

התאמת מבחני המשכל וודקוק-ג'ונסון III לעברית ולערבית בישראל

דוח פיתוח המבחנים ומחקר החלוץ



המינהל הפדגוגי
שירות פסיכולוגי
ייעוצי (שפ"י)



תהליך התרגום וההתאמה של סוללת מבחני וודקוק-ג'ונסון III לעברית ולערבית בישראל הוא מיזם של אגף פסיכולוגיה חינוכית במשרד החינוך. המיזם התבצע ונוהל על ידי ראמ"ה – הרשות הארצית למדידה והערכה בחינוך. הוא יצא לפועל בזכות שיתוף פעולה פורה בין ראמ"ה ובין גופים רבים שליוו את המחקר וביצעו שלבים שונים במהלכו: צוות הפיתוח במנהלת מדידה והערכה במטח (המרכז לטכנולוגיה חינוכית) היה אחראי לביצוע השלב של הכנת סוללת המבחנים (תרגום והתאמה של המבחנים השונים לעברית ולערבית), ביצוע מחקר חלוץ וגיבוש הנוסחים הסופיים של המבחנים. שלבים אלה מתוארים בפירוט בדוח זה. חברת "טלדור" הייתה אחראית לאיסוף הנתונים בשלב מחקר הנורמות. מאל"ו (המרכז הארצי למדידה והערכה) היה אחראי לביצוע השלב של חישוב הציונים ופיתוח הנורמות על בסיס הנתונים שנאספו במחקר הנורמות¹. תודתנו העמוקה נתונה לכל השותפים הרבים שלקחו חלק בפרויקט משמעותי זה. לרשימת השותפים המלאה והמפורטת ראו נספח 4.

¹ לקבלת דוח המתאר את תהליך פיתוח הנורמות, אפשר לפנות לראמ"ה: [/https://rama.edu.gov.il](https://rama.edu.gov.il)

תוכן עניינים

1. מבוא.....	5
א. הרקע לפרויקט ועקרונות היסוד לתהליך התרגום וההתאמה.....	5
ב. מטרת כלי ההערכה.....	7
ג. הבסיס התאורטי של המבחן.....	7
2. מבנה המבחן.....	12
3. הליכי התרגום וההתאמה של מבחן וודקוק-ג'ונסון.....	18
א. מידע כללי.....	18
ב. מיפוי ראשוני וקביעה של היקף הפיתוח הנדרש במבחנים.....	19
ג. הליך תרגום והתאמת המבחנים.....	26
ג.1. שלבי הפרויקט.....	26
ג.2. סוגיות מרכזיות בתהליך התרגום וההתאמה.....	30
ד. מחקרי חלוץ מצומצמים ומחקר חלוץ.....	45
ד.1. מחקרי החלוץ המצומצמים.....	45
ד.2. מחקר חלוץ (פיילוט).....	50
ד.3. גיבוש הנוסחים הסופיים.....	60
ד.4. היקף השינויים ואפיונם במבחנים השונים.....	67
4. נספחים.....	68
נספח 1 – שמות המבחנים בגרסאות בישראל (בעברית ובערבית).....	70
נספח 2 – המבחנים שלא הותאמו לגרסאות בישראל.....	72
נספח 3 – עיבוד נתוני תת-מבחן 'ד1'.....	73
נספח 4 – קרדיטים.....	83
5. בבליוגרפיה.....	87

רשימת טבלאות

- טבלה 1: מדדי המודל CPM, היכולות הרחבות והצרות במבחן וודקוק-ג'ונסון III והמבחנים המודדים אותם בגרסה הישראלית.....13
- טבלה 2: תיאור הליכי ההתאמה במבחנים שהותאמו לשפות העברית והערבית22
- טבלה 3: התפלגות הנבחנים במדגם של מחקר החלוץ לפי משתני הרקע- דוברי עברית.....54
- טבלה 4: התפלגות הנבחנים במדגם של מחקר החלוץ לפי משתני הרקע- דוברי ערבית.....54
- טבלה 5: דוגמה – ריכוז מידע לקראת דיון בנושא גיבוש הנוסח הסופי של תת-מבחן 1א בעברית.64
- טבלה 6: היקפי הפיתוח של הפריטים בכל שלבי המחקר67

רשימת תרשימים

- תרשים 1: השלבים העיקריים בהליך התרגום וההתאמה של המבחנים.....27
- תרשים 2: הליך התרגום וההתאמה של מבחן בוודקוק-ג'ונסון למבחן בישראל28
- תרשים 3: הליך ההרכבה (הפיתוח) של מבחן בוודקוק-ג'ונסון בישראל29

רשימת תיבות

- תיבה 1: היכולות הרחבות הנמדדות במבחן וודקוק-ג'ונסון III11
- תיבה 2: השיקולים להכללתם או לאי-הכללתם של מבחנים בגרסאות בישראל12
- תיבה 3 : מאימוץ להרכבה – התרגום וההתאמה על פני רצף.....20
- תיבה 4: התאמה תרבותית למגזר החרדי.....38
- תיבה 5: התאמה תרבותית למגזר הבדואי.....39
- תיבה 6: ניתוח DIF.....52

1. מבוא

א. הרקע לפרויקט ועקרונות היסוד לתהליך התרגום וההתאמה

מזה כשני עשורים ויותר הורגש הצורך בכלים למדידת אינטליגנציה (משכל) וכישורי חשיבה של תלמידים במערכת החינוך בישראל, הן בתחנות השירות הפסיכולוגי החינוכי, הן בקרב פסיכולוגים חינוכיים. הצורך הורגש בייחוד בקרב דוברי ערבית, שכן אין מבחן אינטליגנציה עדכני לאוכלוסייה זאת. בשנת 2011 רכש משרד החינוך את הזכויות לתרגום, להתאמה ולשימוש במבחן וודקוק-ג'ונסון III (Woodcock Johnson III). הרשות הארצית למדידה ולהערכה (להלן ראמ"ה), המובילה והמנחה את מערכת החינוך בתחום המדידה וההערכה, נתבקשה על ידי השירות הפסיכולוגי הייעוצי במשרד החינוך (להלן שפ"י) לתרגם את המבחן לשתי השפות הרשמיות של מערכת החינוך בישראל – עברית וערבית, ולהתאימו לתרבות בישראל. הפרויקט וההערכות אל הפרויקט בוצעו ונוהלו על ידי ראמ"ה, והוא יצא לפועל באמצעות המרכז של ראמ"ה לפיתוח, תרגום והתאמה של מבחנים וכלי מדידה בתחום החינוך. הזכיינית במרכז הייתה חברת מטח. עם תחילת הפרויקט הוקמה ועדת היגוי שמונתה על ידי ראמ"ה. הוועדה ליוותה את הפרויקט מבחינה מתודולוגית וכיוונה את הפיתוח. בוועדה היו אנשי ראמ"ה, אנשי שפ"י ממשרד החינוך ומומחים מתחומי הפסיכומטריקה והלינגוויסטיקה בארץ. המומחים היו מאוניברסיטאות, מהמרכז הארצי לבחינות ולהערכה, ממאגר היועצים של ראמ"ה ומשפ"י.

כדי לבצע את הפרויקט הוקם במטח צוות פיתוח אינטרדיסציפלינרי בשתי השפות (עברית וערבית), שהיו בו פסיכולוגים קליניים וחינוכיים, פסיכומטריקאים ומומחים לפסיכולינגוויסטיקה. הליכי התרגום, ההתאמה ופיתוח הגרסאות נעשו לפי הסטנדרטים המקובלים בתחום הפסיכומטריקה. מנהל-על, ששימש גם ראש צוות הפיתוח בעברית, הוביל את הצוות. דוח זה מתאר את תהליך התרגום וההתאמה של מבחן וודקוק-ג'ונסון לעברית ולערבית בישראל, מן השלב של למידת המבחנים והבנת התכונות והמבנים התיאורטיים הנמדדים, עד השלב של פיתוח המבחנים ששימשו במחקר הנורמות. תהליך זה התחיל בשנת 2013, ונמשך עד שנת 2017. דוח נפרד מתאר את תהליך פיתוח הנורמות למבחנים.

תרגום המבחנים והתאמתם

Hambleton (2005) כתב: "אדפטציה של מבחן פסיכולוגי מתרבות ומשפה כלשהן לתרבות ולשפה אחרת היא תהליך מורכב ומאתגר. בניגוד לתרגום פשוט, התאמה היא תהליך רחב, שתחילתו בבירור השאלה אם העברת המבחן לתרבות אחרת תשמר את המבנה התאורטי הנמדד. בהמשך התהליך מתקבלת החלטה בנוגע לסוגי ההתאמות התרבותיות והשפתיות הנדרשות, ובהתאם לכך נבנה המבחן בשפה החדשה ('שפת המטרה'). בסוף התהליך נדרשת בחינה מעמיקה של התוצר אם הוא אכן אקוויולנטי לכלי המקורי, גם במבנה התאורטי הנמדד וגם באופן המדידה."

תהליך התרגום וההתאמה של מבחנים וכלי מדידה פסיכולוגיים וחינוכיים נעשה מתוך זהירות והקפדה רבה. יש דוגמות רבות של תהליכי התאמה ותרגום שלא הצליחו, שכן התוצאה הייתה כלים שתקפותם הייתה מוטלת בספק (Hambleton & Lee, 2013). Hambleton and Lee סברו שיש לכך כמה

סיבות, וביניהן אי-הבנה וחוסר מודעות למורכבות התהליך הנדרש, נקיטת אסטרטגיה של תרגום פשוט וישיר מתוך תפיסה שגויה שלפיה תרגום טוב לבדו יישמר את תקפות הכלי ככלי מדידה של תכונות פסיכולוגיות. כמו כן קוצר זמן, מחסור בתקציב וידע מקצועי חסר בנוגע לצעדים הנדרשים כדי להוציא לפועל הליכי תרגום והתאמה ראויים.

ועדת המבחנים הבינלאומית (International Test Commission) פרסמה קווים מנחים לתרגום ולהתאמה של מבחנים (2005, 2010). קווים מנחים אלה מסייעים להבנות תהליך מקצועי להפקת מבחנים שיש בהם מאפיינים פסיכומטריים דומים למבחנים שבשפת המקור (Hambleton, 2005). התהליך עוסק בהתאמה של מבחנים מילוליים, של מבחנים שאינם מילוליים (שבהם יש להתאים גם את הגרפיקה, את התמונות ואת הציורים), של ההנחיות, של אופן מספור הפריטים ולמעשה של כל דבר שעלול ליצור תרבותיות (על הטיה תרבותית, ראו בסעיף ג.2.יא). במשך השנים פורסמו מאמרים שעסקו בקווים המנחים האלה

(למשל, Gudmunsson, 2009; Hambleton & Patsula, 1998, 1999; Hambleton, 2005). תהליך הפיתוח, שכלל תרגום והתאמה של המבחן וודקו-ג'ונסון III לשפות המטרה (עברית וערבית), היה מעמיק ומורכב והתבצע לפי השלבים האלה:

1. בחינת המבנה התאורטי והמשמעות שלו בארצות הברית ובישראל (בתרבות העברית ובתרבות הערבית).
2. בדיקת הביטוי האופרציונלי של המבחן בתרבויות השונות ובשפות השונות (במהלך התרגום וההתאמה ולאחר התרגום וההתאמה).
3. פיתוח מחדש של מבחנים ופריטים (אם אי אפשר לתרגם אותם או שאין טעם לתרגם), בהתחשב בהיבטים התרבותיים, המגדריים והפסיכומטריים הרלוונטיים.
4. בדיקת הפריטים המותאמים והחדשים במדגם קטן של נבחנים כדי לזהות בעיות במבחן בשפות המטרה (להלן "מחקרי חלוץ מצומצמים"), ותיקון או שינוי פריטים לפי המסקנות ממחקרים אלה.
5. ביצוע מחקר חלוץ נרחב (להלן "מחקר חלוץ, פיילוט") ועיבוד נתונים מקיף.
6. הסקת מסקנות מהתוצאות של מחקר החלוץ וגיבוש סופי של המבחנים לפי הנתונים האמפיריים שנאספו במחקר זה.
7. ביצוע מחקר נורמות וקביעת ציוני הנורמות במבחן (ראו תיאור בדוח נפרד).
8. כתיבת חומרים גלויים, כגון המדריך לבוחן ודוח הפיתוח (דוח זה).

ב. מטרת כלי ההערכה

מבחן אינטליגנציה (משכל) הוא כלי מדידה לכשרים קוגניטיביים ("רמת משכל"), שמטרתו לסייע לפסיכולוג החינוכי לעמוד על החוזקות הקוגניטיביות של הנבחן ועל הקשיים שלו. בעזרת מבחן משכל הפסיכולוג אומד את הדרכים שבהם הנבחן מתמודד עם משימות קוגניטיביות שונות. הפסיכולוג החינוכי יכול להסתייע בתמונת התפקוד המתקבלת ממבחן המשכל לשם בניית תוכנית התערבות שתעזור לנבחן לחזק את תפקודו בתחומים שבהם הוא מתקשה, וכמו כן למצות את יכולתו בתחומים שבהם הוא חזק. מבחן משכל הוא אחד הכלים שלפיהם קובעים לקות למידה. נוסף על כך מבחן משכל יכול לשמש לאיתור אוכלוסיות קצה שונות (פיגור שכלי, מחוננות), כך הוא אחד הכלים שבהם הפסיכולוג החינוכי מסתייע בבואו לקבל החלטות בנוגע לשיבוץ ילד כלשהו במסגרת חינוכית שתיתן מענה לצרכיו.

מטרת פרויקט התרגום וההתאמה של כלי ההערכה למדידת משכל לעברית ולערבית בישראל הייתה להעמיד ברשות הפסיכולוגים החינוכיים במערכת החינוך בישראל כלי תקף ועדכני למדידת כשרים קוגניטיביים, אשר ישמש אותם בעבודתם השוטפת.

ג. הבסיס התאורטי של המבחן

מבחן וודקוק-ג'ונסון I (הגרסה הראשונה) (Woodcock-Johnson Psychoeducational Battery) פורסם בשנת 1977 באנגלית על ידי הפסיכולוג ריצ'רד וודקוק ושותפתו מרי ג'ונסון (Richard Woodcock & Mary E. Bonner Johnson). המבחן יצא לאור בהוצאת Resources Corporations, מינסוטה, ארצות הברית. בשנים 1989, 2001, ו-2014 זכה המבחן לגרסאות מעודכנות. המבחן שתורגם והותאם לישראל בפרויקט זה הוא גרסת וודקוק-ג'ונסון III (הגרסה השלישית). הפיתוח שלה החל בשנת 1997 והיא יצאה לאור בארצות הברית בשנת 2001 בהוצאת Riverside Publishing (Woodcock, R. W., McGrew, K. S., & Mather, N., 2001) Company. לפיתוח גרסה זו חבר ריצ'רד וודקוק לשותפים חדשים: פרדריק שרנק, קווין מק'גרו וננסי מת'ר (Fredrick A. Shrank, Kevin S. McGrew, Nancy Mather).

מבחן וודקוק-ג'ונסון הוא מקבץ (סוללה) של מבחנים הנחלק לשני חלקים: חלק אחד בוחן כשרים קוגניטיביים וחלק אחר בודק הישגים לימודיים (קריאה, כתיבה ומתמטיקה). חלקים במבחן מתאימים להעברה החל מגיל שנתיים, ולמבחן נורמות מגיל שנתיים עד גיל 90. כמענה לצורך העיקרי של הפסיכולוגים העובדים בשירותים הפסיכולוגיים החינוכיים, הוחלט שבפרויקט זה יתבצעו התרגום וההתאמה לישראל רק של **החלק הקוגניטיבי** ורק של **קבוצות הגיל 3.5-18** (גיל התלמידים במערכת החינוך). לפיכך דוח הפיתוח הזה עוסק אך רק בחלק הקוגניטיבי של מבחן וודקוק-ג'ונסון III.

התאוריה CHC שייכת לגישה הפסיכומטרית למשכל. התאוריה רותמת שיטות סטטיסטיות ובעיקר ניתוחי גורמים כדי לפתח ולגבש מודלים פסיכולוגיים וכדי להבין את המבנה הפסיכולוגי (construct)

שאותו מעוניינים למדוד. שני מקורות מחקר עיקריים, הנוגעים למבנה של היכולות הקוגניטיביות האנושיות, עומדים בבסיס תאוריה זו.

מקור המחקר הראשון כולל את הממצאים מניתוחי גורמים שערכו ריימונד קטל וג'ון הורן, והוא מבוסס על התאוריה הנקראת Gf-Gc (Horn, 1985, 1988, 1991; Woodcock, 1990, 1994). האותיות Gf מייצגות יכולת קוגניטיבית פלואידית (היכולת להבין קשרים ויחסים בין מושגים), והאותיות Gc מייצגות יכולת קוגניטיבית קריסטלית (ידע מגובש – גוף ידע והרגלים המצטברים בעקבות הפעלה של היכולת הפלואידית). קטל היה אחד הראשונים שהבחין בין שתי היכולות הללו (Cattell, 1941, 1943, 1950). לאחר מכן זיהה הורן (Horn, 1965) יכולות קוגניטיביות רחבות אחרות, וביניהן Gsm – הזיכרון לטווח קצר (היכולת לקודד ולשמור מידע ולבצע מניפולציה מנטלית במידע הנמצא במודעות המיידית), Glr – אחסון ושליפה לטווח ארוך (היכולת לאחסן, לגבש ולשלוף מידע בטווח של דקות, שעות, ימים ושנים), Gs – מהירות העיבוד (היכולת לבצע משימות קוגניטיביות פשוטות החוזרות על עצמן במהירות ובשטף) – Gv-i – עיבוד חזותי-מרחבי (היכולת להשתמש בדימויים חזותיים כדי לפתור בעיות). כעבור שנים הוסיפו החוקרים את היכולת Ga – עיבוד שמיעתי (היכולת לזהות ולעבד מידע שמיעתי, Horn & Stankov, 1982), ובשנת 1989 נכללו שבע היכולות הרחבות האלה במבחן WJ-R – Woodcock Johnson test of Cognitive Ability Revised (Woodcock, 1990; Woodcock & Johnson, 1989a).

הורן (1988, 1989) ערך עוד מחקרים והוסיף למודל הזה את Gq – היכולת הכמותית, עומק הידע הקשור למתמטיקה והיקפו. כעבור כמה שנים הוסיף וודקוק (1998) את Grw – יכולות הקריאה והכתיבה (עומק הידע והכישורים הנוגעים לשפה כתובה והיקפם).

מקור המחקר השני כולל את הממצאים ממחקרים שביצע ג'ון קרול בשנים 1993 ו-1998. קרול (Carroll, 1993) ניתח מחדש את הנתונים שהתקבלו מ-461 מחקרים על היכולות הקוגניטיביות שבוצעו החל משנת 1925. הוא השתמש בארבעה מאגרי מידע של המהדורה הראשונה של מבחני וודקוק-ג'ונסון משנת 1977. בספרו הציג קרול את תוצאות מחקרו. לספר הייתה השפעה רבה על ההבנה של מבנה היכולות הקוגניטיביות האנושיות (McGrew, 2005). קרול המשיג את היכולות הקוגניטיביות במבנה היררכי של שלוש שכבות. הוא זיהה 69 יכולות צרות שאותן הוא שייך לשכבה התחתונה בהיררכיה. את היכולות הצרות האלה הוא קיבץ ליכולות קוגניטיביות רחבות יותר, שהיו בשכבה האמצעית בהיררכיה. בשכבה זו כלל קרול יכולת פלואידית, ידע מגובש, זיכרון ולמידה, עיבוד חזותי, עיבוד שמיעתי, יכולת שליפה, מהירות קוגניטיבית ומהירות עיבוד. בשכבה העליונה, בראש ההיררכיה, זיהה קרול גורם כללי, שאותו כינה **רמת המשכל הכללית – g**.

