במקרים רבים.

נתוני תוקף הניבוי נאספו לראשונה בשנת 1986, במסגרת עבודת גמר לתואר שני של גב' ג'קי בטל, מאוניברסיטת חיפה. כקריטריונים עיקריים לתיקוף, שימשו המדדים הבאים: א. ממוצע הקורסים ב-M.A. (סולם 0-100). ב. ציון גמר סופי ב-M.A. (כולל תיזה, סולם 0-100) ג. הערכת מנחים בפרקטיקום (סולם 1-6). ד. ציון עבודת גמר (תיזה, סולם 0-100). כחזאים שימשו ציוני מכא"ל הכלליים וששת מרכיביהם (בציוני תקן, בגירסת מכא"ל המקוצצת) וכן ציוני ה-B.A. של הנבחנים (בסולם 0-100). להבדיל מציון ה-B.A. מלוח 9.4.1, אשר דווח ע"י הנבחנים (לא כולם), ולא היה סופי, הרי שציון ה-B.A. בלוח 9.4.2 הינו ציון סופי ורשמי, כפי שדווח ע"י המחלקות לפסיכולוגיה. הנתונים נאספו מארבע אוניברסיטאות, מקרב כ-250 תלמידי M.A. בפסיכולוגיה שנבחנו במכא"ל בשנים 1983 ו-1984. מובן שלא כל הקריטריונים היו קיימים עבור כל התלמידים. גם הקריטריונים, וגם ציון ה-B.A. הממוצע, סובלים

מחוסר אחידות בדרך חישובם במוסדות השונים. עם כל המיגבלות הללו, דווחו התוצאות הבאות:

ליח 8.4.2 - מתאמי פירסון בין חזאים לקריטריונים. בשוליים - ממוצעים וסטיות תקן.

קריטריונים חזאים	ממוצע קורסים M.A.	ציון גמר M.A.	ציון עבודת גמר	שכיחות כללית (n)	ממוצע תקן	סטיית תקן
ציון מכא"ל כללי	.36	.60	.32*	254	0.38	0.79
1. אישיות	.23	.49*	.35*	"	0.29	0.58
2. סטט' ומחקר	.26	.36*	.29*	"	0.32	0.82
3. קוג', למ', תפ'	.22	.64	.26*	"	0.29	0.86
4. חברתי	.22	.51	.25*	"	0.26	0.51
5. התפתחותית	.29	.63	0	"	0.31	0.50
6. פיזיולוגית	.17	.19*	.26*	"	0.27	0.70
ממוצע ב.א.	.44	.53	.24*	237	85.56	3.59
שכיחות כללית ממוצע סטיית תקן	251 85.37 3.70	14 89.00 3.23	21 90.45 4.35			

\* - לא מובהק,  $p > 0.05$

תוקף הניבוי של מבחן מכא"ל עומד על 0.36, כנגד ממוצע הקורסים ב-M.A. תוקף הניבוי של ממוצע ה-B.A., כנגד אותו קריטריון, הינו 0.44. יש לזכור את מגבלת אי-האחידות של הקריטריון, המטילה חסם עליון על תוקף הניבוי. כמו כן, מדובר במצב של קיצוץ תחום החל על קבוצת המחקר מלכתחילה: זוהי קבוצה ברמה גבוהה והומוגנית, (גבוהים ב-0.38 סטיות תקן מאוכלוסית כל נבחני מכא"ל) ופיזור הישגיהם קטן ב-20 אחוזים מזה של כל הנבחנים. הישוב המתאם של ציון מכא"ל הכללי עם ממוצע ציוני הקורסים ב-M.A. לפי נוסחת פירסון (1903) לקיצוץ תחום, נותן מקדם תוקף מתוקן שערכו 0.44.

תוקף הניבוי של מכא"ל כנגד ציון גמר ה-M.A. העומד על 0.60, הינו גבוה ביותר, אולם מפאת מיעוט הנבחנים שעליו הוא מבוסס (14 איש) - ראוי לסייג את מידת הביטחון במידת יציבותו. עוצמת הניבוי של מכא"ל ושל ממוצע ה-B.A. לגבי הערכות המנחים בפרקטיקום הינה אפס.

השימוש בצירוף של מכא"ל וממוצע ה-B.A. נותן תוצאה טובה מכל אחד מהחזאים בנפרד: מתאם מרובה (במשקלות אופטימליים) של 0.46 (לעומת 0.44 של ממוצע ה-B.A. בנפרד, ו-0.36 של ציון מכא"ל הכללי בנפרד).

תוצאות אלו מתיישבות עם מימצאים דומים בספרות, והן אף מפתיעות לטובה, בהתחשב בציפיות הנמוכות, שנבעו הן מהפיזור הנמוך של ציוני הקריטריון והחזאי, הן מהמהימנות הנמוכה של הקריטריון כתוצאה מריבוי השיטות לחישובו במחלקות השונות. מדריך ה-GRE, המוצא לאור ע"י ה-ETS (84-1983), מדווח בעמ' 33, על מקדם תוקף חציוני של 0.30, שנמצא מעבר לשמונה קולג'ים, בין השנים 1952-1981, עבור מבחני "נושא" (subject tests) בניבוי דירוג הסטודנטים במחלקות. פסיכולוגיה הוא אחד ממבחני ה"נושא", והוא מקביל בשימושו למכא"ל. כאשר הקריטריון המנובא הוא ציון בחינת הגמר במחלקה, עומד התוקף החציוני (ב-2 קולג'ים) על 0.48. באותו פירסום, תוקף החיזוי החציוני של ציוני ה-B.A., לגבי דירוג התלמידים במחלקות (ב-30 קולג'ים), הינו 0.35.