הדמיון בין שני מקורות המידע העצמאיים הללו, של קטל והורן ועמיתיהם ושל קרול, נתן תמיכה לתאוריה CHC המשולבת. קרול עצמו תיאר את המודל CHC כמודל הזמין הטוב ביותר של מבנה המשכל האנושי, אך הוסיף הסתייגות: "מודל זה אינו מציב בראש ההיררכיה את רמת המשכל הכללית, את הגורם g".

המודל CHC פותח בשנת 2000 על ידי קווין מק'גרו (Kevin McGrew), חוקר משכל ופסיכולוג חינוכי, והוא היה הבסיס למבחני וודקוק-ג'ונסון במהדורתם הנוכחית. המודל CHC, כמו המודל של קרול, רואה במשכל מבנה היררכי בן שלוש רמות. ברמה העליונה ניצב ה-g, רמת המשכל הכללית, שאותה

מנסים לאמוד באמצעות הציון IQ. ה-g מורכב מכמה יכולות קוגניטיביות רחבות, שכל אחת מהן מורכבת מכמה יכולות קוגניטיביות צרות. המבחנים במבחני המשכל בודקים את היכולות הקוגניטיביות הצרות.

המודל CHC של מק'גרו מאפשר השתנות במשך הזמן, בהתאם למחקר, ובזה כוחו. ככל שהמחקר מתפתח הוא מצביע על מבנים חדשים ויכולות חדשות התורמות להבנת התכונה והמבנה הפסיכולוגי הנמדדים. בגרסאות הראשונות שלו היו עשר יכולות רחבות. בגרסה האחרונה, שפותחה בשנת 2012 על ידי שניידר ומק'גרו (Schneider & McGrew), כלולות 16 יכולות רחבות ו-74 יכולות צרות.

בשנות ה-90 הציע ריצ'ארד וודקוק מודל תאורטי נוסף להמשגת היכולת הקוגניטיבית והאינטליגנציה – מודל הביצוע הקוגניטיבי (CPM, Cognitive Performance Model). המושג "ביצוע קוגניטיבי" מחליף במודל התאורטי הזה את המושגים "אינטליגנציה" ו"יכולת קוגניטיבית". לפי מודל זה אפשר להסיק מסקנות בנוגע למשכל של אדם רק מתוך תצפיות בביצוע שלו. הסיבה לכך היא שהביצוע הנצפה מושפע לא רק מרמתו הקוגניטיבית של הנבחן, אלא גם ממצבו הרגשי, מהסביבה שבה הוא גדל, מהמצב האבחוני שלו ומהאינטראקציה שלו עם הבוחן. הפסיכולוג המאבחן נדרש לבחינה אינטגרטיבית של כל ההיבטים הללו בעת פירוש תוצאות המבחן (Mother & Woodcock, 2001). וודקוק טען שלוש טענות מרכזיות שבעזרתן אפשר להבין את המושג "ביצוע קוגניטיבי" ואת המבחן המודד אותו, הן מבחינת תכניו, הן מבחינת אופני ההיבחנות המגולמים בו. שלוש הטענות הן אלה:

א. למידה חדשה נשענת על יכולת החשיבה של האדם. במושג "יכולת חשיבה" אפשר לכלול שלוש יכולות קוגניטיביות: (1) היכולת הפלואידית (היכולת של האדם לחשוב בצורה מופשטת, להסיק מסקנות ולהכליל הכללות), ואיכות התפקודים הניהוליים של האדם (יכולתו לבצע משימות מורכבות בדרך מתוכננת ומבוקרת, בגמישות מחשבתית ובחתיירה למטרה); (2) היכולת של האדם לאחסן מידע ולשלוף אותו מהזיכרון לטווח ארוך; (3) היכולת של האדם לעבד מידע חזותי ושמיעתי.

ב. כל ביצוע, הן במצבים שבהם אדם נחשף לתהליכים של למידה חדשה, הן במצבים שבהם הוא מבצע פעולות אוטומטיות (כמו קריאת מילים וכתיבתן או ניווט ליעד מוכר), נשען על "מאגרי הידע" הרלוונטיים העומדים לרשותו. המושג "מאגרי ידע" עוסק בשלוש יכולות: (1) הידע המגובש (ההיכרות של האדם עם התרבות שבה הוא חי, היכולת הלשונית שלו – הרחב והעומק של הלקסיקון שלו, השליטה שלו בדקדוק ובתחביר והיכולות שלו להביע רעיונות ולהבין רעיונות המובעים באמצעות שפה); (2) הידע המתמטי; (3) יכולות הקריאה והכתיבה. מאגרי הידע הנרכש ניתנים לשינוי: אסטרטגיות הוראה והזדמנויות להעשרה יכולות להשפיע על רמות הביצוע של אדם.

ג. הביצוע האוטומטי מוגבל על ידי הזיכרון לטווח קצר וגם על ידי מהירות העיבוד של האדם. הזיכרון לטווח קצר ומהירות העיבוד קובעים יחד את יעילות הביצוע הקוגניטיבי של האדם. יכולות אלה הן מעין "צוואר בקבוק" ויש להן תפקיד מרכזי הן בקליטת מידע וגירויים מהסביבה, הן בעיבוד שלהם, הן בקישור שלהם לסביבה.

את הביצועים של הנבחן במבחן וודקוק-ג'ונסון III אפשר לפרש גם לפי המודל CPM של וודקוק. כך ציוני הנבחן במבחנים מתקבצים לשלושה "מדדי-על", שכל אחד מהם מותאם לאחת משלוש היחידות התפקודיות האלה: **(א)** המדד ליכולת החשיבה; **(ב)** המדד ליכולת הוורבלית²; **(ג)** המדד ליעילות הקוגניטיבית.

התקדמות המחקר השפיעה על מספר היכולות המרכיבות את המשכל והמגדירות אותו מחדש. גיליון והמשגתן של יכולות חדשות הרחיבו את מספר היכולות הכלולות בתאוריה CHC. לפי התאוריה כל יכולת קוגניטיבית רחבה, למשל "עיבוד חזותי" ו"ידע מגובש", מורכבת מיכולות קוגניטיביות צרות יותר (נמצאות במיקום היררכי נמוך יותר). המבחן וודקוק-ג'ונסון בנוי ממבחנים רבים שכל אחד מהם בודק יכולת צרה אחת, וביחד הם יוצרים תמונה כללית של משכל האדם. הדרך הטובה ביותר למדידת יכולת רחבה היא למפות אותה ולפרק אותה לשתיים-שלוש יכולות צרות יותר, ולבנות מבחנים כך שכל אחד מהם יתמקד בבדיקת יכולת צרה אחרת. במבחן וודקוק-ג'ונסון III נמדדות שבע יכולות קוגניטיביות המוגדרות כיכולות רחבות: **ידע מגובש, אחסון לטווח ארוך ושליפה, עיבוד חזותי-מרחבי, עיבוד שמיעתי, חשיבה פלואידית, מהירות עיבוד וזיכרון לטווח קצר.**

² החלק הקוגניטיבי של המבחן אינו מכסה את כל ההיבטים של מאגרי הידע שכן הקריאה, הכתיבה והמתמטיקה נבדקים באמצעות מבחני ההישגים של מבחן וודקוק-ג'ונסון III, והם כאמור לא תורגמו ולא הותאמו בפרויקט הזה.

תיבה 1: יכולות CHC רחבות הנמדדות במבחן וודקוק-ג'ונסון III

1. ידע מגובש – ידע עולם הזמין לאדם. הרוחב והעומק של הידע והכישורים המוערכים על ידי התרבות שהאדם חי בה.
2. אחסון ושליפה לטווח ארוך – היכולת לאחסן ולגבש מידע חדש בזיכרון לטווח ארוך ולשלוף ממנו את המידע.
3. עיבוד חזותי-מרחבי – תפיסה מרחבית, היכולת לתפוס גירויים חזותיים, להחזיק אותם בראש ולבצע בהם מניפולציות.
4. עיבוד שמיעתי – היכולת להבחין בין גירויים שמיעתיים, לערוך אנליזה וסינתזה לגירוי שמיעתי.
5. היכולת הפלואידית – היכולת לשלוט באופן מכוון וגמיש בקשב כדי לפתור בעיות חדשות שאי אפשר לבצען באופן אוטומטי: לחשוב בצורה מופשטת, להסיק מסקנות, להעלות השערות, להבין השלכות ולהכליל כללים.
6. מהירות העיבוד – היכולת לבצע באופן אוטומטי ושוטף, אחרי תרגול רב, משימות קוגניטיביות קלות יחסית, החוזרות על עצמן, שנדרש בהן מיקוד הקשב.
7. הזיכרון לטווח קצר/זיכרון עבודה – היכולת לקודד מידע, לשמור מידע ולבצע מניפולציה במידע הנמצא במודעות המיידית.

2. מבנה המבחן

במבחן המקורי באנגלית, בחלק המודד כשרים קוגניטיביים, יש 31 מבחנים הנחלקים לשלוש סוללות: 10 מבחנים בסוללה "הבסיסית", 10 מבחנים בסוללה "המורחבת" ו-11 מבחנים בסוללה "המשלימה". בגרסאות של המבחן בישראל תורגמו והותאמו 23 מבחנים הנחלקים לשתי סוללות – **סוללה בסיסית** שבה המבחנים 1-14 (המספרים בגרסאות בישראל), ו**סוללה משלימה** שבה המבחנים 15-23 (המספרים בגרסאות בישראל). בנספח 1 אפשר לראות את המספרים ואת שמות המבחנים, בגרסה בישראל ובגרסת המקור.

בתיבה 2 שלהלן כתובים השיקולים להכללתם או לאי-הכללתם של מבחנים מגרסת המקור בגרסאות בישראל.

תיבה 2: השיקולים להכללתם או לאי-הכללתם של מבחנים בגרסאות בישראל

1. נבחרו שניים – שלושה מבחנים לכל יכולת רחבה.
2. נבחרו מבחנים המייצגים יכולות צרות שונות של אותה יכולת רחבה.
3. נבחרו (במידת האפשר) מבחנים הבודקים אותה יכולת והמייצגים סוגי גירויים שונים (מילולי/חזותי/מספרי).
4. לא נבחרו מבחנים שנמצאו לקויים מבחינה פסיכומטרית או שלא היה ברור לצוות המומחים אם הם מודדים את מה שהם צריכים למדוד (דוגמה למבחן כזה הוא מבחן 23 בגרסת המקור (sound pattern-voice) – מבחן שבו הנבחן שומע זוגות של דפוסי קול ונדרש להגיד אם הדפוסים זהים או שאינם זהים).

בסוללה הבסיסית (מבחנים 1-14) יש שני מבחנים למדידת כל אחת מהיכולות הרחבות. בסוללה המשלימה יש מבחנים נוספים שאפשר להיעזר בהם לפי הצורך כדי לבחון היבטים נוספים של ביצועי הנבחן. הבוחן לא מעביר בדרך כלל את כל 23 המבחנים שבגרסה בישראל לכל נבחן. המבחן בנוי באופן מודולרי ולבוחן יש גמישות בהעברת המבחנים. לאחר שהבוחן מנסח את שאלת האבחון, הוא בוחר את המבחנים שסייעו לו להשיב עליה, ובאמצעותם יבדוק את הנבחן.

זמן ההעברה של כל מבחן בסוללה הוא 2 עד 10 דקות. זמן זה משתנה בין מבחן למבחן ועלול להיות ארוך יותר או קצר יותר לפי מאפייני הנבחנים. זמן העברה הכולל של המבחנים בסוללה הבסיסית הוא כשעה וחצי עד שעתיים.

בטבלה 1 שלהלן מוצגות היכולות הנמדדות במודל CPM, וכמו כן היכולות הרחבות והצורות הנמדדות במבחן וודקוק-ג'ונסון III – לפי התאוריה CHC. לצדן כתובים שמות המבחנים בגרסה הישראלית המודדים כל אחת מהיכולות.

טבלה 1: מדדי המודל CPM, היכולות הרחבות והצרות במבחן וודקוק-ג'ונסון ווו

והמבחנים המודדים אותם בגרסה הישראלית

מדדי CPM	יכולות CHC רחבות והגדרתן	יכולות CHC צרות והגדרתן	הסוללה הבסיסית		הסוללה המשלימה	
			מספר המבחן	שם המבחן	מספר המבחן	שם המבחן
ידע נרחב	ידע מגובש - הרוחב והעומק של הידע והכישורים המוערכים על ידי התרבות שהאדם חי בה.	ידע לקסיקלי - הידע השייך לאוצר המילים בשפה.	א1 ב1 ג1 ד1	הבנה מילולית		
			א8 ב8	ידע כללי		
יכולת חשיבה	אחסון ושליפה לטווח ארוך - היכולת לאחסן, לגבש ולשלוף מידע לטווח זמן ארוך של דקות, שעות, ימים ושנים. מבחינים בשתי יכולות ביניים: יעילות בלמידה (שבה כלולה בין היתר למידה אסוציאטיבית), ושטף בשליפה (שבה כלולים בין היתר שטף רעיוני וקלות שיום). שטף בשליפה מוגדר כקצב שבו אנשים יכולים לגשת למידע השמור בזיכרון לטווח ארוך.	למידה אסוציאטיבית - היכולת לזכור צמדים של פרטי מידע שאינם קשורים זה לזה בדרך כלל.	2	למידה חזותית-שמיעתית	15 23	זכירת שמות זכירת שמות מושהית
			9	שטף סמנטי		
		שטף רעיוני - היכולת להפיק במהירות סדרות של רעיונות, מילים או ביטויים הקשורים לנושא כלשהו או לקטגוריה כלשהי. הדגש הוא על מספר הרעיונות ולא על איכותם.			21	שיום תמונות מהיר
		קלות שיום - היכולת לכנות במהירות אובייקטים בשמותיהם; יכולת זו מכונה גם "שיום אוטומטי מהיר" או "מהירות הגישה ללקסיקון".				
	עיבוד חזותי-מרחבי - תפיסה מרחבית, היכולת	עיבוד מרחבי / מהירות רוטציה - היכולת לפתור	3	עיבוד מרחבי		

הסוללה המשלימה		הסוללה הבסיסית		יכולות CHC צרות והגדרתן	יכולות CHC רחבות והגדרתן	מדדי CPM
שם המבחן	מספר המבחן	שם המבחן	מספר המבחן			
				במהירות בעיות תוך כדי שימוש ברוטציה מנטלית של דימויים חזותיים פשוטים.	לעשות אנליזה וסינתזה לגירוי חזותי והיכולת להחזיק בראש	
		זיהוי ציורים	10	זיכרון חזותי - היכולת לזכור דימויים חזותיים מורכבים במשך פרק זמן הקצר משלושים שניות.	דימויים ולבצע בהם מניפולציה. היכולת להשתמש בדימוי חזותי (לעיתים קרובות בד בבד עם גירויים חזותיים	
סיבוב עצמים	22			ויזואליזציה - היכולת לתפוס דפוסים חזותיים מורכבים, לערוך סימולציה מנטלית ולדמות איך הם ייראו אם ישנו אותם (למשל יסובבו אותם או יסתירו אותם חלקית).	אחרים) כדי לפתור בעיות.	
תכנון	16			סריקה חזותית - היכולת לבצע במהירות סריקה חזותית של דפוס או של שדה מרחבי מורכב.		
		מיזוג צלילים	4	סינתזה (מיזוג/חיבור) פונטית - היכולת לעבד צלילי דיבור: לזהות צלילים, ולמזג צלילים או לערוך סינתזה שלהם.	עיבוד שמיעתי - היכולת להבחין בין גירויים שמיעתיים, לערוך אנליזה וסינתזה לגירוי שמיעתי. בעיבוד השמיעתי כלולות יכולות הקשורות	

מדי CPM	יכולות CHC רחבות והגדרתן	יכולות CHC צרות והגדרתן	הסוללה הבסיסית		הסוללה המשלימה	
			מספר המבחן	שם המבחן	מספר המבחן	שם המבחן
	למודעות הפונולוגית.	אנליזה (פירוק/הפרדה) פונטית – היכולת לעבד צילי דיבור: לזהות צלילים, לבדד צלילים או לערוך אנליזה שלהם.	11	מילים לא שלמות		
		הבחנה בין צלילים – היכולת למצוא הבדלים בין צילי דיבור בתנאים של הסחה או עיוות.			17	קשב שמיעתי
	יכולת פלואידית – שליטה מכוונת וגמישה בקשב כדי לפתור בעיות חדשות, שאי אפשר לפתור אותן על סמך הרגלי למידה, על סמך סכמות או על סמך תסריטים קודמים.	אינדוקציה – היכולת לצפות בתופעה ולגלות את העקרונות או את הכללים העומדים בבסיסה והקובעים את התנהגותה.	5	יצירת כללים		
		דדוקציה / חשיבה סדרתית / יישום כללים – היכולת לחשוב באופן לוגי תוך כדי שימוש בהנחות ובעקרונות ידועים.	12	אנליזה-סינתזה	16	תכנון
		חשיבה כמותית – היכולת לחשוב באמצעות אינדוקציה או דדוקציה על מספרים, על יחסים מתמטיים ועל פעולות מתמטיות.			18	סדרות מספרים

מדדי CPM	יכולות CHC רחבות והגדרתן	יכולות CHC צרות והגדרתן	הסוללה הבסיסית		הסוללה המשלימה	
			מספר המבחן	שם המבחן	מספר המבחן	שם המבחן
יעילות קוגניטיבית	מהירות עיבוד – היכולת לבצע משימות קוגניטיביות פשוטות החוזרות על עצמן, במהירות ובשטף.	מהירות תפיסה – היכולת להשוות במהירות בין גירויים חזותיים כדי למצוא דמיון או שוני ביניהם.	6	התאמה חזותית	19	איתור ציורים מהיר
		מהירות חשיבה – היכולת לבצע משימות בקצב או בשטף בזמן מוגבל (למשל מהירות ההפקה של כמה שיותר כללים אפשריים או של כמה שיותר פתרונות אפשריים לבעיה).	13	מהירות החלטה		
		קלות שיום – היכולת לכנות במהירות אובייקטים בשמותיהם; יכולת זו מכונה גם "שיום אוטומטי מהיר" או "מהירות הגישה ללקסיקון".			21	שיום תמונות מהיר
	זיכרון לטווח קצר / זיכרון עבודה – היכולת לקודד מידע, לשמור מידע ולבצע מניפולציה במידע הנמצא במודעות המיידית; יכולת זו בודקת את קיבולת הזיכרון	טווח זיכרון – היכולת לקודד מידע, לשמור אותו בזיכרון הראשוני ולהפיק אותו מחדש באופן מיידית, לפי סדר הצגתו.	7	זכירת מילים		
		זיכרון עבודה/שליטה בקשב – היכולת	14	זכירת ספרות לאחר		

הסוללה המשלימה		הסוללה הבסיסית		יכולות CHC צרות והגדרתן	יכולות CHC רחבות והגדרתן	מדדי CPM
שם המבחן	מספר המבחן	שם המבחן	מספר המבחן			
זיכרון עבודה שמיעתי	20			לכוון את מוקד הקשב לביצוע מניפולציות – שילובים ושינויים פשוטים יחסית במידע הנמצא בזיכרון הראשוני (בזמן חסימה של גירויים מסיחים).	הראשוני ואת יעילות השליטה בקשב.	

3. הליכי התרגום וההתאמה של מבחן וודקוק-ג'ונסון

א. מידע כללי

התרגום וההתאמה של המבחנים השונים נעשו על בסיס של רקע תאורטי נרחב ועדכני ועל בסיס סטנדרטים פסיכומטריים גבוהים המקובלים בתחום של פיתוח מבחנים בארץ ובעולם. מפתחי המבחנים נעזרו בספרות המתעדת את פיתוח מבחן וודקוק-ג'ונסון המקורי. במקרים בודדים הם התייעצו עם אחדים מבין אלה שפיתחו את המבחנים בארצות הברית ועם מומחים בעולם להתאמה ולפיתוח מבחנים. במקרים שלא היה אפשר למצוא בספרות את הרציונל התאורטי והמעשי של פיתוח המבחנים בשפת המקור, ולא היה אפשר לקבל הנחיות ברורות מאלה שפיתחו אותם, נעשה ניתוח התוכן והמבנה של המבחן על ידי המומחים וצוותי הפיתוח בישראל. הם גם ניסחו את העקרונות הנדרשים לתרגום ולהתאמה של הפריטים בכל אחד מהמבחנים שנותחו. כמו כן בהליך התאמת המבחנים הובא בחשבון ההקשר החינוכי – בית ספרי, למשל מינוחים המקובלים בבתי הספר או רמת הלמידה במערכת החינוך בישראל. נוסף על כך הובאה בחשבון העובדה שהשפות עברית וערבית נכתבות מימין לשמאל, ועלולות להיות לכך השלכות על עימוד רכיבים טקסטואליים וגרפיים בדפי המבחן.

המבחנים הותאמו לשני המגזרים, העברי והערבי, מתוך התחשבות בסממניו של כל מגזר, כלליו וערכיו, ועל בסיס ההיכרות של הייחוד שבכל שפה ותרבות. בכל מגזר הובאו בחשבון הגיוון מבחינה תרבותית וערכית: במגזר דוברי העברית – דתיים, חרדים, עולים חדשים מארצות מוצא שונות; במגזר דוברי הערבית – בני דתות שונות, מאזורים שונים בארץ, בדואים ודרוזים. עם זאת בגלל מגבלות של משאבים ובגלל הצורך בסולם אחיד לכל האוכלוסייה, נבנתה גרסה אחת בעברית וגרסה אחת בערבית.

ועדת ההגוי³, שהשתתפו בה מומחים מן האקדמיה, מראמ"ה, ממוסדות ארציים לפיתוח מבחנים וממשרד החינוך, ליוותה את הפיתוח. המומחים היו מתחומי הפסיכומטריקה ופיתוח מבחנים, הפסיכולוגיה, השפה והפסיכו-לינגוויסטיקה. המומחים ניתחו את מבחני המקור, הנחו וביקרו את צוות הפיתוח בשלבים השונים ודנו בדילמות עקרוניות שהתעוררו במהלך התרגום וההתאמה. כמו כן הם יעצו לראמ"ה בסוגיות שונות הנוגעות לתרגום ולהתאמה של המבחן, ובנוגע לשימוש העתידי הראוי בו ולפרויקט בכללותו. למשל הוועדה המליצה על המבחנים שיש "להרכיב מחדש"⁴ ועל המבחנים שבהם

³ באופן פרקטי ועדת ההיגוי נחלקה לשתי ועדות: ועדה מתודולוגית שדנה בנושאים כללים הנוגעים להליך התרגום-ההתאמה לישראל וסייעה בהתוויתם של דרכים אלו, וועדת היגוי שדנה במבחנים השפתיים בלבד.

⁴ הליך ההרכבה מחדש דומה מאוד להליך של פיתוח של פריטים ומבחנים באשר הם וכולל שלבי פיתוח דומים למעט שלב הגדרת המבחן (מבנה המבחן והפריטים, סוגם וכו') והמסגרת הקונצפטואלית שלו אשר קיים כבר במבחן המקורי.

נדרשת רק התאמה או שנדרש בהם רק אימוץ ישיר (תרגום פשוט). הוועדה המליצה על שיעור הפריטים העודפים שיש לפתח בכל מבחן, וכמו כן דנה בשאלות כגון אלה: מה יהיה מנעד ההבדלים בין הגרסה הערבית ובין הגרסה העברית? האם להתאים תת-גרסאות לאוכלוסיות נפרדות? האם יהיה לשתי גרסאות השפה סולם ציונים משותף? ועוד. כמו כן המליצה הוועדה על גודל המדגם במחקרי החלוץ ועל אופיו של המדגם, ויעצה בסוגיות הקשורות לעיבוד הנתונים, לבניית הנורמות ועוד.

צוות הפיתוח במטח, שעסק בתרגום ובהתאמה בפועל, הורכב מפסיכומטריקאים, מפסיכולוגים חינוכיים בכירים ומפסיכו-לינגוויסטיקאים, אשר התמחותם העיקרית היא רכישת שפה בקרב צעירים, בשפות העברית והערבית. הם נתמכו בעורכי לשון ובגרפיקאים. הצוות חולק לשני צוותי משנה, האחד של מומחים דוברי עברית והאחר של מומחים דוברי ערבית.

שני הצוותים עבדו בנפרד ובמקביל, לפי אותם עקרונות תרגום, התאמה ופיתוח, אך בממשק ובניהול משותף תוך כדי העשרה, חשיבה משותפת, השראה הדדית והתבוננות ביקורתית על תוצרי התהליך. לצוותים מונה מנהל-על אחד אשר היה אמון על ניהול הצוותים והפרייקט כולו והיה הגורם המקשר בין שני הצוותים. כדי ליצור סינכרון ודיון משותף בסוגיות פיתוח שונות, נערכו ישיבות פיתוח של צוותי שתי השפות יחד, בתחילת הפיתוח ובמהלך הפיתוח, בשלבים שהוגדרו מראש. כמו כן ראשי הצוותים בעברית ובערבית, שיכלו לראות תמונה רחבה של התאמת המבחנים והרכבתם בשתי השפות, סייעו ליצור שתי גרסאות בשתי שפות, שיש הלימה ביניהן (עוד בנושא זה, ראו בהמשך תרשימים 1, 2 ו-3). צוותי הפיתוח עבדו במשך כל הפרויקט לפי המתווה של ועדת ההיגוי והמלצותיה, בתיאום מלא עם מינהלת הפרויקט בראמ"ה, אשר הובילה, פיקחה וניהלה את הפרויקט בכללותו.

ב. מיפוי ראשוני וקביעה של היקף הפיתוח הנדרש במבחנים

תרגום והתאמה של מבחן עוסקים בהליך המתחיל בבירור השאלה אם העברת מבחן משפה לשפה ומתרבות לתרבות מצליחה לשמר את המבנה התאורטי הנמדד. לאחר מכן נשאלת השאלה בנוגע להתאמות התרבותיות והשפתיות הנדרשות. בשלב הבא מתבצעת בנייה בפועל של המבחן בשפה החדשה בהתאם להחלטות שמתקבלות. לבסוף נבדק אם הכלי שהתקבל בשפת המטרה אקוויולנטי לכלי המקורי מבחינת המבנה התאורטי הנמדד ומבחינת דרך המדידה (Hambleton, 2005). נהוג להשתמש במונח "התאמה" של מבחן במובן הרחב של המילה, לכל הליך שבו נערכים שינויים בכלי כלשהו כדי להשתמש בו בשפה ובתרבות אחרת. אולם למעשה יש להבחין בין שלושה מונחים המגדירים הליכים שונים הנמצאים ברצף, החל מהליך "אימוץ" (adoption), אחר כך הליך "התאמה" (adaptation) וכלה בהליך "הרכבה" (assembly) (Hambleton & Patsula, 1998, 1999;). בתיבה 3 שלהלן מוצגים שלושת ההליכים. (Malada et al., 2008).

תיבה 3 : מאימוץ להרכבה – התרגום וההתאמה על פני רצף

תהליכי התרגום וההתאמה של מבחנים ושל כלי הערכה שונים הם אוסף של הליכים על פני רצף, לפי מורכבותם. בקצה האחד של הרצף נמצא הליך האימוץ (adoption), שהוא התרגום הפשוט ובקצה השני נמצא הליך הפיתוח מחדש בשפת המטרה, הליך ההרכבה (assembly). יש להביא בחשבון שלא תמיד הגבול בין סוגי ההליכים השונים ברור וחד.

להלן תיאור קצר של שלושת ההליכים:

הליך האימוץ (adoption). הליך **תרגום מדויק**, ישיר ופשוט של הפריטים משפת המקור לשפת היעד. הליך כזה מיושם למשל במבחנים העוסקים בצורות, שבהם הצורות נשארות כמו שהיו בשפת המקור, אבל נדרש לתרגם את ההנחיות לבוחן, את ההוראות לנבחן ואת המחווונים.

הליך ההתאמה (adaptation). הליך הנמצא בתווך בין הליך האימוץ ובין הליך ההרכבה ויש בו מרכיבים מכל אחד מהם: תרגום מדויק וישיר של חלק ממרכיבי המבחן אם מניחים שהם יתפקדו באופן דומה בתרבויות השונות (חלק מהפריטים וכן ההנחיות לבוחן וההוראות לנבחן), והתאמה של פריטים שבהם אימוץ כפשוטו (תרגום פשוט) אינו מתאים מסיבות תרבותיות, שפתיות או פסיכומטריות. בהליך זה נדרשים לעיתים שינויים קלים בתכנים כדי שיתאימו לשפה ולתרבות המקומית, ולעיתים שינויים גדולים או פיתוח פריטים חדשים במקום אלה שאינם מתאימים כלל לרוח המבחן בשפת המטרה.

הליך ההרכבה (assembly). הליך **פיתוח מחדש** של רוב הפריטים או של כולם ו/או פיתוח מחדש של גריינים כך שיתאימו לשפת המטרה ולתרבות המטרה. פיתוח כזה נעשה בדרך כלל אם מניחים כי התרגום או ההתאמה של הכלי אינם אפשריים, או שהתוצרים של שני ההליכים האלה חסרי משמעות בשפת המטרה ו/או חסרי תוקף (למשל כלי המודד בסופו של דבר תכונה שונה מזו הנמדדת במקור). כדי לפתח מבחן נדרש להבין היטב את המבנה (construct) הנמדד במבחן המקורי ולהכיר היטב את הרקע התאורטי שלו. כמו כן יש להכיר את התהליך של בניית כלי המדידה המקורי, כלומר את הרציונל ואת מכלול השיקולים שהביאו ליצירת הכלי הספציפי ולבחירת הפריטים שבו. אם מידע כזה אינו זמין, יש לערוך ניתוח מעמיק לאחור של הכלי המקורי (מעין תהליך פירוק והרכבה, כמו reverse engineering), ולגזור את עקרונות הפיתוח של המבחנים השונים. בחלק מהמקרים יש לגזור את ההגדרה של היכולת הנמדדת וגם את אופן מדידתה. לפי ניתוח זה יש לבנות את הכלי מתוך התחשבות במגבלות השפתיות והתרבותיות של שפת המטרה. כמו כן יש להבין מהם הפרמטרים המשפיעים על רמת הקושי של הפריטים כדי לבנות פריטים ברמות קושי שונות. הליך הרכבה נדרש בדרך כלל במבחנים הבודקים ידע מגובש (למשל מבחני ידע כללי ומבחנים שפתיים) או כאלה המתבססים על ידע כזה. כמו כן נדרש הליך הרכבה במבחנים שבהם השפה היא משתנה דומיננטי. בהליך ההרכבה של מבחן נדרשות גם התאמות של ההנחיות לבוחן וההוראות לנבחן.

בתחילת הפרויקט נקבע לכל מבחן ההליך הנדרש במעבר לתרבות ולשפה בישראל (העברית והערבית): אימוץ, התאמה או הרכבה. לכל אחד ממבחני המקור נערכו ניתוחים של התוכן ושל המבנה התאורטי הנבדקים בו, נשקלו אפשרויות התאמתו לשפות המטרה ונבחר ההליך המתאים ביותר. לאור ניתוחים אלה סוכם הנחיות ראשוניות ושלבי העבודה בכל אחד מהמבחנים, ונקבעו דגשים ספציפיים בכל אחת מהשפות. כמו כן לכל מבחן נקבעה מידת הדמיון הנדרש בין הנוסח בעברית לנוסח בערבית, ובין שני הנוסחים האלה לנוסח המקור. במקרים שבהם לא היה אפשר לאמץ מבחן על כל פריטיו, הוחלט על בסיס אילו קריטריונים יש לשמר פריטים מן המקור או לערוך בהם התאמות תרבותיות. אם הוחלט על התאמות תרבותיות, נקבעו ההתאמות שייעשו. בתכנון הליכי התרגום וההתאמה דנו גם במחווה בהקשר התרבותי והשפתי שלו (למשל בפריטי מבחנים שבהם ההיבט המילולי-שפתי הוא הדומיננטי, נקבע מה ייחשב תשובה נכונה ומה ייחשב תשובה שגויה). היבטים תרבותיים נשקלו גם בנוגע להתאמת ההנחיות ובנוגע להתאמת העזרים הנדרשים בחלק מהמבחנים, כגון הקלטות שמע ודפי עבודה לנבחן.

יש לציין כי פותחו פריטים חדשים גם במבחנים שבהם נעשה הליך אימוץ או הליך התאמה כדי להגדיל את מספר הפריטים למקרה שיתברר בהמשך הפרויקט כי מסיבה כלשהי חלק מהפריטים ממבחן המקור בארצות הברית מתפקדים אחרת בישראל (למשל רמת קושי שונה או תפקוד לקוי).

בניתוח התוכן של המבחנים נמצאו 10 מבחנים שנדרש בהם הליך אימוץ, 5 מבחנים שנדרש בהם הליך התאמה ו-12 מבחנים שנדרש בהם הליך הרכבה/פיתוח⁵. המבחנים שאומצו היו ברובם מבחנים צורניים (צורות ולא ציורים) או מספריים. אומנם במבחנים אלה אומצו הפריטים מגרסת המקור, אך כאמור גם להם פותחו פריטים חדשים כדי להגדיל את מספר פריטים. בקבוצת המבחנים שבהם הוחלט לבצע התאמה, היו גירויים חזותיים (ציורים) שלא התאימו לתרבויות בארץ. קבוצת המבחנים שפותחו מחדש היו בעיקר אלה המושגות על שפה. המבחנים האלה בודקים יכולת מילולית (למשל אוצר מילים, מילים נרדפות, הפכים), חשיבה מילולית (למשל אנלוגיות) ועיבוד שמיעתי של גירויים שפתיים (מילים). מבחנים אחרים שפותחו מחדש הם מבחנים הבודקים יכולות כמו זיכרון לטווח קצר וזיכרון לטווח ארוך, שבהם בדיקת היכולת נשענת באופן דומיננטי על השפה (מילים, משפטים). נוסף על כך נערך הליך התאמה של ההוראות לנבחן. בהליך זה הושם דגש על חידוד ההוראות כדי שיהיו חד-משמעיות ועל הפשטות והבהירות שלהן, בעיקר לנבחנים צעירים. בכל המבחנים נעשה ניסיון לשמור על העיצוב הגרפי של מבחן המקור, אבל חלק מן הציורים היו מעט אנכרוניסטיים ולא התאימו לתקופה וחלק מהם לא התאימו לתרבות בישראל. לכן במבחנים מסוימים נעשו שיפורים גרפיים והתאמות למה שמוכר היום בישראל.

⁵ תת-המבחנים של מבחן 1 (4 תת-מבחנים) ושל מבחן 8 (2 תת-מבחנים) נמנו כאן כמבחנים נפרדים מכיוון שבכל אחד מהם התבצעו תהליכי "הרכבה" מורכבים; לכן נמנו כאן 27 מבחנים ולא 23.

בתהליך הפיתוח הושקעה חשיבה רבה בבניית חוברת הציון ודפי העבודה לנבחן. הם נכתבו ועוצבו מחדש בגרסאות בישראל כך שיהיו נוחים וידידותיים לבוחנים שישתמשו במבחן בעתיד.

בטבלה 2 שלהלן מוצג ההליך שנערך בכל אחד ממבחני וודקוק-ג'ונסון III, שתורגמו והותאמו לעברית ולערבית בפרויקט זה.

טבלה 2: תיאור הליכי ההתאמה במבחנים שהותאמו לשפות העברית והערבית⁶

המבחן	תת-המבחן	אופן ההתמודדות בפיתוח	הערות
1 הבנה מילולית	א1 אוצר מילים על-פי תמונות	הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוזנים. עודכנו ציורים כדי שיתאימו לתרבות בישראל והוחלפו ציורים רבים. נעשה ניסיון לבחור פריטים המתאימים לשתי השפות.
	ב1 מילים נרדפות	הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוזנים. רוב הפריטים נכתבו מחדש, שכן פריטי המקור לא התאימו לישראל ולא אפשרו למדוד את מה שהמבחן היה צריך למדוד.
	ג1 הפכים	הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוזנים. רוב הפריטים נכתבו מחדש, שכן פריטי המקור לא התאימו לישראל ולא אפשרו למדוד את מה שהמבחן היה צריך למדוד.
	ד1 אנלוגיות מילוליות	הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוזנים. רוב הפריטים נכתבו מחדש, שכן פריטי המקור לא התאימו לישראל ולא אפשרו למדוד את מה שהמבחן היה צריך למדוד.
2 למידה חזותית שמיעתית		הרכבה	<u>פיתוח חדש</u> נכתבו סמלים חדשים לצד שימוש בחלק מסמלי המקור ונכתבו "משפטים חדשים". הוחלט לפתח גרסאות שישענו על המבנה התחבירי של השפות עברית וערבית. כתוצאה מכך נכתבו בכל אחת מהשפות "משפטים" חדשים ונבחרו סמלים שיתאימו להם.

⁶ עוד על הליכי האימוץ, ההתאמה וההרכבה, ראו בתיבה 3 לעיל.