נתוני התוקף הנראה (face validity), מקורם בשני סוגי שופטים, המצויים משני צידי המיתר: מצד אחד, המיסד המשתמש בכלי לצורך הברירה והמיון ללימודי ה-M.A. במחלקות לפסיכולוגיה בארץ, והכוונה היא למורים, ראשי החוגים, ונציגי המחלקות בוועדות מכא"ל מדי שנה. בעיני מרביתם, וכן בעיני כותב פרק זה, נתפס המבחן כבעל רלוונטיות מיידית ובלתי מוטלת בספק למטרה שלה הוא נועד: הוא בודק הישגים וידע בתחומים שהבקאות בהם היא היעד שהציבו לעצמם המחלקות לפסיכולוגיה ביחס לתלמידיהם. לימודי המוסמך נועדו גם הם בעיקר לאותה מטרה - הקניית והעמקת הידע באותם תחומים. לכן, מיון או קונקורס באמצעות כלי כמכא"ל הינו הוגן ביותר. מובן שתמיכה לעמדה זו מתקבלת מהנתונים האמפיריים של תוקף הניבוי והתוקף הבו-זמני שנסקרו לעיל. מצידו השני של המתרס נמצאים נבחנים - "קורבנות" השיטה. דעותיהם בנושא זה, ובנושאים אחרים הנוגעים למבחן, נבדקו באמצעות שאלוני משוב, שהועברו לנבחני השנתיים האחרונות. תוצאות שאלון המשוב משנת 1987 מוצגות בלוח 8.4.3.

כ-230 איש ענו על שאלון זה.

לוח 8.4.3 - משוב לנבחן (1987): התפלגות תשובות באחוזים, ממוצעים ו-ס.ת.

שאלה	תשובה	במידה רבה מאד	במידה רבה	במידה בינונית	במידה מועטה	כלל לא	ממוצע	סטיית תקן
		5	4	3	2	1		
	<u>באיזו מידה:</u>							
	1. התכוננת למבחן	21.6	50.2	21.6	5.6	0.9	3.9	0.85
	2. המבחן נראה לך קשה	18.1	45.7	36.2	-	-	3.8	0.72
	3. מצאת עניין בפתירתו	4.5	23.9	45.0	22.5	4.1	3.0	0.90
	4. המבחן הרגיש או תסכל אותך	7.6	17.9	35.9	30.5	8.1	2.7	1.05
	5. הצלחת במבחן	1.4	10.6	74.3	13.3	0.5	3.0	0.56
	6. נראה לך המבחן כלי מתאים לבדיקת הידע שלך בפסיכולוגיה	0.9	13.2	38.2	38.6	9.1	2.6	0.86
	<u>באיזו מידה, לדעתך, מתאימות שיטות המיון שלהלן לקבלת תלמידים ללמודי המוסמך בפסיכולוגיה:</u>							
	7. בחינת ידע בפסיכולוגיה	5.0	27.3	45.0	20.5	2.3	2.9	0.87
	8. ראיון אישי	33.8	43.7	16.2	5.4	0.9	4.0	0.89
	9. ממוצע ציוני B.A.	13.5	45.0	38.7	2.3	0.5	3.7	0.75
	10. המלצות	15.8	35.6	26.6	18.5	3.6	3.4	1.07
	11. מבחן מצב (משחקי תפקיד, סימולציות)	16.7	29.0	27.6	17.1	9.5	3.3	1.20

יש לזכור שאת שאלון המשוב מילאו התלמידים בסמיכות רבה למועד היבחנם במכ"ל, בעודם נתונים במתח ובלחץ של הבחינה, ובטרם ידועות להם תוצאות המבחן. בתנאים אלה, מכל כלי המיון לתואר השני המוצעים לתלמידים, מבחן הידע בפסיכולוגיה "נהנה" מהפופולריות הפחותה ביותר. יתרה מזו - מכ"ל לא נראה לנבחנים כלי מתאים במיוחד לבדיקת הידע שלהם בפסיכולוגיה, אף ללא מטרות מיון! הראיון האישי נראה בעיניהם מתאים ביותר למיונם לתואר השני, ואפילו המלצות ומבחני מצב עדיפים בעיני הנבחנים על פני מבחן ידע. מימצא זה עומד בסתירה עם נתונים אמפיריים מוכרים, המעידים על כך שמבחני הידע עדיפים מבחינת נתוני מהימנות ותוקף חיזוי, בהשוואה למדדים הפחות אובייקטיביים החליפיים שהוזכרו. עם זאת, התנגדות התלמידים למבחן אינה מוחלטת:

רק כ-23% מכלל המשיבים, סבורים כי בחינת ידע בפסיכולוגיה מתאימה במידה מעטה, או כלל לא, למיון תלמידים לתואר השני. ובאשר ליתר מימצאי המשוב: כ-72% התכוננו למבחן במידה רבה ורבה מאוד, לכ-64% נראה המבחן קשה או קשה מאוד. כ-28% מצאו ענין רב ורב מאוד בפתירתו, ורק 26% מהנבחנים תוסכלו או חשו רוגז רב ורב מאוד בעת פתירת מכא"ל. כ-12% בלבד יצאו מן המבחן בהרגשה שהצליחו בו במידה רבה ורבה מאוד. מידת ה"ענין" שמצאו בפתירתו, גבוהה ממידת ה"תסכול ורוגז" שחשו בעת פתירתו.

#### 9. השוואה עם פרויקט מקביל - G.R.E - פסיכולוגיה

לסיום פרק זה, ראוי להשוות את מכא"ל עם תוכניות מקבילות בעולם. שירות המבחנים החינוכיים האמריקאי (E.T.S) מקיים פרויקט מבחני ידע, המוצעים על ידו, יחד עם קריטריונים נוספים כגון מבחן כושר השיבה, ככלי למיון מועמדים ללימודים אקדמיים מתקדמים (לתואר שני ומעלה) ולחלוקת מילגות הצטיינות לתלמידים. מערכת זו נקראת G.R.E, ובין יתר הנושאים במבחני הידע קיים גם נושא הפסיכולוגיה, אשר יקרא להלן "G.R.E - פסיכולוגיה". למבחן זה ניגשים מדי שנה כ-15,000 מועמדים ללימודי תואר שני ושלישי בפסיכולוגיה בארה"ב: 59% מתוכם נשים ו-41% גברים. במכא"ל - 72% מהנבחנים עד היום הן נשים, ו-28% גברים. "G.R.E - פסיכולוגיה" מקביל למכא"ל מבחינת תוכנו, אופן שימוש, ואוכלוסיית היעד שלו. למעשה, זהו המבחן הידוע והרווח ביותר בסוגו.

השוואת מכא"ל עם "G.R.E - פסיכולוגיה" חייבת לצאת מתוך נקודת מוצא פרואית ולכאורה שלא-ממין-העניין - הנקודה הכלכלית. הבדלי התקציב האדירים שבין שני הפרוייקטים, מתבטאים בהכרח בהבדלים ברמה המקצועית, ואם לא במוצר הסופי, הרי לפחות באופן שהדברים צריכים להעשות. דמי הבחינה המשולמים ע"י הנבחנים, גם עבור מכא"ל וגם עבור "G.R.E - פסיכולוגיה" עומדים על כ-\$30. התקציב השנתי של מכא"ל, הינו לפיכך כ-\$8,000. תשבון גס מצביע על תקציב אמריקאי שנתי משוער של כ-\$450,000 (רק של נושא הפסיכולוגיה!!!).

במבחן ה-"G.R.E - פסיכולוגיה" יש 200 שאלות רבות ברירה, להן מוקצבות 170 דקות; במכא"ל 160 שאלות, ולהן מוקצבות 180 דקות! ה-"G.R.E - פסיכולוגיה" נחלק לשלושה אשכולות של נושאים:

הראשון - אשכול מדעי הטבע והניסוי, שמחציתו עוסקת בלמידה, קוגניציה ותפיסה, ומחציתו השניה עוסקת באתולוגיה ופסיכולוגיה השוואתית, תחושה ופיזיולוגיה. כל אלו מרכיבים את תת הציון המכונה "פסיכולוגיה ניסויית".  
אשכול שני - מדעי החברה. פריטי קבוצה זו נחלקים שווה בשווה בין תחומי אישיות, קלינית ואבנורמלית, התפתחותית, ופסיכולוגיה חברתית. תחומים אלה מרכיבים את תת-הציון "פסיכולוגיה חברתית".

אשכול שלישי הינו כללי. בו כלולים נושאים כמו הסטוריה של הפסיכולוגיה, פסיכולוגיה יישומית, מדידה וסטטיסטיקה. פריטים אלה תורמים לציון הכללי, אך אינם מרכיבים תת-ציון משל עצמם. כל תת-ציון מהנזכרים לעיל מבוסס על כ-80 פריטים. זהו אורך מספיק עבור מבחן שמהימנותו אינה נופלת מ-0.80. ששת הסולמות החלקיים של מכא"ל אינם מהימנים דיים כדי לאפשר בהם שימוש דיפרנציאלי. אם רוצים שימוש כזה - חייבים לאחד גושים של תחומים, ולא להרבות תת-חלקים, כמוצע ע"י גורמים שונים במסודות החברים במכא"ל