המבחן	תת-המבחן	אופן ההתמודדות בפיתוח	הערות
3 יחסים מרחביים		אימוץ	<u>אימוץ מהמקור ותוספת בשיעור של כ-30% פריטים חדשים</u>
4 מיזוג צלילים		הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוננים. כדי להתאים את המבחן לשפות המטרה, נדרש ניתוח של דרך פירוק המילים במבחן המקור. בעקבות זאת נבחרו מילים ייעודיות מהמקור, והמילים בעברית ובערבית פורקו לפי אותם עקרונות של שפת המקור.
5 יצירת כללים		אימוץ	<u>אימוץ מהמקור ותוספת בשיעור של כ-20% פריטים חדשים</u> הותאמו ההנחיות לבוחן וההוראות לנבחן, שופרו ההנחיות וההוראות במקומות בודדים ועודכנו התשובות במחונן.
6 התאמה חזותית		אימוץ	<u>אימוץ מהמקור</u>
7 זכירת מילים		הרכבה	<u>תרגום פריטים בודדים ובעיקר פיתוח חדש של פריטים ופריטים עודפים</u> נבחרו מילים המתאימות למבחן המקור מבחינת משלב הלשון, מספר ההברות ועוד.
8 ידיעות כלליות	א8 איפה אפשר למצוא?	הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוננים. כדי לבדוק את הידע השפתי והפונקציונלי במבחן, נדרשה התאמה לשפה ולתרבות בישראל. לכן רוב הפריטים פותחו בנפרד לכל אחת משפות המטרה.
	ב8 בשביל מה משתמשים ב...?	הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוננים. כדי לבדוק את הידע השפתי והפונקציונלי במבחן, נדרשה התאמה לשפה ולתרבות בישראל. לכן רוב הפריטים פותחו בנפרד לכל אחת משפות המטרה.
9 שטף סמנטי		הרכבה	<u>פיתוח חדש של חלק מהפריטים</u> שנו 2 פריטים מתוך 3 שהיו במבחן המקור. נכתב מחונן חדש לנוסחים בעברית ובערבית.
10 זיהוי ציורים		התאמה	<u>התאמה גרפית של חלק מהציורים לציורים מודרניים ותוספת של פריטים עודפים</u>
11 מילים לא שלמות		הרכבה	<u>פיתוח חדש</u> נכתבו פריטים חדשים, פריטים עודפים ומחוננים. נדרש ניתוח של מבחן המקור כדי להתאים אותו לשפות המטרה. נבחרו מילים מתאימות בשפות המטרה שיש בהן אופני השמטה של חלקי מילה

המבחן	תת-המבחן	אופן ההתמודדות בפיתוח	הערות
			כפי שהיו בשפת המקור. הובאו בחשבון המאפיינים הלשוניים של שפות המטרה.
12 אנליזה- סינתזה		אימוץ	<u>אימוץ מהמקור</u>
13 מהירות החלטה		התאמה	<u>התאמה ושינוי בשיעור של כ-10% מהציורים משיקולי התאמה תרבותית</u> <u>הוחלפו ציורים מהמקור בציורים עדכניים.</u>
14 זכירת ספרות לאחר		אימוץ	<u>אימוץ מהמקור</u>
15 זכירת שמות		התאמה	<u>החלפה של כל שמות הפריטים במבחן</u> <u>פותחו שמות משותפים לעברית ולערבית</u> <u>שהתאימו להגייה בשפות אלה. נבחרו מילים שאינן שמות אמיתיים בעברית ובערבית.</u>
16 תכנון		אימוץ	<u>אימוץ מהמקור ותוספת של פריטים עודפים בשיעור של 25%</u>
17 קשב שמיעתי		הרכבה	<u>פיתוח חדש</u> <u>נותחו פריטי המקור מבחינת אורך המילה ומבחינת דמיונה למילים המסיחות. כמו כן נבדק עד כמה קל לזהות את המילה בציור. כל הפריטים נכתבו מחדש בכל אחת משתי שפות המטרה, וציורו ציורים חדשים לכל אחד מהפריטים האלה.</u>
18 סדרות מספרים		אימוץ	<u>אימוץ מהמקור ותוספת של פריטים חדשים בשיעור של 25%</u>
19 איתור ציורים מהיר		אימוץ	<u>אימוץ מהמקור</u>
20 זיכרון עבודה שמיעתי		התאמה	<u>תרגום פריטים בודדים ובעיקר פיתוח חדש של פריטים ופריטים עודפים</u>
21 שיום תמונות מהיר		התאמה	<u>התאמה של חלק מהציורים או החלפה של חלק מהציורים לציורים עדכניים המותאמים מבחינה תרבותית</u>
22 סיבוב עצמים		אימוץ	<u>אימוץ מהמקור ותוספת של פריטים עודפים בשיעור של 25%</u>

המבחן	תת-המבחן	אופן ההתמודדות בפיתוח	הערות
23 זכירת שמות מושהית		אימוץ	שימוש בשמות שפותחו במבחן 15 פותחו שמות משותפים לעברית ולערבית שהתאימו להגייה בשפות אלה. נבחרו מילים שאינן שמות אמיתיים בעברית ובערבית.

ג. הליך תרגום והתאמת המבחנים

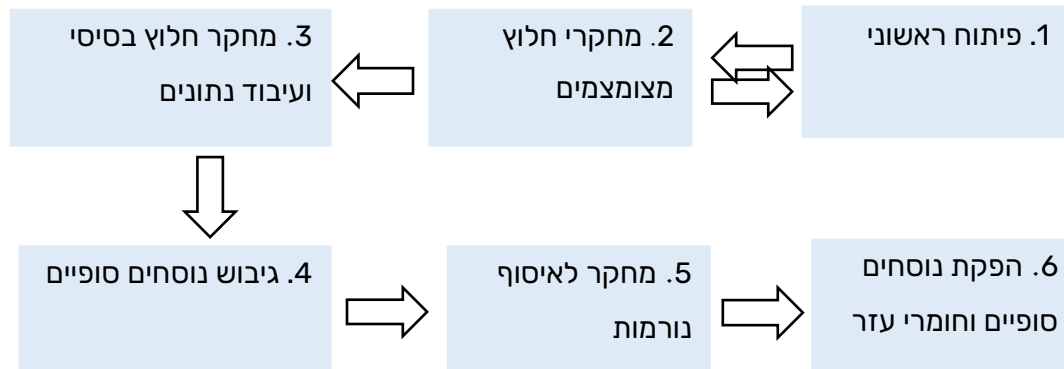
ג.1. שלבי הפרויקט

פרויקט התרגום וההתאמה כלל את השלבים האלה:

- שלב 1: הליך ראשוני של אימוץ/התאמה/הרכבה (על פי מה שנקבע בתחילת הפרויקט) של מבחן המקור לעברית ולערבית (לרבות פיתוח פריטים עודפים).
- שלב 2: מחקרי חלוץ מצומצמים ("מיני-פילוט") – נוסחי המבחן הועברו בגרסתם הראשונית למספר מצומצם של נבחנים דוברי כל אחת מהשפות, בכמה קבוצות גיל. נערכה בדיקה איכותנית של תקינות הפריטים, לרבות בדיקת התקינות וההתאמה של המחוונים. נבדקו בהירות ההוראות ומידת הבנתן על ידי הנבחנים. תוקנו המבחנים ונמשך הפיתוח בהתאם לממצאים. המבחן הועבר פעם נוספת לכמה נבחנים אחרים. המבחנים תוקנו, עודכנו ונמשך הפיתוח בהתאם לממצאים.
- שלב 3: מחקר חלוץ בסיסי ("פילוט") – המבחנים הועברו לקבוצה נרחבת של נבחנים בקבוצות גיל שונות. בוצע עיבוד נתונים לפי הדגשים ספציפיים בכל מבחן. נעשתה אינטגרציה בין הממצאים של ניתוחי הפריטים והמבחנים ובין מידע איכותני שנאסף על הפריטים במחקר החלוץ.
- שלב 4: גיבוש נוסחים סופיים בהתאם לממצאים שעלו במחקר החלוץ – גובשו סופית ההנחיות לבוחן וההוראות לנבחן, נקבעו נקודות הכניסה והיציאה, והקריטריונים לגבול התחתון ולגבול העליון, הושמטו פריטים עודפים ונקבע מספר סופי של פריטים בכל מבחן.
- שלב 5: מחקר הנוגע לאיסוף הנורמות – המבחנים הועברו למדגם מייצג של אלפי נבחנים מכל קבוצות הגיל ומכל המגזרים המייצגים את אוכלוסיות המטרה. חושבו הנורמות באוכלוסיות המטרה על סמך ממצאי המחקר. מחקר זה מתואר במפורט בדוח נפרד.
- שלב 6: הפקת אוגדני המבחן וחומרי העזר – דפי העבודה לנבחן, חוברת הציון, קובצי השמע, השקפים לציון והמדריך לבוחן, כל אלה הוכנו לקראת הוצאתם לאור והטמעתם במערכת החינוך.

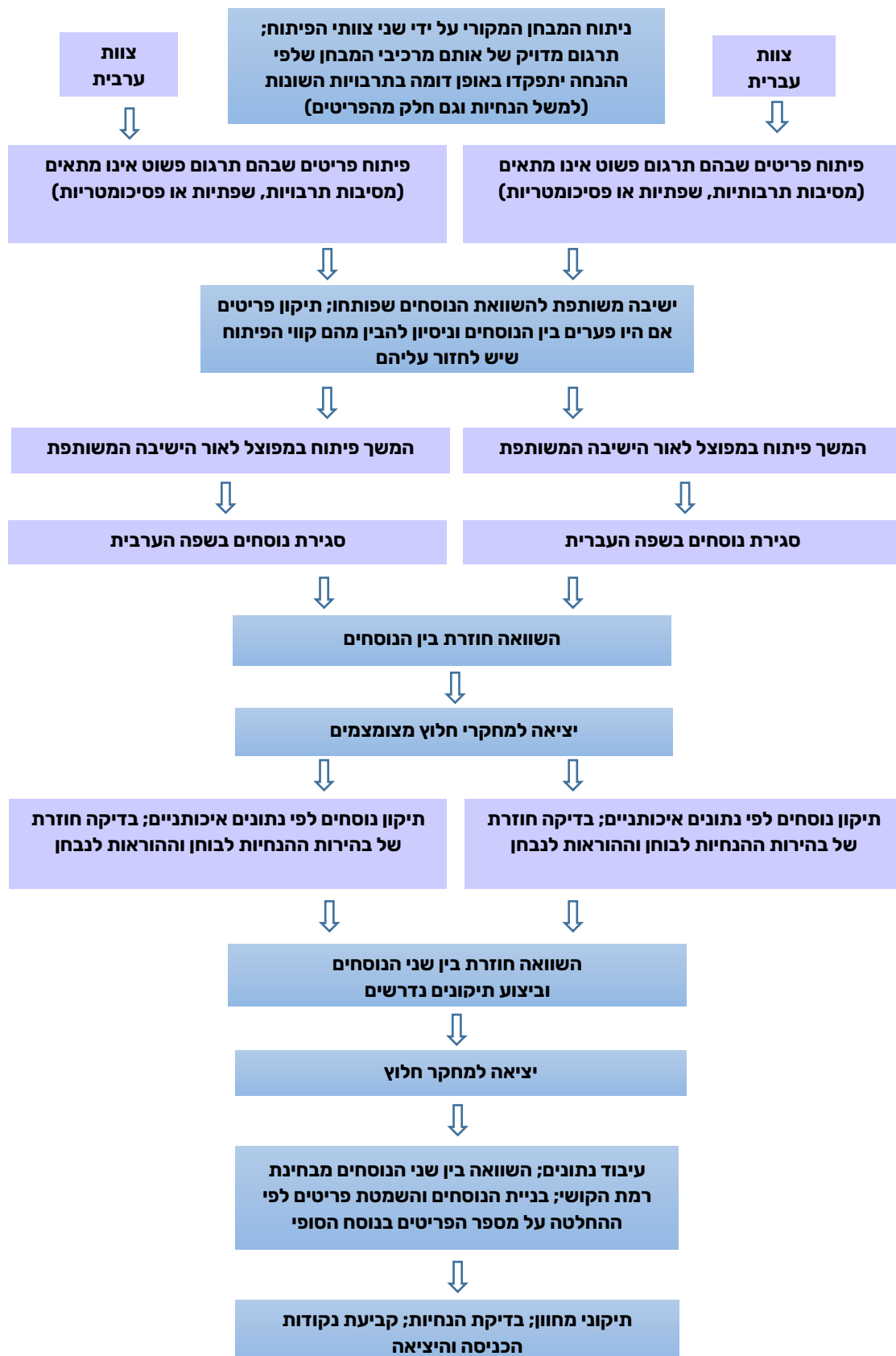
בתרשים 1 שלהלן מתוארים שלבי הפרויקט באופן כללי.

תרשים 1: השלבים העיקריים בהליך התרגום וההתאמה של המבחנים

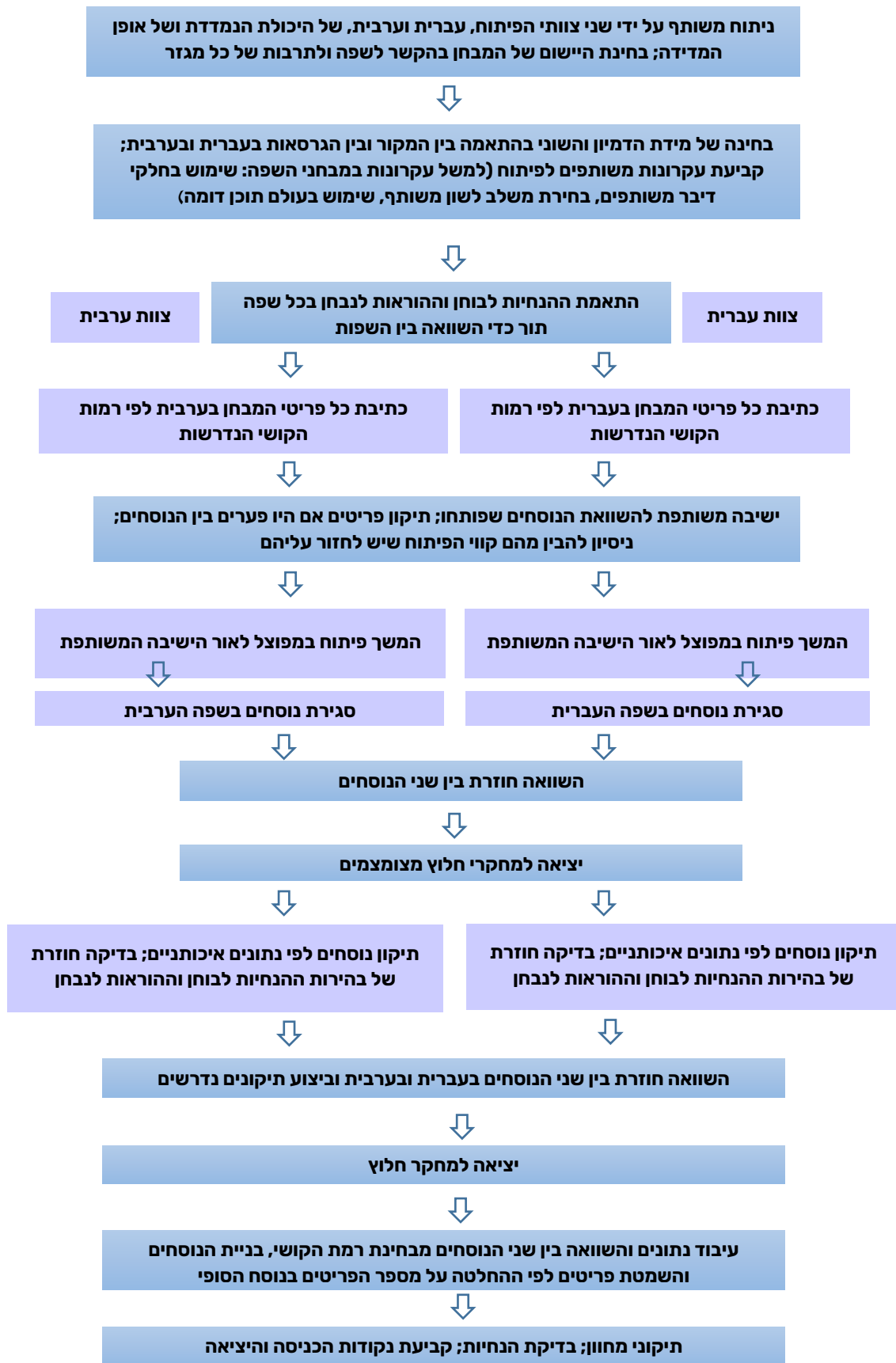


בתרשים 2 ובתרשים 3 שלהלן מתוארים בפירוט שלבי האימוץ/התאמה/הרכבה כפי שיושמו בפרויקט זה של התאמת מבחן וודקוק-ג'ונסון III לעברית ולערבית בישראל. כאמור לעיל, תהליכים אלו בוצעו במקביל ובנפרד בשתי השפות, בתיאום מלא.

תרשים 2: הליך התרגום וההתאמה של מבחן בוודקוק-ג'ונסון III למבחן בישראל



תרשים 3: הליך ההרכבה (הפיתוח) של מבחן בוודקוק-ג'ונסון III למבחן בישראל



ג.2. סוגיות מרכזיות בתהליך התרגום וההתאמה

לא מעט דילמות והתלבטויות ליוו את העבודה בפרויקט, לרבות בהליך הבירור והניתוח הראשוני. חלקן היו כלליות וחלקן היו ייחודיות לכל מבחן. הדילמות והסוגיות העיקריות בפרויקט התרגום וההתאמה של מבחן וודקוק-ג'ונסון III לישראל מתוארות להלן. בחלק מן הנושאים מובאות דוגמות קונקרטיות כדי להמחיש סוגיה כלשהי אבל כמובן, אין הדוגמות ממצות את כל המקרים.

ג.2.א. מידת ההיצמדות לתוכן של מבחן המקור

בפיתוח הפריטים במבחנים השונים הוגדר מראש אם לאמץ את פריטי המקור או לפתח חדשים. כאמור לעיל, ההגדרה בכל מבחן נבעה מרמת ההתאמה הנדרשת (אימוץ/התאמה/הרכבה) וממידת התאמתו של תוכן מבחן כלשהו במקור לשפות ולתרבות בישראל. השאיפה הייתה לשמר ככל האפשר את התכנים המקוריים. עם זאת היה צריך לאזן בין השאיפה הזאת ובין השאיפה למדוד בשפת המטרה את מה שהפריט מודד בשפת המקור, ולהקפיד שהפריט החדש יתפקד כראוי מבחינה פסיכומטרית ומבחינת ההיבטים השפתיים והתרבותיים שלו. בחלק מן המבחנים היה פער בין שתי השאיפות האלה. למשל במבחני הידע המגובש נבחן כל פריט מהמקור אם אפשר לאמץ אותו. אם היה אפשר, בוצע הליך אימוץ תוך כדי התאמה מינימלית, כלומר תרגום הפריט לשפת המטרה והתאמה רק של המחוון ושל שאלות הבירור. למרות ההנחיה לאמץ פריטים רבים ככל האפשר, אומצו בפועל רק פריטים בודדים. דבר זה נבע בעיקר בשל הבדל בשלבי רכישת השפה בגילים הצעירים, וכמו כן בשל הבדלים במשלב הלשון, בשכיחותן של המילים וברקע התרבותי – בין שפת המקור ובין שפות המטרה.

במבחנים המודדים עיבוד וקשב שמיעתי (למשל מבחן 4) לא היה אפשר כלל לאמץ פריטים מהמקור מפאת השוני בין שפת המקור ובין שפות המטרה (פריטי המבחן הם מילים). עם זאת העקרונות שלפיהם פותחו המבחנים והפריטים במקור באנגלית (ככל שהיה ידוע) היו גם העקרונות לפיתוח הפריטים בעברית ובערבית. העקרונות יושמו מתוך התחשבות בשונות הלשונית והתחבירית בין שפות המטרה ובין שפת המקור. כמו כן הוחלט על עלייה הדרגתית ברמת הקושי של הפריטים, וכדי להשיגה נדרשה הבנה טובה של הפרמטרים המשפיעים על דרגות הקושי של הפריטים בשפת המקור.

במבחנים שבהם הוחלט על הליך אימוץ, אומצו רוב הפריטים כפשוטם, למשל ציורים או צורות. פריטים שונו או הוחלפו רק אם התברר במהלך הפיתוח (למשל במחקרי החלוץ המצומצמים) שהפריטים אינם מתאימים מבחינה תרבותית, או שאינם מוכרים דיים בישראל, או שאינם שכיחים דיים בישראל (למשל "כדור בייסבול"), או שאינם מתאימים מבחינה ערכית (למשל, "חזיר"), או שמשהו בהם מטעה (למשל ציור "מנעול" שעוצב בסגנון שנות ה-70 והקשה על נבחנים צעירים לזהות אותו). לבסוף תוקנו או הוחלפו פריטים שנמצאו מוטעים מבחינת תרבותית ושפתית (כלומר קלים יותר או קשים יותר באופן מובהק מסיבות לשוניות הקשורות לשפה העברית או לשפה הערבית, או מסיבות אחרות). היו כמה מבחנים שבהם תוכנן תחילה הליך אימוץ אך בסופו דבר בוצע הליך הרכבה. במבחני הזיכרון לטווח קצר (כמו המבחן "זכירת מילים"), שבו נדרש הנבחן לחזור על מילים מושמעות לפי סדר השמעתן,

בוצע הליך הרכבה. התברר כי יש להחליף את המילים במבחן במילים אחרות כדי להתאימן לעברית ולערבית מבחינת השכיחות שלהן בשפה, מבחינת משלב הלשון שלהן ומבחינת אופן הגייתן.

ג.2.ב. פיתוח פריטים עודפים

ברוב המבחנים הוחלט על פיתוח פריטים עודפים בהיקפים משתנים. הרציונל לפיתוח פריטים עודפים היה לאפשר שינויים בהרכב הסופי של המבחנים לפי הצורך (למשל בשל שינויים ברמת הקושי של הפריטים בישראל לעומת רמת הקושי של הפריטים בארצות הברית, או אם חלק מהפריטים בגרסת המקור אינם מתפקדים כראוי מבחינה פסיכומטרית⁷). כמו כן עלה הצורך לאפשר בחירה רחבה יותר למפתחים בעת בניית הנוסחים כדי שיהיה אפשר ליצור נוסחים מאוזנים מבחינת רמות הקושי בתוך קבוצות הגיל. בבניית הפריטים העודפים הוחלט לשמר את אופי התכנים של מבחן המקור, ולכן פותחו פריטים חדשים מאותם עולמות תוכן (למשל במבחני השפה כלי נשיפה, כלי עבודה ביתיים, מילים נרדפות שיש להן שדה סמנטי משותף, שמות תואר המבטאים רגשות דומים, פעלים המתארים פעולות דומות). מספר הפריטים העודפים שנדרש לפתח בכל מבחן נקבע לפי מספר השינויים הצפויים במבנה המבחן ובפריטיו בעקבות התרגום וההתאמה, ולפי מידת המורכבות של התאמת המבחן.

במבחנים שעברו הליך הרכבה הובא בחשבון כי פריטים מסוימים שפותחו ייפסלו בהמשך הפיתוח, למשל בשל היבטים פסיכומטריים. לכן בשל מגבלות הזמן פותח מאגר גדול של פריטים שצומצם בהמשך הפיתוח. דבר זה אפשר לבחור את הפריטים הטובים ביותר ולהשמיט את אלה הטובים פחות (לפי הממצאים בשלבי הפיתוח השונים). יש לציין שדרך זו טובה יותר מפיתוח מספר מדויק של פריטים ותיקונם לפי הצורך. הסיבה לכך היא שאם יידרש לפתח פריט חדש, יהיה צורך לחזור על השלבים הניסיוניים של העברה ובדיקה, וכל אלה יפגעו בלוח הזמנים של הפיתוח. בסופו של תהליך הפיתוח הראשוני לשלב של מחקרי החלוץ המצומצמים נכתבו לכל מבחני השפה פריטים בשיעור של 200%, לכל מבחני הקשב השמיעתי – פריטים בשיעור של כ-150% ולשאר המבחנים – פריטים בשיעור של 120%-150%.

כאמור לעיל, גם כדי לפתח פריטים עודפים במבחנים שאומצו או שהותאמו קלות (במבחנים הצורניים), נדרש ניתוח מעמיק והבנה ברורה של המבנה (construct) הנמדד במבחן: כיצד סוג פריטים ספציפי מודד את המבנה ומה הם הפרמטרים המשפיעים על רמות הקושי של הפריטים. את הפריטים החדשים היה צריך לפתח כך שיהיו דומים במבנה ובצורה ליתר הפריטים כדי שישתלבו "בטבעיות" במבחן הקיים וכדי שייראו חלק בלתי נפרד מהמבחן. כמו כן הפריטים היו צריכים להיות

⁷ ועדת ההגוי הנחתה לפתח פריטים עודפים גם במבחנים הצורניים, שכן אפשר ללמוד מהניסיון ומהתיעוד בספרות כי בהליכי תרגום והתאמה לארצות ולתרבויות אחרות, פריטים במבחנים צורניים (שאינם תלויי שפה) תפקדו לעיתים פחות טוב מאשר במבחן המקור. לתופעה זו אין הסבר מחקרי מספק, אבל קיומה חייב לפתח פריטים עודפים בשיעור של כ-25%.

ברמת קושי המותאמת לגיל (במקרים שבהם היו פריטים קלים מאוד או קשים מאוד). פריטים אלה עברו בהמשך הליכי בקרה ונוסו כמו יתר פריטי המבחן בשלבים הניסיוניים השונים.

ג.2.ג. מספר הפריטים בכל מבחן

עם תחילת הפיתוח ובמהלכו עלתה השאלה האם מספר הפריטים בכל מבחן יהיה כמו מספרם בנוסח המקור, או שיהיה אפשר להאריך או לקצר את המבחן בהתאם לצרכים בישראל. הוחלט שמספר הפריטים יהיה כמו במבחן באנגלית כדי לשמור על הדמיון למבחן המקור. כמו כן הוחלט על מספר זה של פריטים בנוסח העברי ובנוסח הערבי. למרות זאת במבחנים אחדים הוגדל מספר הפריטים לעומת המקור ובמבחן אחד צומצם מספר הפריטים לעומת המקור. שינויים אלה נעשו על סמך עיבודי הנתונים במחקר החלוץ ובשל הצורך לבנות מבחנים שיבחנו היטב גם בין נבחנים המצויים בקצוות (נבחנים צעירים מאוד ונבחני החינוך המיוחד לעומת נבחנים בעלי יכולת גבוהה). ניתוח הנתונים במחקר החלוץ הצביע על מחסור בפריטים קשים מאוד ועל מחסור בפריטים קלים מאוד (ולעיתים גם על מחסור בפריטים בינוניים). לכן נדרש להוסיף פריטים ו"להגדיל" את יכולת המדידה של הכלי ואת תוקפו בקצוות. (עוד על מספר הפריטים בכל שלבי הפרויקט, ראו בטבלה בסעיף ד.4. להלן).

ג.2.ד. התאמת ההנחיות לבוחן והוראות לנבחן

במבחנים רחבי היקף, שבהם חשוב לשמור על סטנדרטיזציה מרבית בתנאי ההעברה, יש חשיבות קריטית להוראות הנאמרות לנבחן על ידי הבוחן (או המוקראות לנבחן). הוראות שאינן אחידות, שאינן ברורות או שמטעות יכולות לפגוע במדידה ולמנוע מן הנבחן לבטא את יכולותיו באופן מיטבי. הוראות מפורטות מדי או מכוונות מדי עלולות לשפר את ביצועי הנבחן עד כדי כך שתוקף המדידה ייפגע. לכן תרגום ההוראות הוא הליך קריטי בהתאמת מבחן משפה אחת לאחרת. כבר בראשית הפרויקט זוהה הליך זה כהליך מושכל המצריך עבודה משולבת של פסיכומטריקאים, של פסיכולוגים בעלי ניסיון מעשי בשדה האבחון ושל עורכים לשוניים. הקושי התעורר במיוחד בשל הצורך להתאים את ההוראות לקבוצות הגיל השונות, 3.5 עד 18. ההוראות היו צריכות להיות ברורות ונהירות לנבחנים בכל קבוצות הגיל: שלא יישמעו ילדותיות מדי מצד אחד ושלא ינוסחו במשלב לשון גבוה מדי מצד אחר. דגש מיוחד הושם על הניסוח שיהיה פשוט, ברור ומותאם לשפה הטבעית המדוברת של ילדים. ההתמודדות המרכזית בהקשר זה הייתה כיצד לשמור על התקינות הלשונית אם המבחן הוא "מדובר". היו מקרים שבהם הייתה התלבטות בין ניסוח בשפה תקינה, שהיה במשלב לשון גבוה ובמבנה תחבירי מורכב, ובין שפה המותאמת לילדים. במקרים אלה הייתה העדפה לניסוח המתאים לשפת הדיבור היומיומית של הנבחנים ולמשפטים קצרים וברורים, על פני ניסוחים לשוניים תקינים מבחינה אקדמית אך גבוהים מדי. נוסף על כך הושקע מאמץ מיוחד כדי לשמור על אחידות התכנים ועל אחידות הסגנון בין ההוראות לנבחן בעברית ובין ההוראות לנבחן בערבית.

להלן שלבי הליך תרגום ההוראות:

- ההוראות של כל מבחן עברו תרגום בסיסי לשתי השפות על ידי מתרגם מוסמך. התרגומים האלה היו הבסיס להמשך תהליך העבודה בשתי השפות.
- אחרי התרגום עברו ההוראות התאמה ובמידת הצורך נוסחו מחדש על ידי צוותי הפיתוח, בשילוב עורכי לשון בשתי השפות. ההוראות בעברית הושוו לאלה שבערבית. בניסוח ההוראות בערבית נבחנו הן הנוסח המוצע בעברית, הן נוסח המקור באנגלית.
- במבחנים שבהם היה מלל רב בהוראות לנבחן (למשל במבחן 5 – "יצירת כללים"), דנו מחדש על ניסוח ההוראות. ההוראות נבחנו בתשומת לב מיוחדת, במחקרי החלוץ המצומצם ובמחקר החלוץ, ונבדק האם הן ברורות דין.
- לאחר שגובשו נוסחים מוסכמים של ההנחיות לבוחן וההוראות לנבחן, בכל מבחן, התקיימה ישיבה משותפת להאחדת ההנחיות וההוראות בעברית ובערבית, בהשתתפות כל אנשי הצוותים בשתי השפות. בדיונים אלה חודדו ניסוחים שונים והתקבלו החלטות נקודתיות במקומות שבהם היו קשיי ניסוח או מחלוקות לגבי הניסוח הרצוי.
- ההנחיות לבוחן וההוראות לנבחן בכל המבחנים, בשתי השפות, עברו בקרת איכות על ידי ראמ"ה.

ג.2.ה. סוגיית החינוך המיוחד

במטרה למדוד את היכולת של הנבחנים תלמידי החינוך המיוחד ולתת מענה לשונות הרבה בקבוצה זו הוגדל מספר הפריטים הקלים ביותר בחלק מהמבחנים ונעשתה בדיקת התאמה של ההוראות ושל הפריטים במבחנים לתלמידי החינוך המיוחד. אוכלוסייה זו לא הייתה חלק במחקר החלוץ מטעמי נוחות אך הוותה חלק ממחקר הנורמות. (על שיתוף נבחני החינוך המיוחד, ראו בדוח העוסק במחקר הנורמות).

ג.2.ו. שאלות ברור

במבחנים הבודקים כישורי חשיבה, כמו מבחן וודקוק-ג'ונסון, ובמיוחד במבחנים הבודקים ידע מגובש, נהוג לשאול את הנבחן שאלות ברור לפי הצורך ולפי שיקול דעתו של הבוחן. טיב שאלות הברור, עומקן ומספר הפעמים שהבוחן חוזר עליהן יכולים להטות את תוצאות המבחן ולפגוע בתוקף שלו. למשל ניסוח מנחה או ניסוח ישיר מדי של שאלת הברור יכולים להיות לפעמים רמז לנבחן. ניסוחים כאלה עלולים לסייע לו להשיב על פריט או שהם יכולים להשפיע על תגובותיו בפריטים ובמבחנים הבאים. לכן כדי לשמור על תנאי העברה סטנדרטיים ככל האפשר, יש להגדיר את המקרים שבהם הבוחן רשאי לברר את תגובת הנבחן. לפי הכללים המוגדרים במבחן וודקוק-ג'ונסון III המקורי, הבוחן רשאי לשאול את הנבחן שאלת ברור אם תשובת הנבחן עמומה, אם הבוחן אינו בטוח כיצד עליו לציין את התשובה ואם התשובה קרובה לתשובה הנכונה, אך אינה מדויקת דיה, כדי לקבל עליה ניקוד. למשל אם הנבחן עוסק בתשובתו רק בחלק מהגירוי, או במבחן של זיהוי ושיום חפצים, אם הנבחן עונה בשם פעולה במקום לשיים את החפץ.

ניסוח שאלות הבירור היה אחד האתגרים הגדולים בהתאמת מבחני השפה לעברית ולערבית שכן נדרש ניסוח מדויק ומובנה שלהן. לשם כך נבדקו ניסוחים שונים של שאלות בירור, בהתאמה לרוח שאלות הברור בנוסח המקור, והשאלות הותאמו הן לפריטים מסוימים הן לקונטקסט הרחב של תוכן המבחן.

בכל מבחן ותת-מבחן ממבחני השפה התווספו לפריטים שאלות בירור, בייחוד לפריטים שבהם התגלה במחקר החלוץ שיש צורך בהן. למשל בתת-המבחן 1א בפריט ציור "הר הגעש", אם הנבחן ענה "אש" (עסק רק בחלק מן הציור), נדרשה שאלת בירור מסוג "נסה שוב. מה זה?" במקביל לשאלת הבירור הזאת נדרש הבוחן להקיף את כל הציור באצבע כדי להראות לנבחן את האובייקט השלם שאותו עליו לשיים. באותו תת-מבחן, בפריט של ציור "מספריים", אם הנבחן ענה "גוזרים עם זה" (תשובה שעסקה בפעולה במקום בשם החפץ), ניכר כי הוא זיהה את האובייקט וידע מהו, אולם בשל גילו הצעיר או מסיבה אחרת, הוא הגדיר את החפץ כך. שאלת הבירור המותאמת במקרה הזה הייתה "איך קוראים לזה?" והנה עוד דוגמה: הוצג ציור "בת יענה" והנבחן ענה "חסידה". בעקבות תשובה זו שאלת הבירור הייתה "נסה שוב. מה זה?" ההשערה הייתה שהנבחן קרוב לתשובה הנכונה ומכיר אותה, אולם טעה בגלל פרטים בציור. הנחת היסוד של שאלות הבירור הייתה שהנבחן יודע את התשובה הנכונה, ואם תינתן לו הזדמנות נוספת או הכוונה מינימלית, הוא ישלוף אותה. יש לשים לב ששאלות הבירור נועדו למבחני 1 (הבנה מילולית) ולמבחני 8 (ידע כללי).

בעקבות מחקרי החלוץ המצומצים ומחקר החלוץ מופו הפריטים לפי מספר שאלות הבירור שהבוחן שאל בפריט כלשהו. על בסיס המיפוי שופרו ציורים, הושמטו פריטים שהתברר שלא תפקדו היטב (למשל הושמט הציור "מאובן" בתת-המבחן 1א, ציור שנראה לנבחנים רבים כשבלול או כשבר של כלי חרס). כמו כן התווספו הנחיות ברורות לבוחן בנוגע לשימוש מיטבי בשאלות הבירור והודגשו ניסוחים בפריטים מסוימים. (עוד על הנחיות לשאלות הבירור במבחנים מסוימים, ראו בפרק 4 של המדריך לבוחן.)

להלן ההנחיות הכלליות בנוגע לשאלות הבירור ולאופני השימוש בהם:

- ברוב המקרים מספיק לשאול שאלת בירור אחת, מכיוון שאם לאחר ניסיון נוסף עדיין הנבחן אינו עונה תשובה נכונה, יש להניח שאינו יודע אותה או אינו יכול לשלוף אותה. יש להימנע משאלות בירור רבות, שכן הן מכוונות את הנבחן לתשובה הנכונה ועלולות לפגוע בתוקף המבחן.
- במבחני 1 ובמבחני 8 ההחלטה אם לשאול שאלת בירור תלויה בשיקול דעתו של הבוחן בנוגע לתשובת הנבחן. שאלות הבירור אינן מוגבלות לפריטים מסוימים, ואפשר לשאול שאלת בירור אחת לכל פריט.
- במבחנים אלה, בהנחיות לבוחן שבאוגדן, כתובות דוגמות לתגובות שכיחות שיש לשאול עליהן שאלות בירור. דוגמות אלה מציגות תשובות שהן קרובות במידת מה לתשובה הנכונה, כדוגמת ה"חסידה" ו-"בת יענה" שהוזכרה לעיל. אפשר לשאול שאלת בירור גם אם התשובה שענה הנבחן אינה כתובה במחווה, לפי שיקול דעתו של הבוחן. אם שואלים שאלת בירור שאינה כתובה בפריט, יש לנסח את הפנייה בשאלה באותו נוסח כמו זה שבשאלות הבירור באותו מבחן, ולא

לנסח פנייה בניסוח חופשי. הסיבה לכך היא שניסוח חופשי של שאלת בירור יפגע בסטנדרטיות של העברת המבחן: יש חשש שהניסוח יהיה מנחה מדי או רומז מדי.

- בניגוד לכך, במבחן 5 אי אפשר לשאול שאלות הבירור בכל פריט והן אינן נתונות לשיקול דעתו של הבוחן, אלא הן מובנות בתהליך ההיבחנות הסטנדרטי של המבחן. הבוחן משתמש בהן כתגובה לתשובות ספציפיות של הנבחן, רק במקומות שבהם כתוב שיש לשאול שאלות ברור באוגדן.

ג.2.ז. תשובות בשפה אחרת משפת המבחן

ישראל היא מדינת הגירה ובמערכת החינוך שלה לומדים ילדים רבים שהם או הוריהם היגרו אליה ממדינה אחרת. עברית אינה שפת-האם שלהם או שהיא אחת משתי שפות-האם שלהם. לעיתים ילדים אלה מדברים וחושבים בשפה שאינה שפת המבחן (עברית או ערבית) וייתכן שהם מכירים את המילה, את הביטוי או את האובייקט שעליו הם נשאלים בשפה זו, אך אינם מכירים אותם בשפת המבחן. בדיון במבחני הידע המגובש וגם במבחנים אחרים עלתה הסוגיה הזאת לדיון: האם לקבל תשובות נכונות בשפה אחרת משפת המבחן? שכן, אם הנבחנים יודעים את המילה בשפת האם שלהם, אי אפשר לטעון שהם אינם מכירים את המילה, אך הם לא יודעים להגיד אותה בשפת המבחן. לאחר מחקרי החלוץ המצומצמים הוחלט לעקוב אחר סוגיה זו במחקר החלוץ ולאמוד את ממדי התופעה באמצעות תיעוד מלא של תגובות הנבחנים במבחנים השפתיים, בשתי הגרסאות – בעברית ובערבית. מניתוח התגובות שנאספו במחקר החלוץ התברר שיש אחוז כלשהו (נמוך יחסית) של תשובות בשפות אחרות. הדיון בסוגיית המענה בשפה אחרת העלה את שאלת היסוד מה נמדד במבחן, ועל סמך התשובה לשאלה זו נדונה השאלה האם לקבל או שלא לקבל תשובות בשפה אחרת. נוסף על כך עלתה שאלת הסטנדרטיזציה וסוגיית היכולת של הפסיכולוג הבוחן לציין תשובות בשפה שהוא עצמו אינו דובר אותה ואינו מבין אותה (למשל רוסית או צרפתית או ספרדית או אמהרית).

מהדיונים על התגובות שהתקבלו במבחני 1 (ידע מגובש ואוצר מילים), עלה שיש לחלק את מבחני 1 לשלוש קטגוריות מבחינת היכולת הנמדדת בהם: מבחן 1א – ידע עולם, תפיסת קונספט באמצעות תמונה, המבחנים 1ב, 1ג – ידע עולם, רוחב ועומק לקסיקלי, מבחן 1ד – ידע עולם ויכולת פלואידית (Reasoning). כמו כן הוחלט שיש לשייך את המבחנים 8א ו-8ב לקטגוריה הראשונה – ידע עולם (כמו מבחן 1א). על בסיס חלוקה זו הוחלט לקבל גם תשובות נכונות שנאמרו בשפה אחרת במבחנים 1א ו-1ד, שבהם נמדד ידע עולם ויכולת שיום ושליפה, ובמבחנים 8א ו-8ב, שבהם נמדד ידע עולם באמצעות שאלות הנוגעות לשימוש בחפצים או למיקום חפצים. הרציונל להחלטה זו היה שידע עולם ויכולת שיום ושליפה אפשר למדוד גם אם הנבחן עונה בשפה אחרת. לעומת זאת במבחנים 1ב ו-1ג, שבהם נמדד רוחב ועומק הלקסיקון, הוחלט שלא לקבל תשובות בשפה אחרת משפת ההיבחנות. במקרה של תשובה בשפה אחרת, הבוחן יאמר לנבחן: "תגיד לי מה המילה שרצית להגיד בערבית/בעברית" או "תענה לי בערבית/בערבית", ויציין את התשובה לפי התשובה שתינתן רק בשפת המבחן.