תוכן ומבנה המבחן ב-G.R.E נקבעים ע"י וועדת מומחים ממספר אוניברסיטאות הנבחרים לתפקידם לשנתיים עם אופציה להארכה. החלפתם מאפשרת מחשבה מחודשת באשר להרכב ולתכנים במבחן ורביזיה כזו אמנם נעשית על בסיס תקופתי. בניית המבחן והגהתו הקפדניים נערכים ע"י מומחים בתחומם, הנעזרים במומחים לבניית מבחנים מ-E.T.S. על פני הדברים, דומה התהליך המתואר לזה המתרחש במכא"ל. אולם ההבדל הגדול נעוץ בעובדה שמאגר תורמי השאלות במכא"ל מצומצם ביותר, והוא מנוצל לעיפה מדי שנה. יתרה מזאת - הם עושים זאת ב"התנדבות", כמו גם תיפקודה ה"התנדבותי" של וועדת מכא"ל. במכא"ל אין רביזיה תקופתית של תוכן ומבנה המבחן, ואין עבודה מספיק מעמיקה, בתשלום, עם מומחים בתחומי הפסיכולוגיה והפסיכומטריקה.

בנושא פסילת הפריטים - המודל האמריקאי אינו משתמש בטכניקה כגון זו הנהוגה במכא"ל, לא רק מפני שמדובר בהטייה אפשרית בין מאות מוסדות, אלא מפני שהגדרת שדה התוכן של המבחן נעשית ב א-פריורי, כחלק מתפקידה של הוועדה.

הבדל נוסף בין הפרוייקטים נעוץ במאמץ המתודולוגי המושקע ב-G.R.E לכייל את ציוני המבחנים בין הגירסאות השונות, והדאגה לכך שהנוסחים יהיו "מקבילים", במובן המחמיר. טכניקה זו הכרחית במצב בו קיימת יותר מגרסה שנתית אחת, כפי שמתקיימות ע"י G.R.E, וכן במצב של שינוי הדרגתי בתוכן המבחן, כתוצאה משינוי הלכי הרוח בתכניות לימודים. גם נפח הנבחנים השנתי הגדול ב-G.R.E, וכן ריבוי המחלקות לפסיכולוגיה והשונות הרבה ביניהן, מחייבים כיוול זה.

מכך אין משתמע שמכא"ל פטור עקרונית מהצורך לכייל בין נוסחים. מעשית, נתקל הפרוייקט במספיק קשיים בבניית כל נוסח ונוסח, ונראה כי במצבו הנוכחי אינו יכול להוסיף אילוצים המשתמעים מהכיוול בין נוסחים.

המאפיינים הסטטיסטיים של "G.R.E - פסיכולוגיה" עולים באיכותם במעט על אלו של מכא"ל: מהימנות הסוללה כולה המדווחת (במדריכים המצורפים לחומר ההרשמה למבחן) הינה 0.94 ומהימנות תת החלקים הינה כ-0.88. כזכור, מהימנות מכא"ל הינה כ-0.85. חלק מההבדל במהימנות נעוץ בהבדלים באורך: ממכא"ל נשארים אחרי קיצוצו כ-130 פריטים, וב-G.R.E - כ-200. חלק נוסף מההבדל במהימנות, מקורו בקיצוץ התחום המשוער שחל ביכולתה של אוכלוסיית נבחני מכא"ל, לעומת נבחני "G.R.E - פסיכולוגיה": ניתן לומר בבטחון כמעט מוחלט, שרמת היכולת של נבחני מכא"ל גבוהה מזו של כלל נבחני "G.R.E - פסיכולוגיה", ופיזורם קטן יותר, משום שהראשונים עברו סלקציה חמורה כבר בלימודי התואר הראשון שלהם, והאחרונים - הינם באגרי תואר ראשון בכל התחומים, ולא דווקא של חוגים סלקטיביים. אולם אין ספק שהעדיפות באיכות ה-G.R.E מוסברת גם במבחר הרב יותר של פריטים איכותיים שעומד לרשות בוני המבחן, ולמאמצים המקצועיים הממושכים והיקרים המושקעים בבנייתם.



מפתיעה העובדה שאין עדות לעדיפות ה-"G.R.E. - פסיכולוגיה" על פני מכא"ל מבחינת נתוני התוקף! (ראה סעיף 8.4), והרי זהו בעצם הקריטריון העליון להערכת איכותו של מבחן.

מכא"ל יכול לסגל לו כמה מהיתרונות הטמונים בשיטה האמריקאית, בעיקר במה שקשור באופן פעולת הוועדה הממונה, והגדרות התוכן שמתפקדה לבצע. בהינתן המשאבים הכספיים והמקצועיים העומדים לרשותו - עומד מכא"ל בכבוד בהשוואה עם ה-"G.R.E. - פסיכולוגיה".

## 10. סיכום ודיון

התמודדות מכא"ל עם הקושי שבהגדרת עולם התוכן במבחן כפי שתואר בסעיף 4, מבליטה מספר בעיות, שמקורותיהן ברמות שונות:

א. בעיית הגדרת הפסיכולוגיה את עצמה, כדיסציפלינה. ימיה של בעיה זאת הינם, מן הסתם, כימי הפסיכולוגיה עצמה. מכא"ל מציב אתגר מעשי לבעיית ההגדרה העצמית של הפסיכולוגיה, וממחיש את הקושי הטמון בכך.