ג.2.ח. תגובות בסלנג

סלנג הוא שימוש לא תקני אך שכיח בשפה, הנוצר בקבוצות אזוריות או תרבותיות כלשהן, לרבות שימוש במילים חדשות שהומצאו ושאינן נחשבות תקינות, או שימוש במילים משפה אחרת. מילים מסוימות בסלנג הופכות להיות שגורות בשפה יותר ממילים אחרות, וכך הופכות להיות חלק מהשפה "הטבעית" והנפוצה. במחקרי החלוץ המצומצמים ובמחקר החלוץ התקבלו תשובות בסלנג, ועל סמך אותן תשובות נשקלה השאלה איך לציין אותן. לאחר דיון בסוגיה הוחלט שמאחר שהשימוש בסלנג הוא חלק טבעי ובעל משמעות בשפת הילדים, תתקבל תשובה בסלנג כתשובה נכונה במגבלות האלה: בתת-המבחנים 1 ו-11 – אם הנבחן משיב תשובה בסלנג, יש לתת לו את הציון 1 (ניקוד מלא) אם ברור שמשמעות תשובתו בסלנג זהה למשמעות התשובה הנכונה.

בתתי-המבחנים 11, 12, 13 ו-14 – אם הנבחן משיב תשובה בסלנג הקרובה במשמעותה לתשובה הנכונה, יש לשאול אותו שאלת בירור, ולתת לו את הציון 1 אם ברור שמשמעות תשובתו בסלנג זהה למשמעות התשובה הנכונה.

במבחני השפה בערבית תשובה תיחשב נכונה אם הנבחן ענה מילה בערבית מדוברת שיש לה אותה משמעות כמו לתשובה הנכונה הכתובה במחווה, בערבית ספרותית-רשמית.

ג.2.ט. מספר הגרסאות לפיתוח

כאמור לעיל, מדינת ישראל היא מדינה הקולטת הגירה. יש בה ילדים שהם או הוריהם היגרו לארץ מארץ אחרת ושפת האם שלהם אינה עברית. כתוצאה מכך יכולת ההבנה וההבעה שלהם בשפת המבחן נמוכה יותר. את דוברי הערבית בישראל מאפיינות שתי תופעות אחרות: השימוש בדיאלקטים שונים באזורים גיאוגרפיים ותרבותיים שונים והדיגלוסיה.

עם תחילת הפרויקט הוחלט לפתח גרסה אחת בעברית וגרסה אחת בערבית, כלומר לא לפתח גרסאות נוספות, ייחודיות לקהילות ולקבוצות תרבותיות שונות, בשל היתרון של הסטנדרטיזציה ובשל השאיפה להאחדה רבה ככל האפשר בשימוש במבחן. השיקולים נגד יצירת גרסאות שונות לקבוצות תרבותיות שונות היו גם שיקולי עלות, שיקולים הנוגעים למשאבי זמן וכוח אדם, והיה שיקול נוסף: יצירת גרסאות שונות של מבחן לקבוצות תרבותיות שונות תקשה על היכולת להשוות בין הציונים בגרסאות השונות. לכן כפי שצוין לעיל, הוחלט לפתח גרסה אחת בעברית וגרסה אחת בערבית ולשים דגש על סוגיית ההתאמה של המבחן לקהלים מגוונים מבחינה תרבותית וערכית, לבני מגזרים ותת-מגזרים שונים, מתוך רגישות תרבותית ומתוך ניסיון לצמצם עד כמה שאפשר את ההשפעות של הטיה תרבותית במבחנים (ראו להלן בסעיפים ג.2.י. ו-ג.2.יא.).

ג.2.י. רגישות תרבותית

סוגית הרגישות התרבותית בפיתוח המבחן נגעה לכל טקסט, תרשים, ציור, ייצוג גרפי ולכל הנחיה מילולית. ניתנה תשומת לב לכך שלא יופיעו במבחן פריטים העלולים לפגוע רגשית בנבחנים, להעליב אותם, להביך אותם ולהלחיץ אותם מבחינה פסיכולוגית. כמו כן ניתנה תשומת לב לכך שבנושאים אישיים וחברתיים לא יועברו מסרים שליליים (גלויים או סמויים), או כאלה שאינם מתאימים מבחינה ערכית לכל נבחן, מבחינת גיל, מגדר, אורח חיים, העדפה מינית, דת, לאום, חברה וקהילה. כדי להימנע מפגיעה בנבחנים מבחינה רגשית וערכית, וכדי להתאים את המבחן לקהלים שונים, בוצע בכל מבחן הליך שיטתי רב-שלבי שהיו מעורבים בו נציגים של תת-מגזרים תרבותיים שונים. הנציגים עברו על כל הפריטים והעירו על פריטים, ציורים והוראות אשר עלולים לפגוע ברגשות הנבחנים או כאלה שאופיינו כטעונים מבחינה תרבותית, ערכית או דתית. הפסיכולוגים בצוותי הפיתוח בחנו את ההיבטים האלה במהלך הפיתוח עצמו וגם בעת העברת הפריטים למספר מצומצם של נבחנים בשלבי הפיתוח הראשונים. במחקרי החלוץ המצומצמים ובמחקר החלוץ הועברו המבחנים לנבחנים ממגוון זרמים בעדה החרדית ולנבחנים במגזר הבדואי, ותוקנו פריטים בהתאם לממצאים שהתקבלו על בסיס איכותני-התרשמותי.

ג.2.יא. הטיה תרבותית

בפיתוח או בתרגום והתאמה של מבחנים יש להבחין בין רגישות תרבותית, ובין הטיה תרבותית. הטיה תרבותית שלילית בפריט כלשהו משמעותה היא שנבחן הניחן ביכולת הנמדדת, נכשל במענה על הפריט המודד את היכולת הזו בשל היבט תרבותי בפריט המונע ממנו לבטא את ידיעותיו או את יכולותיו. חשוב לזכור שלא כל כישלון בפריט נובע מהטיה תרבותית. לעיתים מה שנראה כהטיה תרבותית הוא חלק אינטגרלי ממה שהפריט נועד למדוד.

תשומת לב מיוחדת ניתנה לסוגיית ההטיה התרבותית במהלך כל התהליך של התאמת המבחנים לאוכלוסייה בישראל, מתוך כוונה לצמצם ככל האפשר את מספר ההטיות התרבותיות השליליות כלפי קבוצות תרבותיות, מגדריות או אזוריות. בהתמודדות עם סוגיית ההטיה התרבותית עלו שאלות כגון אלה: האם הפריט עלול להיות מוטה תרבותית? אם כן – מהי מידת ההטיה? האם הפריט מודד את התכונה שאותה התכוונו למדוד באמצעותו, מבחינה פסיכומטרית ואבחנתית? לאחר דיון בכל פריט הוחלט אם להשאיר אותו, לשנות אותו או להשמיט אותו. למשל בתת-מבחן 1א הנבחן מתבקש לשיים מילים שונות המוצגות לו באמצעות ציורים. האם בפריט של ציור "מזרקה" אפשר לצפות מנבחן בכל קבוצה תרבותית לזהות מזרקה ולדעת כיצד לשיים אותה? במילים אחרות, האם המילה מזרקה היא חלק מאוצר המילים המצופה שיהיה לנבחנים? האם יש בפריט זה הטיה תרבותית שלילית כלפי נבחנים שבתחום יישובם אין מזרקה, או שהוריהם לא לקחו אותם מחוץ ליישובם ולא הזדמן להם לראות מזרקה? או שאין בפריט זה הטיה תרבותית וזיהוי המזרקה הוא מה שרוצים למדוד במבחן. במילים אחרות, האם הנבחן נפגע בשל חסך סביבתי? או שאי זיהוי המזרקה משקף ידע מגובש נמוך והמטרה היא שהמבחן ימדוד זאת. מי קובע מהו הידע המגובש המצופה? האם הידע נקבע לפי קבוצת הרוב? לפי נתון אחר? כדי לזהות הטיה תרבותית בפריטים נעזרו המפתחים גם בנתונים אמפיריים שעלו מניתוח DIF (Differential Item Functioning) (עוד על נושא זה, ראו בפרק ד.2. העוסק במחקר

החלוץ). ניתוח זה נערך לכל פריטי המבחנים במחקר החלוץ. היו פריטים שבהם נתגלה פער גדול בתפקוד הפסיכומטרי, כלומר פער גדול יחסית ברמת הקושי של פריט כלשהו במבחן בין שתי הקבוצות שנבדקו בניתוח ה-DIF לעומת הפערים ברמות הקושי של יתר הפריטים באותו מבחן בין שתי הקבוצות. אם נמצאו פריטים שבהם היו פערים בין מגדרים שונים או בין קבוצות תרבותיות שונות, וצוות המבחן התרשם כי יש בהם הטיה מגדרית או תרבותית – הם הושמטו מהמבחנים בנוסחם הסופי. דוגמה של ניתוחי DIF של הפריטים במבחן T1, ראו בנספח 3 "נוסחים למחקר הנורמות".

בתיבה 4 שלהלן מתואר הליך ההתאמה התרבותית למגזר החרדי.

תיבה 4: התאמה תרבותית למגזר החרדי

קבוצת המגזר החרדי היא אחת הקבוצות התרבותיות בישראל. רוב אנשי המגזר דוברים עברית ויידיש ויש כאלה הדוברים רק יידיש. אנשי המגזר החרדי ייחודיים באדיקותם הדתית, מתגוררים בשכונות נפרדות ומתבדלים מבחינה חינוכית ותרבותית. תלמידי המגזר החרדי לומדים בבתי ספר נפרדים ואינם חשופים לתרבות שמחוץ למגזר שלהם. מובילי הדעת במגזר (רבנים, מחנכים, מורים ומלמדים) רגישים לכל חשיפה של תלמידיהם לחומרי לימוד או חומרי הערכה, שמא אינם בעלי מסר חינוכי המתאים לערכיהם. מסיבות אלה, אפשר היה לצפות שבמגזר זה יהיו איים שפתיים, ועולם אסוציאטיבי שונה שישפיעו על הביצועים של נבחנים מהמגזר הזה.

כיוון שתלמידי המגזר החרדי הם חלק ממערכת החינוך בישראל, בפיתוח גרסת המבחן לשפה העברית ניתנה תשומת לב מיוחדת להתאמת המבחנים והפריטים לנבחנים ממגזר זה. הוחלט לא לכלול במבחנים פריטים העלולים לפגוע בערכיהם וברגשותיהם של נבחנים מהמגזר. נציגות של שפ"י מהמגזר החרדי בדקו את כל הפריטים במבחנים שפותחו למחקרי החלוץ המצומצמים והעירו על פריטים שלדעתן לא התאימו לנבחנים מהמגזר החרדי בגלל היבט הקשור לרגישות תרבותית. בעקבות הערותיהן תוקנו פריטים מילוליים וציורים במבחנים שונים. עיקרון נוסף שהנחה את צוות הפיתוח בדיונים על ההתאמה התרבותית למגזר החרדי היה שאין לפסול פריטים רק בשל חוסר מוכרות או בשל היבטים הנוגעים ישירות ליכולת הנמדדת במבחנים. במקרים אחדים, על סמך הנתונים שהתקבלו ממחקרי החלוץ המצומצם וממחקר החלוץ, הוחלט להכניס למחווה גם דוגמות של תשובות ייחודיות שהתקבלו מנבחנים חרדים מתוך התחשבות בהבדלים התרבותיים בין המגזר הכללי ובין המגזר החרדי ועל מנת להקל על הבחון.

להלן דוגמות להתאמות שנעשו מתוך רגישות למגזר החרדי :

- הוחלפו פריטים המעוררים אי-נוחות במגזר החרדי, כמו ציור "חזיר" או המילה "דינוזאור".
- במבחנים שבהם צוירו בגרסה הישראלית ציורים ייחודיים, הובאו בחשבון רגישויות של המגזר החרדי. כך למשל צויר "קוף" שאינו "אנושי מדי" במראהו ובתנוחת גופו. במבחן שבו מוצגים

ציורים מקטגוריות שונות, הוחלט לבטל את הקטגוריה של פרצופים מאחר שלא הייתה הסכמה בנוגע למראה הציורים למגזר החרדי.

- במבחנים שבהם נדרש הנבחן לזהות ולזכור ציורי קומיקס, הוחלף השם "יצורי חלל" שבמקור ל"יצורים" בגרסה הישראלית.
- בתת-המבחן 1ג "הפכים" התקבלו בפריט "טבעי" תשובות לא צפויות מנבחני המגזר החרדי, כגון "פלאי" או "נסי". תשובות אלה המחישו את ההבנה הסמנטית הייחודית של נבחני המגזר הזה. בהקשר התרבותי והחברתי תשובות אלה יכולות להיחשב נכונות אף על פי שאינן מבטאות את מה שצוות הפיתוח התכוון אליו. למרות זאת עקב עודף פריטים הוחלט בסופו של דבר להשמיט את הפריט הזה כדי למנוע התלבטויות וחוסר סטנדרטיות בציון.
- הפריטים במבחני הידע המגובש שעסקו בידע מדעי או ביולוגי, נצפו כמוטים תרבותיים באופן מובהק לרעת הנבחנים מהמגזר החרדי, שכן נבחני המגזר הזה אינם חשופים כלל לעולם תרבותי זה. פריטים אלו הושמטו (למשל "אלמוג" ו"טפטפות" במבחן 8א "ידע כללי").
- דוגמה הפוכה לפריט שהיה מוטה באופן מובהק לטובת הנבחנים מהמגזר החרדי היה הפריט "אביון" במבחן 1ב "מילים נרדפות". מילה זו שמקורה בתנ"ך שגורה ומוכרת מאוד במגזר החרדי, ומוכרת פחות ביתר המגזרים. גם פריט זה הושמט.

בתיבה 5 שלהלן מתואר הליך ההתאמה התרבותית למגזר הבדואי.

תיבה 5: התאמה תרבותית למגזר הבדואי

האוכלוסייה הבדואית הינה קבוצה מובחנת בתוך החברה הערבית בישראל, והיא בעלת מאפיינים תרבותיים הייחודיים לה. מרבית האוכלוסייה הבדואית בישראל נמצאת בנגב וחיה בסביבה מדברית. בפיתוח המבחנים בשפה הערבית ניתנה תשומת לב מיוחדת הן למאפיינים תרבותיים הן למאפיינים השפתיים (הדיאלקט הבדואי), והפריטים המילוליים והקשורים לידע עולם נבדקו בקפדנות. פסיכולוגית בכירה, העובדת במגזר הבדואי, בדקה את הפריטים ואת המחוונים במהלך מחקרי החלוץ המצומצמים, ופעם נוספת במהלך מחקר החלוץ. כמו כן נערכו התיעצויות עם אנשי שפה דוברי הדיאלקט הבדואי בשלבי הפיתוח. בתהליך זה הושם דגש על מידת הבהירות והמוכרות של הפריטים השונים, נוסף לרגישות התרבותית. ניכר היה למשל שבפריטים הקשורים לעולם הים התגלו רמות קושי גבוהות יחסית בקרב הנבחנים מהמגזר הבדואי (למשל הפריט *مرساة* "עוגן" – מבחן 1א מחקר החלוץ). ההסבר לעליה ברמות הקושי היה שהמוכרות או החשיפה לפריטים מסוג זה הינה נמוכה באופן משמעותי בקרב הנבחנים הבדואים בשל סביבת החיים המדברית והתרבות המרוחקת מעולם הים.

להלן דוגמות להתאמות שנעשו מתוך רגישות למגזר הבדואי:

- הפריט *نافورة* מזרקה (תת-מבחן 1 א "אוצר מילים") התגלה כמוטה לרעת הנבחנים מהמגזר הבדואי. אוביקט זה כמעט ואינו נוכח בסביבת החיים הטבעית של אוכלוסייה זו. הפריט הושמט.
- הפריט *حنفية* ברז (תת-מבחן 8 א "ידע כללי") התגלה כמוטה לרעת הנבחנים מהמגזר הבדואי, ונמצא כי חלק גדול מהתשובות שהתקבלו בקבוצה זו היה *ماسورة* (צינור). זו דוגמה לכך שבדיאלקט הבדואי לאוביקט מסוים יש שם אחר, בעל משמעות שונה, ביחס לשאר הדיאלקטים של הערבית ובמקרה זה בהשוואה לשפה הסטנדרטית. הפריט הושמט.
- הפריט *شمعة* נר (תת-מבחן 1 א "אוצר מילים") התגלה כקשה והתברר שישנה דו-משמעות למילת גריין זו בדיאלקט הבדואי (*طشت* קערת פלסטיק). הפריט הושמט.
- הפריט *قارب لشراع مثل طائرة* ל... סירה למפרש זה כמו מטוס ל... (תת-מבחן 1 ד "אנלוגיות מילוליות") נמצא מוטה לרעת הנבחנים מהמגזר הבדואי, וההסבר שוב, בדומה לדוגמה *مرساة* עוגן (מובא בפסקה למעלה) הוא שפריט זה שייך לעולם הים, והוא פחות מוכר בתרבות הבדואית. הפריט הושמט.

ג.2.יב. דיגלוסיה

המונח "דיגלוסיה" מתאר מציאות שבה בתוך אותה חברה מתקיימים שני קודים לשוניים זה לצד זה: הדיאלקט המדובר היומיומי והשפה הסטנדרטית, שהיא שפת הכתיבה והקריאה. לפי הגדרתו הקלאסית של פרגוסון, זוהי "סיטואציה לשונית יציבה יחסית, שבה לצד הדיאלקטים המרכזיים של השפה יש שפה נוספת, בעלת מעמד גבוה יותר, עשירה יותר, המשרתת מאגר גדול ומכובד של ספרות כתובה, השייך לתקופה מוקדמת יותר או לקהילת דוברים מסוימת. שפה זו נלמדת ברובה במסגרת החינוך הפורמלי, היא משרתת מטרות רשמיות של כתיבה ודיבור, אך אינה נמצאת בשימוש באף קבוצה של הקהילה בשיח היומיומי" (Ferguson, 1959, p. 336).

השפה הערבית היא דוגמה קלאסית למציאות דיגלוסית, שבה "הערבית המדוברת" היא השפה הדבורה, הנמוכה, הנתפסת כפשוטה והנפוצה בתקשורת היומיומית הלא-פורמאלית. היא השפה הטבעית והאינטואיטיבית השגורה בפי כל ילד ערבי, השפה שבה הוא הוגה את מילותיו הראשונות, מבטא את צרכיו, את תחושותיו ואת רעיונותיו; לעומתה "הערבית הסטנדרטית" היא הגוון הרשמי והכתוב של השפה, שאותה כל ילד וילדה ערבים רוכשים באופן פורמלי בבית הספר, בדרך הדומה לרכישת שפה שנייה (Maamouri, 1998; Holes, 2004). בין שני גווני השפה, הדבור והכתוב, קיים מרחק לשוני רב, העלול להעמיד מכשול בפני דוברי השפה הצעירים בדרכם לרכישת האוריינות הלשונית בשפה הסטנדרטית. המרחק הלשוני הבולט בין שני הדיאלקטים מתבטא בהבדלים בתחומי השפה השונים: בפונולוגיה, במורפולוגיה, בסמנטיקה ובתחביר (Saiegh-Haddad, 2003).

כתוצאה מהדיגלוסיה המאפיינת את השפה הערבית נוצרת תופעה של "צמדי פריטים לקסיקליים": מילה אחת בשפה המדוברת והאחרת בשפה הסטנדרטית. לתופעה זו היו השלכות על הליך בחירת

הפריטים במבחנים המילוליים ובמבחני הזיכרון וכמו כן על הנחיות ההעברה והצינון במחווה. מבחינים בשלושה סוגים עיקריים של צמדי פריטים לקסיקליים ("מדוברת – סטנדרטית"): הסוג הראשון הוא "צמד מילים שונות" (Unique Words). בצמד זה המרחק הלשוני גדול ושתי המילים שונות מאוד זו מזו. כך לדוגמה המילה *Ša:f/شاف* בשפה המדוברת לעומת המילה *ra?a/رأى* בשפה התקנית (ראה). הסוג השני הוא "צמד מילים זהות" (Identical Words). בצמד מילים זה יש חפיפה מוחלטת בין השפה המדוברת ובין השפה התקנית, לדוגמה המילה *balad/بلد* (ארץ). הסוג השלישי הוא "צמד מילים דומות" (Cognate Words) והוא השכיח ביותר. בצמד זה יש קשר בין שתי המילים, אך הן עדיין רחוקות מלהיות זהות. ההבדל ביניהן מתבטא בחלק מהפונמות או במבנה ההברתי, כמו *ra:s/راس* בשפה המדוברת לעומת *ra?s/رأس* בשפה הסטנדרטית (ראש). בשל הסוג הראשון ("צמד מילים שונות"), במטלה שבה התבקש הנבחן לשיים תמונה (כגון התמונה "כובע" במבחן 21 – "שיום מהיר"), גם המילה הסטנדרטית (*قبعة*) וגם המילה המדוברת (*طقيّة/برنيطة*). נחשבו במחווה תשובות נכונות. לעומת זאת בשל הסוג השלישי ("צמד מילים דומות"), שבו המילים קשורות זו לזו אך אינן זהות, היה צריך להסביר לבוחן בכל פריט מהי המדיניות של הקראת הגריין ומהו הכלל לצינון התשובה לפי המאפיינים של כל מבחן. למשל במבחן 8 – "ידיעות כלליות", בפריט *sullam/سُلَّم* (סולם) בשפה הסטנדרטית), התבקש הבוחן לומר את הגריין בדיאלקט הדבור של השפה (*sillam/سِلَّم*) כדי למנוע מצב שבו הידע השפתי יתערב במדידה.

השלכה משמעותית אחרת של הדיגלוסיה הייתה על השמעת הפריטים במבחן 14 – "זכירת ספרות לאחר" ובמבחן 20 – "זיכרון עבודה שמיעתי". במבחנים אלה הוחלט שהבוחן יקרא בקול את הפריטים בדיאלקט הדבור של הנבחן במקום להשמיע אותם במכשיר ההשמעה (כמו בעברית ובאנגלית). כך נמנעה השפעת הדיגלוסיה על תפקודו של הנבחן במטלות הנוגעות לזיכרון. עוד השלכה של ההתחשבות במציאות הדיגלוסית בתהליך הפיתוח הייתה על הסוגיה הוראות לנבחן. הוסכם שההוראות, ההסברים והכללים ייכתבו בערבית הסטנדרטית, אך יוקראו לנבחן בערבית המדוברת, בדיאלקט המקומי שלו.

ג.2.יג. דיאלקטים

הערבית הסטנדרטית היא השפה הרשמית ב-23 מדינות במזרח התיכון ובצפון אפריקה. כאמור לעיל, בגלל תופעת הדיגלוסיה השפה הערבית המדוברת משתנה מאזור לאזור. הערבית הפלסטינית היא השם הגנרי של קבוצת הדיאלקטים הדבורים בפי האוכלוסייה הערבית בישראל ובשטחי הרשות הפלסטינית. קבוצה זו נחלקת לשלושה דיאלקטים עיקריים: הכפרי, העירוני והבדואי. הדיאלקט הכפרי נחלק לדיאלקט באזור הגליל (הצפוני) ולדיאלקט באזור המשולש. הערבית הפלסטינית, בדומה לכל דיאלקט ערבי, מתקיימת זו בצד זו עם השפה הערבית הסטנדרטית שהיא השפה הרשמית המשותפת לכל ארצות ערב. בדרך כלל אנשים הגרים באזור גאוגרפי מסוים מדברים בדיאלקט האופייני להם, השונה מזה של שכניהם מאזורים אחרים. הדיאלקטים יכולים להיות שונים זה מזה במידה רבה מאוד עד כדי כך שדוברי דיאלקט באזור מסוים יתקשו להבין במדויק דוברי דיאלקט באזור אחר. השוני בין הדיאלקטים בולט בעיקר בהגייה של פונמות מסוימות ובפריטים לקסיקליים. למשל

בנושא הפונמות, הצלילים הבין-שיניים (ث, ذ, ظ) כמעט אינם נשמעים בדיאלקטים העירוניים, אבל נשמעים בדיאלקטים הכפריים. ברמה הלקסיקלית המילה "ילד" יכולה להיכתב בכמה גרסאות, כמו عجي/aj', ولد/walad' או صبي/sabi'.

לריבוי הדיאלקטים היו השלכות על תהליך הפיתוח: על הכנת המחווים לפריטים (היה צריך לכלול תשובות בכמה שיותר דיאלקטים, במיוחד במבחנים המילוליים), על ההנחיות לבוחן (צריך לקרוא את ההנחיות בדיאלקט של הנבחן), על מבחנים ספציפיים (כמו מבחני "ידע כללי", שבהם היה צריך לקרוא בקול גם את הגריין בדיאלקט של אותו נבחן) ועל בחירת הפריטים (כמו במבחן 2 – "למידה חזותית-שמיעתית", שבו בחירת מילים המשותפות לכל הדיאלקטים היה אתגר גדול).

ג.2. יד. כללי כניסה ויציאה

בדומה למבחני משכל אחרים, גם מפתחי המבחן וודקוק-ג'ונסון האמריקני קבעו בכל מבחן נקודות כניסה (באיזה פריט יש להתחיל את המבחן) וכללי יציאה או נקודות יציאה (באיזה פריט יש לעצור את המבחן). כללי הכניסה והיציאה הם קווים מנחים שנועדו לצמצם ככל האפשר את זמן המבחן, כלומר להתאים את מספר הפריטים המועברים לנבחן כלשהו על בסיס הערכה המביאה בחשבון את גיל הנבחן ואת אומדן הטעויות הצפויות שלו במבחן. כללי הכניסה והיציאה נקבעו כדי להימנע ככל האפשר מלעורר תסכול בקרב נבחנים צעירים או בקרב נבחנים המתקשים להתמודד עם פריטים קשים מאוד, ומאידך גיסא כדי לחסוך מנבחנים מבוגרים או מנבחנים בעלי יכולת גבוהה את ההיבחות בפריטים קלים מאוד, שאינם מאתגרים אותם ושעלולים להיראות בעיניהם טפשיים/קלים מדי. מאחר שהפריטים במבחני וודקוק-ג'ונסון III מסודרים בדרך כלל לפי רמת קושי עולה, נקבע עקרון העברה אדפטיבי שלפיו נבחנים מבוגרים יותר יתחילו את המבחנים בנקודות כניסה מאוחרות יותר, כלומר לא מהפריט הראשון אלא מאחד הפריטים לאורך המבחן (למעט במבחנים הקצובים בזמן, או במבחנים שבהם נקבע שכל הנבחנים מתחילים את המבחן בפריט הראשון). בנוגע לכללי היציאה נקבע שנבחנים הנכשלים בכמה פריטים רצופים, או שלא צברו כמות נקודות מסוימת עד לפריט מסוים, בחינתם תיעצר והם יעברו למבחן הבא. בחישוב הציון הגולמי מניחים שהנבחן פתר נכון את כל הפריטים שמתחת לנקודת הכניסה שלו ושגה בכל הפריטים שמעל נקודת היציאה שלו.

נקודות הכניסה מוגדרות על פי דרגת הכיתה. נקודות היציאה מוגדרות לפי מספר הפריטים הרצופים שבהם שגה הנבחן (למשל לאחר כישלונות בשלושה פריטים רצופים) או לפי מידת (אי) הצלחתו בקבוצת פריטים או לפי מספר הנקודות שצבר הנבחן מתחילת המבחן ועד לפריט מסוים שנקבע כנקודת יציאה.

בקביעת נקודות הכניסה והיציאה לנוסחים בישראל הובאו בחשבון נקודות הכניסה והיציאה בנוסח המקור. נקודות הכניסה והיציאה נקבעו לפי הנתונים שהתקבלו במחקר החלוץ בהתאם לרמות הקושי של הפריטים בקבוצות גיל שונות.

גיבוש נקודות הכניסה: בכל מבחן רלוונטי הוחלט כמה נקודות כניסה יהיו, בהתאמה למבחן המקור במידת האפשר. נקודות ההתחלה האופטימליות לכל דרגת כיתה התקבלו על סמך הנתונים של ניתוח הפריטים במחקר החלוץ. כלל אצבע שהנחה את צוות הפיתוח בקביעת נקודות כניסה היה לבדוק בכל קבוצת גיל באיזה פריט ירדה רמת ההצלחה מתחת ל-80% או 85% מתוך הנחה כי נבחנים בעלי יכולת גבוהה יצליחו בפריטים שלפני פריט זה. בנוסף נבדק בכל קבוצת גיל באיזה פריט הייתה עלייה ברמת ההבחנה (לפי פרמטר ההבחנה point biserial), מתוך הנחה שאין טעם להעביר את הפריטים שלפני פריט זה לאותה קבוצת גיל שכן רמת ההבחנה של הפריטים האלה נמוכה (כיוון שרוב הנבחנים ענו עליהם נכון).

גיבוש נקודות היציאה: בכל המבחנים שבהם בנוסח המקור כללי היציאה היו פשוטים, למשל "הפסק אחרי שלוש טעויות רצופות" – נבדקה התקפות של כללי יציאה אלה על סמך הנתונים שהתקבלו במחקר החלוץ מחישוב אחוז ההצלחה של הנבחנים אחרי רצף הטעויות שהוגדר. הבדיקה נערכה עבור הנבחנים בשתי השפות עברית וערבית. דבר זה התאפשר מאחר שהוחלט שבמחקר החלוץ כללי היציאה יהיו גמישים יותר והנבחנים יפסיקו את המבחן אחרי רצף של שש או שבע טעויות (ולא אחרי שלוש טעויות). כך אפשר היה לעקוב אחר הצלחות הנבחנים לאחר רצפים שונים של טעויות. במבחנים שבהם יש נקודות עצירה (כמו במבחנים 2, 3, 5, 10 ו-16) חושבו מספר הנקודות שהנבחנים קיבלו עד לנקודות העצירה השונות, כפי שהוגדרו במקור, ואחרי נקודות העצירה והוגדרו כללי יציאה לאחר פריט או קבוצת פריטים שאחריהם שיעור ההצלחה היה נמוך מאוד. הרציונל היה לתת לנבחנים הזדמנות לפתור פריטים קשים יותר אם יש סיכוי שיצליחו בהם, אבל להפסיק עבורם את המבחן אם הסיכוי שיצליחו לפתור פריטים נוספים הוא נמוך מאד.

שני שיקולים נוספים שהובאו בחשבון בקביעת נקודות הכניסה והיציאה: השיקול הראשון היה קשור למיקום הפריטים בעמודים מסוימים: על מנת להקל על הבוחנים הוחלט שנקודות ההתחלה של המבחנים יהיו רק בתחילת עמוד. במבחנים שבהם הפריטים מוצגים בעמוד הפונה לנבחן, הבוחן יפסיק את העברת הפריטים רק בסוף עמוד על מנת לא לתסכל את הנבחנים שיראו שיש פריטים נוספים בעמוד שהם אינם עונים עליהם (גם אם הושג הכלל לעצירת המבחן). השיקול השני היה קשור להתאמה בנקודות הכניסה והיציאה בין הנוסח בעברית ובין הנוסח בערבית. יש לציין שלא תמיד הדבר הסתייע בגלל שיקולים הנוגעים לממצאים מניתוחי הפריטים במחקר החלוץ.

ג.2.טו. התאמת המדריך לבוחן לנוסחים בעברית ובערבית

במדריך לבוחן יש מידע מפורט על אופן השימוש במבחני וודקוק-ג'ונסון III. עבודת ההתאמה של המדריך לבוחן בשפה האנגלית לנוסחים בעברית ובערבית נעשתה בשלבים: בתחילה המדריך תורגם מאנגלית לעברית, צוות הפיתוח עבר עליו וחלקים ממנו נכתבו מחדש (תוך התייעצות עם מומחים נוספים כמו מומחים לנבחנים בעלי לקויות שמיעה ולקויות ראייה העובדים עם האוכלוסייה בישראל), התווספו חלקים שהיו חסרים וחשובים בעיני הצוות, ובסופו של התהליך המדריך עבר עריכת לשון והפקה. לאחר שנכתב המדריך בעברית הוא תורגם גם לערבית. המטרה הייתה להתאים את המדריך לבוחן למבחנים בכל שפה ולדגשים בכל שפה תוך שמירה על הדגשים בנוסח המקור.

הפרק הראשון במדריך לבוחן הוא סקירה כללית. הפרק השני עוסק בתאוריה שעליה מושתתים המבחנים וכמו כן בתיאור מורחב של כל אחד מ-23 המבחנים ו-22 המדדים שאפשר להפיק מהם. כמו כן, מתוארים הציונים שאפשר להפיק מהכלי הממוחשב. בפרק השלישי כתובים נהלים כלליים הנוגעים להעברת המבחנים ויש בו גם התייחסות להתאמות לאוכלוסיות מיוחדות. בפרק הרביעי כתובות הנחיות ספציפיות הנוגעות לאופן העברת המבחנים ולאופן הציון שלהם. בחלק מהמבחנים מובא פירוט שהוא מעבר למה שכתוב באוגדני המבחנים.

במדריך לבוחן בגרסאות בעברית ובערבית יש שינויים ושיפורים לעומת נוסח המקור. השינויים באו לידי ביטוי בעיקר בפרקים השלישי והרביעי. התווספו דגשים ודוגמות שנגעו להעברת המבחנים ולהליכי התייעוד והציון, בעברית ובערבית. למשל נכתבו כללים לציון מבחני השפה אם לנבחנים יש קשיי הגייה, נוסחו כללים לציון הקשורים לחלקי הדיבר במבחנים 1 ו-1ג והתווספו הסברים בנוגע לדיאלקטים השונים ובנוגע לדיגלוסיה בשפה הערבית.

ד. מחקרי חלוץ מצומצמים ומחקר חלוץ

ד.1. מחקרי החלוץ המצומצמים

מחקר חלוץ מצומצם ("מיני-פיילוט") הוא תהליך העברה ניסיוני של מבחן, שבו מועברים פריטי מבחן בצורתם הראשונית למספר מצומצם של נבחרים. נהוג לשלב במחקר שאלות לנבחן שעיקרן הבהרה ותחקור ("מדוע השבת ככה ולא אחרת?" "מה הבנת מההוראות?") כדי להבין את שיקולי הנבחן בעת מתן התגובה. נהוג גם להשתמש בשיטה של "חשיבה בקול רם" – *thinking aloud method*, השאלה מהמעבדה הקוגניטיבית: הנבחרים אומרים בקול רם מה הם חושבים ומה הם עושים תוך כדי פתרון שאלה, בעיה או משימה. שלב זה הוא חשוב לבדיקה ראשונית של תפקוד הפריטים ושל ההוראות הנוגעות להם. בשלב זה בודקים אם הנבחרים מבינים את המטלה כפי שהתכוונו אליה המפתחים ואם הם יכולים לפתור אותה. המסקנות העולות משלב זה מבוססות על ניתוח איכותני (ולא כמותי) של תגובות הנבחרים.

עם סיום שלבי התרגום וההתאמה הראשוניים של המבחנים החל השלב של מחקרי החלוץ המצומצמים. המבחנים שעברו הליך הרכבה או הליך התאמה עמדו במרכז המחקר עקב השינויים הרבים שבוצעו בהם לעומת נוסח המקור באנגלית ועקב המורכבות השפתית שלהם. במבחני השפה בוצעו שני מחקרי חלוץ מצומצמים, בכל פעם למקבץ מבחנים אחר. נוסף על כך בוצעו מחקרי חלוץ מצומצמים למבחנים הצורניים (מבחנים שבהם השפה אינה מרכיב מרכזי או שאין השפעה לשפה מבחינת מורכבות ההוראות לנבחן). מבחן 2 – "למידה חזותית-שמיעתית" נבדק בשני מחקרי החלוץ המצומצמים. המבחן תוקן בעקבות מחקר החלוץ המצומצם הראשון ולאחר מכן נבדק הנוסח החדש במחקר החלוץ המצומצם השני.

מבחני השפה הועברו לנבחרים על ידי פסיכולוגים בוחנים שגויסו במיוחד לכך והוכשרו על ידי צוות הפיתוח (עוד בנושא הכשרת הבוחנים, ראו בסעיף ד.1.ד. להלן). המבחנים הצורניים הועברו לנבחרים על ידי אנשי צוות הפיתוח בשתי השפות.

ד.1.א. המטרות של מחקרי החלוץ המצומצמים

מטרת-העל של מחקרי החלוץ המצומצמים הייתה הכנה וגיבוש המבחנים לקראת ההעברה הנסיונית רחבת ההיקף – מחקר החלוץ. היה חשוב לגבש את המבחנים לקראת מחקר החלוץ ולהעביר אותם כנוסחים בשלים ככל האפשר כדי לא להידרש למחקר חלוץ נוסף וכדי למנוע מנבחרים רבים להיבחן על פריטים שאינם מוצלחים.

המטרות הספציפיות של מחקרי החלוץ המצומצמים היו אלה:

- בדיקה של תהליך ההעברה בכללותו, וזיהוי בעיות בהעברתם של מבחנים ספציפיים.
- בדיקה של מורכבות ההעברה של מבחנים כמו 2, 5 ו-16. בחינה של דרכי התיעוד והציגון במבחנים האלה כדי שיהיו מיטביים וגם קלים ככל האפשר לבוחנים.

- בדיקת ההוראות לנבחנים בשכבות הגיל השונות מבחינת הבהירות שלהן וגיבוש הצעות לשיפורן. הושם דגש על בהירות ההוראות לנבחנים צעירים (קבוצות הגיל 3.5-7).
- בדיקת הפריטים בכל מבחן אם הם אכן "עובדים" ופתירים, ואם הנבחנים עונים עליהם כמו שהתכוון צוות הפיתוח. איתור פריטים בעייתיים, למשל פריטים שבהם הצירים אינם ברורים או פריטים שיש בהם כפל משמעויות. האיתור נעשה למשל על בסיס תגובות בלתי צפויות, רמת קושי בלתי צפויה וגיוון רב מדי בתשובות הנבחנים.
- בדיקת טיב המחוונים באמצעות ניתוח של תגובות הנבחנים, הצפויות והבלתי צפויות (במחוונים מוגדרות באופן שיטתי תשובות נכונות ותשובות שגויות).
- בדיקה של שאלות הבירור במבחני השפה כדי לחדד אותן, כדי להתאים את הניסוח שלהן לתגובות הנבחנים וכדי להוסיף שאלות בירור לפריטים בהם נדרשות שאלות כאלה.
- זיהוי פריטים שעולות בהן סוגיות הקשורות לרגישות תרבותית, להטיה תרבותית או להטיה מגדרית.
- זיהוי בעיות הנובעות מתשובות שונות על אותם פריטים בדיאלקטים השונים, או בעיות הנובעות מדיגלוסיה בנוסח בערבית.
- אומדן ראשוני של הזמן הנדרש להעברת כל אחד מהמבחנים.
- בדיקה כמותית של חלק קטן מהמבחנים, שהיו בהם גם מדדים כמותיים (למשל מבחן 2 - "התאמה חזותית שמיעתית" ומבחן 10 - "התאמה חזותית שמיעתית - מושהית").

ד.1.ב. אוכלוסיית הדגימה ומאפייני המדגם

במחקר החלוץ המצומצם הראשון הועברו המבחנים 1א, 1ב, 1ג, 1ד, 2-7 ומבחן 20 ל-107 נבחנים בעברית וכמו כן ל-111 נבחנים בערבית. נוסף על כך הועבר המבחן "התאמה חזותית-שמיעתית מושהית" (מבחן 10 בסוללת המקור, שלא נכלל בסופו של דבר בגרסה הישראלית). **במחקר החלוץ המצומצם השני** הועברו המבחנים 2, 8א, 8ב, 9, 11, 14, 15 ו-17 ל-80 נבחנים בעברית וכמו כן ל-82 נבחנים בערבית. נוסף על כך הועבר המבחן "זכירת משפטים" (מבחן 27 בסוללת המקור, שלא נכלל בסופו של דבר בגרסה הישראלית).