ב. בעיית הקושי שביישום מבנה תאורטי מושכל כמשפט המיפוי שהוצג לעיל, על בנייתם המעשית של פריטים למבחן. הדרך שנתגלתה כיעילה למיפוי שדה הפסיכולוגיה וליישום המיפוי בבניית השאלות - היא באמצעות "תפריט", או "רשימת מצרכים" פשטנית, בלתי-מימדית, בעלת מבנה של אשכולות.

ג. בעיית הקונפליקט בין הזיקות המשמשות במכא"ל בערבוביה: הזיקה לנורמה מול הזיקה לקריטריון. כאן יש צורך בהכרעה. בחירת השאלות במכא"ל מושפעת מהכרעה זו: במבחן נורמטיבי (norm referenced) אין אילוצים כה תמורים על אופי השאלות כבמבחן הישגים טהור, שהינו criterion referenced.

ד. בעיית ההסכמה בין המוסדות החברים במכא"ל, באשר לתוכן המבחן. מאחר שההסכמה נקבעת בעיקר אמפירית, בהסתמך על ניתוח שלאחר מעשה של הבדלי הביצוע בין בוגרי המוסדות השונים בכל פריט, הרי שמעט מאוד מאמצים מוקדשים לניסוח הקונצנזוס לפני בניית המבחן. הפועל היוצא מכך הוא, שהגדרת התוכן נעשית מעגלית, והיא בבחינת שיקוף של הלכי רוח שיש בהם אלמנט של מקריות, וחסר בה תכנון מוקדם. תכנון פעיל כזה עשוי להתרחש מתוך זיקה לתכנון מקביל של תכניות הלימודים ב-B.A. מדובר, כמובן, גם בהיבט הדינאמי של קביעת מבנה ותוכן הבחינה. רצוי שהוועדה תדון ותעייין בתפריט מכא"ל מדי שנה, תוך הצהרה על שינויים במידה שהחלטות על-כך מתקבלות.

ה. בעייתם של הסטודנטים המתכוננים לקראת הבחינה: הגדרת נושאי המבחן הינה, לגביהם עמומה, בלתי ספציפית ובלתי מחייבת. ההבטחה הניתנת להם בדבר המנגנון של פסילת שאלות אשר מעידות על כך שכקבוצה, בוגרי מוסד כלשהו מופלים לרעה - הינה בחזקת אליה וקוץ בה. הסטודנט נקלע למצב פאראדוכסלי, דמוי דילמת האסיר: הוא מתמודד בסיטואציה תחרותית גם עם בני מוסדו. אין לו ודאות שחבריו לא ילמדו בעצמם נושאים שכקבוצה הם לא למדו, ומחמת הספק ומחמת התחרותיות - הוא נדחף להרחיב את מעגל הקריאה והנושאים הנלמדים, מעבר לאלה המזוהים על ידו ככלולים ב"חומר". גם ההגדרה של נושא ככלול ב"חומר" או לא כלול ב"חומר" הינה עמומה: אולי בשנים מסוימות הוא נלמד במוסד מסוים, ובשנים אחרות, באותו מוסד - לא נלמד. אין בהכרח הומוגניות במסלול הלימודים אפילו בקרב בני מחזור אחד, במחלקה אחת. המצב המתואר מעורר מרירות בקרב חלק מהסטודנטים, והוא תורם להתנגדותם הבסיסית (והמובנת) לאוסף המכשולים הנערמים בדרכם המקצועית.

הבעיות הנזכרות מצביעות על צורך במספר שינויים באידיאולוגיה המנחה את הפרוייקט, ובדרכי פעולתה של ועדת מכא"ל. עקרונית, השינויים יהיו בכיוון של ריכוז מאמצים ליצור גוף מקצועי בין-אוניברסיטאי, שיביא לצמצומו (עד ביטול) של מנגנון ההגדרה הרטורואקטיבית של תוכן הבחינה. הגוף המקצועי יהיה מורכב מצוות מוביל בכל תחומי הפסיכולוגיה אשר יבצע את מרבית העבודה מראש: הגדרת מבנה המבחן והרכב תחומיו, הענקת משקלות לתחומים השונים, שינויים שנתיים בהדגשים השונים תוך שיקוף שינויים בהלכי הרוח במחלקות, ואישור מירבי של פריטי המבחן כנגד קריטריונים שנקבעו מראש. במצב זה לא יהיה צורך בפסילת פריטים בדיעבד, אלא במקרים נדירים ביותר של כשלונות סטטיסטיים. מובן שאבן הנגף העיקרית לקיומו של מנגנון זה הינה הצורך לממנו.

על "מפת" סוגי המבחנים נמצא מכא"ל קרוב למבחני ההישגים או הידע (achievement tests). אולם אופן השימוש במבחן והאילוצים המלווים את הפרוייקט מייחדים אותו במספר הבדלים ממבחני ההישגים הטהורים, תוך יצירת תכונות כלאיים ייחודיות. ההישגים המדווחים במכא"ל הינם יחסיים, דהיינו, מספקים מידרג של המועמדים הנבחנים בשנה מסוימת. אין כמעט התייחסות אל רמת הידע המוחלטת שההישגים במכא"ל משקפים. בכך, שלא כמרבית מבחני הידע, אין מכא"ל משמש למדידת האפקטיביות של תכנית לימודים - במקרה זה, לימודי התואר הראשון. מכא"ל נשפט יותר מתוך זיקה לנורמה, ופחות (שלא כמרבית מבחני הידע) מתוך זיקה לקריטריון. מבחן הישגים קלאסי אינו משמש בדרך כלל כחזאי לביצוע עתידי, אלא כקריטריון בסופו של תהליך חינוכי, המשמש למדידת יעילותו. לפיכך, איכותו תימדד יותר ע"י תקפות התוכן והמבנה, ופחות ע"י תקפות הניבוי. מכא"ל, לעומת זאת, ראוי שיקיים בו-זמנית את שתי האיכויות: תקפות החיזוי - משום היותו כלי ממיין; ותקפות התוכן - משום היותו בנוי מעולם תוכן בעל רלוונטיות מיידית למשימה שאליה הוא ממיין - הצלחה אקדמית בלימודי הפסיכולוגיה.

באילוץ של פסילה בדיעבד של הפריטים המוטים לרעת אחד או יותר מהמוסדות החבריים בפרוייקט, יש ויתור מראש הן על מקצת מתקפות התוכן, הן על מקצת מתקפות החיזוי, המתווכת באמצעות מהימנות המבחן, אשר נפגעת מקיצוץ. זהו כנראה הייחוד ה"ייחודי" ביותר של מכא"ל בקרב משפחת מבחני ההישגים: הגדרה מעגלית של עולם התוכן המכוסה ע"י המבחן; בבחינת "הפסיכולוגיה הינה מה שמלמדות במשותף כל המחלקות לפסיכולוגיה בארץ".

לסיום, ראוי לדון בנקודה חשובה, הנוגעת לביקורת אפשרית כלפי המבחן. מדובר בהצהרת הכוונות שלו:  
מכא"ל אינו מתיימר לנבא איזהו הפסיכולוג הטוב. מטרתו ויתרונותיו הינם בתחומים הבאים:

א. מכא"ל הוא כלי אשר מתמודד בעיקר עם המימד הסכולסטי של ההצלחה המקצועית. ככזה, יתרונו הגדול באחידות הסולם שאותו הוא מספק, המשוחרר מתלות ברקע האקדמי הייחודי של הנבחן בו. השימוש במכא"ל מייצג נאמנה את הגישה על-פיה מתמקדת האקדמיה בתפקיד של הקניית ידע, כתנאי הכרחי (ואולי לא מספיק) להצלחה מקצועית. ההצלחה במדידת או אף הגדרת התכונות האישיות האחרות הנחוצות או מנבאות הצלחה מקצועית - אינה רבה, ובמקרים רבים, מוותרת האקדמיה על נסיונות אלה. עם זאת, רצוי (ואף נהוג) להשתמש במכא"ל בנוסף לשורה של מדדים רלוונטיים נוספים. ההחלטה בדבר המשקל היחסי שינתן למכא"ל בתהליך קבלת ההחלטות הינה ערכית בעיקרה, והיא יכולה להעזר אך מעט בנתוני מחקר. הסיבה לכך היא שהגדרת הקריטריונים להצלחה במשימה שאליה בורר המבחן, הינה בעיקרה ערכית.

ב. מכא"ל נהנה מתוקף נראה (face validity) שאין רבים (שלא מקרב הנבחים) המערערים עליו. גם אילו נמצא שמדדי התוקף האחרים שלו אינם גבוהים ביותר, הרי שבמצב של ברירת מועמדים ובהנחה של אי-המצאותם של כלי מיון חליפיים טובים יותר, קיימת מידה רבה של הוגנות בהצבת מכשיר ממין אשר לו מידת רלוונטיות מיידית למשימה העתידה - לפחות בהגדרה צרה שלה, קרי - הצלחה בלימודי המוסמך.

ג. מכא"ל יכול לספק מדד לגבי היקף ורמה מינימליים של תכניות הלימודים במחלקות השונות ולהבטיח רמת מוכנות אקדמית מינימלית לקראת לימודי התואר השני.

ד. ולבסוף, מכא"ל עשוי לעזור במיון, יעוץ והכוונה עצמית של המועמדים, ולא רק בברירתם. ניתן להיעזר בסולמות החלקיים שמכא"ל מספק בסיווג לרמות שונות (בסטטיסטיקה למשל), או כאינדקסיה לנבחן לגבי נקודות החוזק והחולשה שלו. אולם הסתייעות זו חייבת להיעשות בזהירות ובהסתייגות רבה, משום שהסולמות החלקיים בהם מדובר מבוססים על חלקי-בחינה קצרים, ומהימנותם אינה גבוהה.

11. נספח: משתני רקע דמוגרפיים של נבחני מכא"ל

הטבלאות הבאות מציינות פרופיל לתאור אוכלוסיית הנבחנים במכא"ל, בעזרת נתונים הנאספים מהם בשאלון פרטים אישיים הצמוד למבחן. פרט לנתוני רקע, מצויינות גם העדפותיהם האקדמיות. כאשר חסרו נתונים, סומנו התאים ב" - ". המידע הסטטיסטי המוצג בסעיף זה אינו חיוני לקורא אשר עיקר עניינו בצדדים העקרוניים יותר של מכא"ל.

11.1 התפלגות נבחנים שנתית, במספרים מוחלטים, לפי מקום בחינה (אין זהות הכרחית בין מקום הבחינה, לבין המוסד בו למד הנבחן תואר ראשון)

מקום	שנה	81 *	82 *	83 *	84 *	85	86	87
בר-אילן	69					77	59	53
חיפה	51					34	55	43
ירושלים	33					46	47	41
תל-אביב	100					91	82	88
בן-גוריון	-					32	27	20
מועד מיוחד	12					8	6	7
סה"כ	265	291	212	271	288	276	252	

\* אין נתונים משנים אלה.

המספרים יציבים, ונעים בין 250 ל-290 (למעט 212 ב-82).

11.2 מין הנבחנים: אחוז נבחנים שנתי

מין	שנה	81	82	83	84	85	86	87
נקבה	74.1	73.1	73.1	71.0	67.5	74.5	72.5	
זכר	25.9	26.9	26.9	29.0	32.5	25.5	27.5	
לא צויין	-	(1)	-	(2)	(5)	(9)	(1)	

הרכב המינים גם הוא יציב, ביחס של 3:1 לטובת הנשים.

11.3 גיל לפי טווחים: אחוז נבחנים שנתי

87	86	85	84	83	82	81	שנה טווח גילים
0.4	-	-	-	0.9	1.1	1.6	19-20
6.6	6.9	8.5	7.5	6.6	7.6	11.0	21-22
32.8	38.0	29.0	30.5	34.2	34.5	29.8	23-24
39.4	33.5	32.1	37.2	33.1	33.0	30.6	25-27
15.0	13.4	20.1	15.2	16.6	15.5	16.8	28-32
5.7	8.3	9.5	7.8	7.6	8.3	10.2	33-52
(8)	(16)	(5)	(2)	(1)	(13)	(8)	לא צויין
25.9	26.2	26.8	26.5	26.5	26.3	26.3	ממוצע

הגיל הממוצע יציב מעבר לשנים, והוא 26 וכמה חודשים.

11.4 מקום לידת הנבחן: אחוז נבחנים שנתי

87	86	85	84	83	82	81	שנה מקום
84.0	78.8	79.0	82.5	78.2	77.5	74.9	ישראל
2.4	4.4	2.4	3.7	5.2	2.8	3.0	אסיה-אפריקה
13.6	16.8	18.5	13.8	16.6	19.7	22.1	אירופה, אמריקה
(2)	(2)	(2)	(2)	(1)	(2)	-	לא צויין

מרבית הנבחנים הינם ילידי הארץ (בין 75 ל- 84 אחוז), ושיעורם בקרב כל הנבחנים עולה בהדרגה. שיעור ילידי אסיה-אפריקה הינו נמוך מאוד - בין 2.5 ל- 5 אחוז.

11.5 מקום לידת אבי הנבחן: אחוז שנתי

87	86	85	84	83	82	81	שנה מקום
30.2	22.1	24.3	26.5	17.9	18.6	13.4	ישראל
17.8	19.0	13.6	13.7	12.8	10.7	11.9	אסיה, אפריקה
52.1	58.8	62.1	59.8	69.4	70.7	74.7	אירופה, אמריקה
(10)	(50)	(45)	(52)	(16)	(1)	(2)	לא צויין

מוצא ההורים משתנה עם השנים: עולה פרופורצית ילידי אסיה/אפריקה, יחסית ליוצאי המערב. זוהי עובדה משמחת, ויש לקוות שהיא משקפת מגמה דמוגרפית כללית בהשכלה הגבוהה, ואינה ייחודית למקצוע הפסיכולוגיה... העליה בשיעור ההורים ילידי הארץ הינה צפויה לאור "הזדקנות" המדינה.

11.6 שפה מדוברת בבית ההורים: אחוז שנתי

שנה	87	86	85	84	83	82	81
עברית	85.8	83.5	80.6	84.4	79.4	79.2	71.9
אנגלית	4.1	2.9	4.2	4.4	6.6	4.9	4.6
ערבית	1.6	1.1	0.7	2.2	1.5	3.5	1.9
יידיש*	2.8*	0.4	0.4	0.4	0.5	1.1	1.5
אחרת	5.7	12.1	14.1	8.5	11.8	11.3	20.2
לא צויין	(6)	(4)	(4)	(1)	(8)	(7)	-

\* ב-87 - ספרדית

בהתאמה עם ממצאי לוחות 11.4 ו-11.5, עולה שיעור יוצאי הבתים דוברי העברית. ראוי לתשומת לב שיעורם הקטן והיציב של דוברי הערבית!

11.7 השכלת אב באחוזים, לפי שנים

שנה	87	86	85	84	83	82	81
יסודית	12.6	13.2	13.9	12.6	13.7	15.3	15.8
תיכונית חלקית	13.4	16.5	22.0	22.7	16.6	16.7	17.3
תיכונית/על-תיכונית	18.7	25.6	20.9	21.9	20.9	32.3	35.0
אקדמית חלקית*	10.6	7.7	9.8	6.7	9.0	-	-
אקדמית מלאה	44.3	37.0	33.4	36.1	39.8	35.8	31.9
לא צויין	(7)	(3)	(1)	(2)	(1)	-	(3)

\* קטגוריה זו לא היתה קיימת ב-1981 וב-1982.

השכלת ההורים עולה עם השנים: מתרבה בקרב האבות שיעור בעלי השכלה אקדמית מלאה, ומצטמצם שיעורם של היתר.

11.8 מועד סיום לימודים קודמים בפסיכולוגיה, באחוזים, לפי שנים

87	86	85	84	83	82	81	שנה מצב אקדמי
17.1 74.8	18.5 73.0	19.1 73.6	19.1 70.0	19.9 71.8	23.8 67.1	33.8 58.9	סיימו B.A. יסימו השנה
0.8	0.7	1.0	1.5	0.5	0.4	1.1	B.A. סיימו לימודי השלמה
6.9	5.9	5.2	9.4	6.8	5.4	4.9	M.A. יסימו השנה
0.4 (6)	1.9 (6)	1.0 -	- (4)	1.0 (6)	3.2 -	1.1 -	M.A. השלמה אחר לא צויין

המצב האקדמי של נבחני מכ"ל יציב על פני השנים (למעט שנת 81' שהיתה ראשונה ונסיונית) ומגלה דפוס כדלקמן: כ-17% סיימו B.A., כ-75% נבחנים נמצאים בשנה האחרונה ללימודי ה-B.A., וכ-6-7 אחוז לומדים לימודי השלמה, ומקורם בחוגים אחרים.

11.9 המוסדות שאליהם ביקשו להעביר תוצאות הנבחן, במספרים מוחלטים, לפי שנים (החל מ-1987, העברת התוצאות למוסדות הינה אוטומטית, למעט כאשר מובעת התנגדות להעברתן למוסד כלשהו, אותה צריך הנבחן לסמן בעת היבחן)

86	85	84	83	82	81	שנה מוסד
63	67	69	45	-	-	בן-גוריון
201	240	207	172	202	189	בר-אילן
137	162	145	114	135	121	חיפה
205	249	213	174	201	190	תל-אביב
153	175	192	130	122	108	ירושלים
16	-	-	-	-	-	טכניון*

\* כאמור, הטכניון הצטרף לפרויקט ב-86'.

הלוח שלעיל מצלם את מצב הביקוש למחלקות השונות, והוא יציב למדי על פני השנים.

11.10 המגמות שסומנו בעדיפות ראשונה, מעבר למוסדות, במספרים מוחלטים, לפי

שנים

(ניתן היה לסמן מגמה ביותר ממוסד אחד).

87	86	85	84	83	82*	81*	שנה מגמה
8	13	19	17	14	8	2	מחקרית
61	68	96	90	54	55	72	חינוכית
358	439	468	389	346	300	309	קלינית
74	63	68	72	48	66	55	קלינית של הילד
38	35	26	19	28	30	26	תעסוקתית/ ייעוצית
10	11	12	11	12	12	6	שיקומית
27	13	14	24	15	10	5	חברתית
-	11	1	3	3	-	3	קוגניטיבית
10	8	12	10	19	4	-	פסיכוביולוגית
7	1	4	4	2	-	-	התפתחותית**
-	-	-	1	1	-	2	למידה
40	25	26	18	20	33	17	חברתית - תעשייתית

\*ב-82 - 81 לא היו מגמות במוסד באוניברסיטה העברית ולכן לא צוינו. כמו כן,

לא התקיימו לימודי מוסד באוניברסיטת בן-גוריון.

\*\*המגמה לא הוצגה ב- '81, '82.

בלוח 11.10 משתקפת מידת הפופולריות של המגמות השונות. התמונה מעבר לשנים יציבה להפליא, כשהביקוש למגמות הטיפוליות (קלינית + קלינית של הילד) מאפיל כמעט על כל היתר.



## מקורות

- Angoff, W.H. (1982). Uses of difficulty and discrimination indices for detecting item bias. In R.A. Berk (Ed.), Handbook of methods for detecting item bias. Baltimore: John Hopkins University Press.
- Angoff, W.H., & Ford, S.F. (1973). Item race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 10, 95-105.
- Camilli, G. (1979). A critique of the chi-square method for assessing item bias. Unpublished paper, Laboratory of Educational Research, University of Colorado.
- Cardall, C., & Coffman, W.E. (1984). A method for comparing the performance of different groups on the items in a test (ETS Research bulletin 64-61). Princeton, NJ: ETS.
- Cleary, T.A., & Hilton, T.L. (1968). An investigation of item bias. Educational and Psychological Measurement, 28, 61-75.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.
- Donlon, T.F. (Ed.), (1984). The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests. New York: College Entrance Examination Board.
- Jensen, A.R. (1980). Bias in mental testing. New York: The Free Press.
- Plake, B.S., & Hoover, H.D. (1977). An analytic method of identifying biased test items. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Scheuneman, J. (1979). A new method for assessing bias in test items. Journal of Educational Measurement, 16, 143-152.