המבחנים הועברו ל-7 קבוצות הגיל האלה: גן חובה, כיתות א', ג', ה', ז', ט', יא'. בכל קבוצת גיל השתתפו 10 עד 15 תלמידים רגילים מן החינוך הכללי. כמו כן הועברו המבחנים למספר מצומצם של נבחנים מהמגזר החרדי ומהמגזר הבדואי.

במחקר החלוץ המצומצם השלישי נבדקו כל יתר המבחנים שהיו מבחנים צורניים. המבחנים הועברו לכ-20 נבחנים בקבוצות גיל שונות, בעברית ובערבית. הושם דגש על בדיקה של התאמת ההוראות לנבחנים צעירים.

ד.1.ג. הכנת חוברת תיעוד ייעודית לביצוע מחקרי החלוץ המצומצמים

- לקראת מחקרי החלוץ המצומצמים הוכנה חוברת ייעודית לתיעוד ולצינון של תשובות הנבחים.
- במבחני השפה נדרשו הבוחנים לכתוב בחוברת התיעוד את כל התגובות המילוליות של הנבחים, לכל הפריטים, וגם את הצינון שלהם. המטרה הייתה להפיק מידע איכותי שיסייע לעדכן את המחוונים ואת שאלות הבירור, ויאפשר לבדוק דפוסים חוזרים של טעויות.
- הבוחנים התבקשו לציין בעמודות ייעודיות תגובות חריגות של נבחים או התנהלות שחרגה מההנחיות (למשל זמן תגובה לפריט שהוא קצר יחסית או ארוך יחסית; תגובה חריגה לפריט מבחינה רגשית, מבחינה תרבותית או מבחינה מגדרית; צורך במתן הכוונה נוספת או בהסבר נוסף מעבר למה שצוין בהנחיות).
- נוסף על חוברות התיעוד כתבו הבוחנים בסיום ההעברות של המבחן דוחות מובנים על סמך סדרת שאלות שהוכנה מראש. הדוחות סיכמו את המידע האיכותי בנוגע לפריטים, להנחיות ולבעיות מיוחדות שעלו במהלך ההעברות.
- המסקנות שהוסקו מהדוחות רוכזו במסמך מפורט ונדונו בישיבות של צוותי הפיתוח וועדת ההיגוי המלווה שעסקו בגיבוש הפריטים והמבחנים לקראת מחקר החלוץ.

ד.1.ד. הבוחנים וסדנאות ההכשרה

- כדי להעביר את מחקרי החלוץ המצומצמים גויסו כ-30 פסיכולוגים וסטודנטים לתואר שני מ-22 שפ"חים (תחנות שירות פסיכולוגי חינוכי) ברחבי הארץ. גויסו כ-15 בוחנים מכל מגזר שפה (במגזר דוברי הערבית כל הפסיכולוגים היו פסיכולוגים מומחים).
- כל בוחן נדרש להעביר 8 מבחנים במהלך תקופה של כחודש וחצי.
- במחקרי החלוץ המצומצמים הראשון והשני, הבוחנים השתתפו בכל מחקר בסדנה שנמשכה יום אחד (שמונה שעות הדרכה) והועברה על ידי החברים בצוותי הפיתוח. ההדרכה התמקדה בכל המבחנים שהיו כלולים במחקר, וניתנו הסברים הנוגעים לאופן העברת המבחנים, לאופן הצינון שלהם, לנקודות הכניסה והעצירה ולשאלות הבירור. בהדרכה ניתנו הסברים גם לגבי חוברת התיעוד ודפי העבודה לנבחן (במבחנים הרלוונטיים) ונעשו סימולציות של העברה, של תיעוד ושל צינון.
- רוב הפסיכולוגים הבוחנים במחקרי החלוץ המצומצמים היו מיומנים בהעברה של מבחנים פסיכולוגיים, ולכן דגש מיוחד הושם בסדנה על ההבדל שבין העברת מבחן לצורך אבחון ובדיקת הנבחן (ניסיון לעמוד על התפקוד הקוגניטיבי שלו ועל מצבו הרגשי) ובין העברת מבחן לצורך בדיקת איכות המבחן ופריטיו (מוקד תשומת הלב אינו הנבחן אלא המבחן).
- במבחני השפה הושם דגש מיוחד על הדרישות לתיעוד של תשובות הנבחים שיאפשר הפקת מידע איכותי על הפריטים. הוסבר עד כמה חשוב להקפיד על תיעוד מלא ומפורט של התשובות ושל התגובות של הנבחים לפריטים במבחנים השונים, תיעוד שיהווה בסיס להבנת הליקויים בפריטים ולקביעת התיקונים הנדרשים.

ד.1.ה. מועדי המחקרים

- מחקר החלוץ המצומצם הראשון נערך בדצמבר 2013.
- מחקר החלוץ המצומצם השני נערך במאי 2014.
- מחקר החלוץ המצומצם השלישי נערך בנובמבר 2014.

ד.1.ו. העברת המבחנים

- המבחנים הועברו בגנים ובבתי הספר בשעות הלימודים או מיד בתום הלימודים. זמן ההעברה תואם ישירות עם מוסדות החינוך. ההעברה הייתה פרטנית ('אחד על אחד') והתקיימה בחדר שקט המתאים להעברת מבחנים.
- העברת המבחנים נמשכה כשעה וחצי עד שעתיים (תלוי בגיל הנבחן וביכולת שלו).

ד.1.ז. שיפור המבחנים ועדכון בעקבות הממצאים של מחקרי החלוץ המצומצמים

והיערכות למבחן החלוץ

על בסיס התיעוד של תגובות הנבחנים ושל דיווח הבוחנים, הוצעו שינויים בפריטי המבחנים ובהנחיות, והתגבשו הנושאים שעליהם יש לתת את הדעת בהמשך התהליך (למשל במחקר החלוץ). ההצעות נידונו בצוותי הפיתוח והוחלט על השינויים הנדרשים, כגון השמטת פריטים ותיקון פריטים. בישיבות הצוותים גובשו הנוסחים למחקר החלוץ. בנוסחים שהוכנו למחקר החלוץ נכללו פריטים עודפים כדי שיהיה אפשר לבחור את הפריטים הטובים ביותר לנוסחים הסופיים על סמך הממצאים של מחקר החלוץ. למידע הנוגע למספר הפריטים בנוסחים של מחקרי החלוץ המצומצמים לעומת מספר הפריטים בנוסחים בשלבים הבאים, ולמידע הנוגע לשינויים בעקבות המחקרים – ראו טבלה 6 סעיף 4ד.

על בסיס הממצאים של מחקרי החלוץ המצומצמים נעשו השינויים האלה:

- שופרו וחודדו ההנחיות לבוחן וההוראות לנבחן בכל המבחנים בהם זה נדרש. בכמה מבחנים נוסחו מחדש ההוראות לנבחן מאחר שנבחנים רבים התקשו בהבנתן. המסקנה העיקרית ממחקרי החלוץ המצומצמים הייתה שצריך להשתמש בשפה פשוטה, יומיומית וברורה לנבחנים צעירים, ולהעדיף את השיקול הזה על פני שיקולים הנוגעים לתקינות הלשון ולבחירת מילים במשלב גבוה.
- במבחנים הצורניים שופרו ההנחיות לבוחן וההוראות לנבחן, תוקנו ציורים והוחלפו פריטים בשל הטיות תרבותיות.
- שופרו והועשרו המחוונים על סמך תשובות שנתנו הנבחנים (שלא היו כתובות מראש במחוונים).
- במבחנים הרלוונטיים סודרו הפריטים מהקל אל הקשה, והושם דגש על מספר מספיק של פריטים קלים לבדיקת נבחנים צעירים. סידור הפריטים מהקל אל הקשה התבסס על נתונים כמותיים ועל הערכת הבוחנים את רמת הקושי של הפריטים.
- הוחלט להוסיף פריטי הדגמה בנוסף לאלה הקיימים בנוסח המקור בחלק ממבחני 1 ובמבחני 8.
- ההחלטה נבעה מהצורך לחדד את מטלת המבחן לנבחנים צעירים (למשל בתת-מבחן 1ב –

"מילים נרדפות" נוסף ציור להמחשת המשימה; בתת-מבחן 8ב – "ידע כללי" נוספה הדגמה לאופן השימוש באובייקט).

- חודדו ההנחיות לבוחן בנוגע לקשיי הגייה במבחנים 1 ו-8 ובנוגע לחלקי הדיבר במבחני 1.
- הוחלט לערוך מחקר כמותי במחקר החלוץ בנוגע למענה בשפה אחרת במבחנים 1 ו-8. ההנחה הייתה שבעזרת נתוני המחקר הזה יהיה אפשר להחליט אם לאפשר מענה בשפה אחרת באותם במבחנים.
- במבחן 2 – "למידה חזותית שמיעתית" בוצע ניתוח של רמות הקושי לפי קבוצות גיל כדי לבדוק את הטענה שהעלו הבוחנים בנוגע לקלות המבחן בשתי השפות. השערה זו נמצאה נכונה במחקרי החלוץ המצומצמים, שכן במחקרים אלו נמצא כי הציון הממוצע של נבחני חטיבת הביניים היה דומה לציון הממוצע של נבחני התיכון ("אפקט תקרה"). ממצא דומה עלה גם מעיון בטבלאות הנורמות של מבחן המקור באנגלית⁸. למרות השאיפה לשמור על דמיון בין הנוסחים בעברית ובערבית ובין נוסח המקור, הוחלט להקשות את המבחן בשני הנוסחים, בעברית ובערבית, כדי ליצור בגרסאות הישראליות הבחנה טובה יותר בין נבחנים בקצה הגבוה של היכולת. לכן, פותח פרק (סיפור) נוסף שכלל ארבעה סמלים חדשים ושלושה משפטים חדשים. פרק זה התווסף לשבעת הפרקים שהיו בנוסח המקור.
- במבחנים המושמעים חודדה ההנחיה לבוחן להעביר את המבחנים באמצעות מכשיר השמעה בלבד כדי לשמור על סטנדרטיות בהעברה. בחלק מהמבחנים נעשו שינויים בפרטי ההדגמה כדי להקל על הנבחנים לעבור ממצב שבו הבוחן קורא בקול את הפריט למצב שבו הפריט מושמע באמצעות מכשיר ההשמעה.
- במקרים רבים במבחנים המושמעים ענו נבחנים בלי להמתין לסוף הפריט המושמע. לכן חודדה ההוראה לנבחן להמתין עד סוף השמעת הפריט ורק אז לענות עליו.
- תוקנו פריטים במבחנים המושמעים בשל הקושי של הנבחנים להבין את המילים. כל הפריטים המושמעים בכל המבחנים המושמעים הוקלטו מחדש בשתי השפות.
- במחקרי החלוץ המצומצמים נבדקו כמה חלופות לקטגוריות המקוריות במבחן 9 (בעלי חיים, מאכלים ושמות פרטיים). מבחן זה בודק "שטף סמנטי", והנבחן צריך לומר בקול כמה שיותר פריטים מקטגוריה נתונה בזמן נתון, למשל שמות של מאכלים. התחושה של צוותי הפיתוח הייתה כי חלק מהקטגוריות במבחן המקור הן "פתוחות" מדי. למשל במגזר של דוברי העברית כל דבר כמעט יכול להשתייך לקטגוריה "שמות פרטיים" (שמות מהתנ"ך, שמות של חיות, שמות של צמחים וגם שמות של דוממים, כמו "אבן" ו"גפן"). מחקרי החלוץ המצומצמים איששו את החשש הזה, שכן התברר כי המבחן אינו מבחין בין הנבחנים. נוסו כמה חלופות ונבחרה הקטגוריה "בעלי

⁸ הן במבחן המקור והן במבחן שנוסה במחקרי החלוץ המצומצמים נרשמה התייצבות ברמת היכולת הנמדדת בגילי חטיבת הביניים והתיכון. עם זאת קשה לדעת האם דמיון זה נובע בהכרח מדמיון ברמת הקושי של כל אחד מ"המשפטים" וה"סיפורים" שנוסו בעברית ובערבית, מכיוון שלא היו בידנו נתונים אודות רמת הקושי של הפריטים (משפטים וסיפורים) במבחן המקור באנגלית.

מקצוע" במקום הקטגוריה "שמות פרטיים". כמו כן שונתה הקטגוריה "מאכלים" לקטגוריה מצומצמת יותר של "פירות וירקות".

- הוגדרו העקרונות שלפיהם יש לציין את פריטי מבחן 9 – "שטף סמנטי" ונכתבו מחדש ההנחיות לצינון המבחן.
- במבחן 9 התברר כי הקטגוריה "בעלי מקצוע" הייתה קשה לנבחנים צעירים. נצפה "אפקט רצפה" מסוים, שכן נבחנים צעירים התקשו לשלוף שמות של בעלי מקצוע. לאחר דיון הוחלט לבחון את מידת ההבחנה של הקטגוריה במחקר החלוץ.
- במבחן 17 – "קשב שמיעתי", תוקנו ציורים וגם הוחלפו ציורים בשל קשיי נבחנים לזהות אובייקטים המיוצגים בחלק מהציורים. לחלק מהציורים התווספו צבעים כדי להקל על הנבחן לזהות את האובייקטים.
- לאור הניסיון שהצטבר במחקרי החלוץ המצומצמים, נקבעו דגשים להליך ההכשרה הנדרש לבוחנים במחקר החלוץ. למשל הוחלט להכין מצגות ייעודיות למבחנים שבהם נדרשת מיומנות גבוהה בהעברת המבחן, בתיעוד שלו ובצינון שלו.

2.2. מחקר חלוץ (פיילוט)

מחקר חלוץ הוא מחקר רחב היקף והוא מועבר באופן ניסיוני לכמה מאות נבחנים המייצגים את אוכלוסיית המטרה. מחקר חלוץ הוא שלב חשוב בפיתוח של מבחן, והוא נועד לבדוק את תפקוד המבחן בכללותו ואת התפקוד הפסיכומטרי של הפריטים הנכללים בו. המטרות העיקריות שלו הן לבחור את הפריטים שייכללו בגרסאות הסופיות של המבחן, לתקן אותם במידת הצורך, להרכיב את המבחן הסופי ולקבוע את סדר הפריטים בו. בנוסף לכך, בעזרת מחקר חלוץ אפשר ללמוד על ההיבטים הלוגיסטיים הכרוכים בהעברת המבחן בהיקף רחב, כל זאת עוד לפני השלב של מחקר הנורמות.

2.2.א. המטרות של מחקר החלוץ

המטרות העיקריות של מחקר החלוץ בפרויקט זה היו לבדוק ולהפיק מידע אמפירי ופסיכומטרי עשיר על הפריטים ועל המבחנים כדי לעמוד על תפקודם וכן ללמוד על סוגיות הנוגעות לאופן העברת המבחנים. המידע הפסיכומטרי שימש לבדיקת איכות הפריטים בהיבטים שונים, לשיפור המחוונים ולקביעת נקודות הכניסה והיציאה במבחנים. בניתוחי הפריטים חושבו מדדי הקושי וההבחנה בתוך קבוצות הגיל ובין קבוצות הגיל, וכמו כן מדדי התפקוד הדיפרנציאלי (ראו הסבר על כך בתיבה 6).

המטרות הספציפיות של מחקר החלוץ היו אלה:

- איתור פריטים פגומים מבחינת התוכן או בהיבט גרפי (למשל פגימה גרפית נמצאה בפריט במבחן "עיבוד חזותי").
- איתור פריטים מוטעים מבחינה תרבותית או מגדרית ופריטים היכולים לעורר רגישות תרבותית.

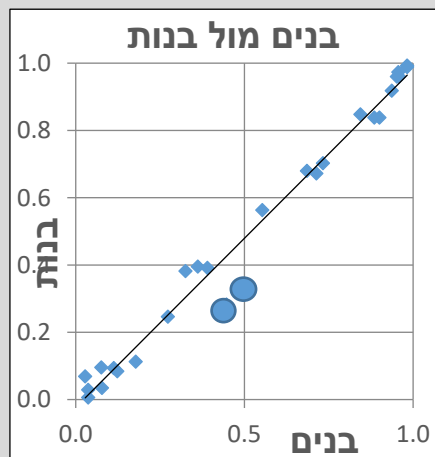
- במבחני השפה והקשב השמיעתי – הוספת תשובות אפשריות של נבחנים כדי לשפר את המחוונים ואת הניסוחים של שאלות הבירור; בדיקת התשובות בדיאלקטים השונים (בשפה הערבית) ובדיקת התשובות הייחודיות (בשפה העברית, בעיקר במגזר החרדי).
- בדיקת הבהירות של ההנחיות לבוחן ושל ההוראות לנבחן, לרבות בדיקת הבהירות של פריטי ההדגמה (בדגש על נבחנים צעירים).
- קביעת מספר הפריטים הסופי בכל מבחן וקביעת נקודות הכניסה והיציאה בכל אחד מהם.
- סידור הפריטים בסדר עולה, מהקל לקשה, במבחנים שבהם הסדר אינו מובנה.

המטרות שלא נכללו במחקר החלוץ היו אלה:

- מאחר שמחקר החלוץ לא נועד לתקף את המודל התאורטי, הוחלט שלא לנסות לשחזר או לאשש את המודל התאורטי של מבחני המקור מתוך הנתונים של מחקר החלוץ (בדיקה כזו התבצעה במחקר הנורמות).
- מאחר שמטרת המחקר לא הייתה בניית נורמות לכל גיל בנפרד, הוחלט שלא לפתח אומדנים נפרדים לכל גיל, אלא לחשב את הנתונים לפי קבוצות של גילים (רק בשלב הנורמות ייקבעו הנורמות לכל גיל בנפרד).
- באשר לבדיקת התפתחות רמות הקושי לפי קבוצות גיל, הניחו כי בדיקת מגמות בלבד בנושא זה, בקבוצות גיל "גסות", תהיה טובה דיה.
- באשר לבניית המדגם, הוחלט על מדגם שייצג באופן טוב דיו את רוב אוכלוסיית המטרה אך לא יכלול אוכלוסיות קצה או אוכלוסיות מיוחדות (כגון נבחנים מהחינוך המיוחד).

תיבה 6: ניתוח DIF

1. תיפקוד פריט דיפרנציאלי (DIF) Differential Item Functioning הוא מאפיין סטטיסטי-פסיכומטרי של פריט מבחן שמצביע על המידה שבה הפריט עלול למדוד יכולות שונות בקרב נבחנים ממגזרי אוכלוסייה שונים (למשל בנים ובנות). בהליך זה משווים בדרך כלל את הציון הממוצע בפריט של נבחנים משני מגזרים שונים שמניחים שהם בעלי יכולת זהה (במבחן המסוים). אם הציון הממוצע בפריט שונה בין שתי הקבוצות, הדבר מרמז על אפשרות שהפריט מוטה לרעת או לטובת אחת משתי קבוצות הנבחנים המשווים.
2. ישנן מספר דרכים לבחינת השאלה אם פריט מסוים במבחן נתון מתפקד באופן דיפרנציאלי או לא. השיטות המקובלות הן רגרסיה לוגיסטית, שיטות מבוססות על תאוריית ה-IRT ושיטת מנטל-הנזל (Mantel-Haenszel). שיטה פופולרית נוספת היא שיטה גרפית שבה משרטטים בגרף דו-ממדי את ממוצעי הציון בשתי קבוצות הנבחנים המשווים כאשר כל נקודה בגרף מייצגת פריט. ערך ציר ה- X מייצג את הציון הממוצע בפריט בקבוצה אחת וערך ציר ה- Y את הציון הממוצע באותו פריט בקבוצה השנייה. אם כל הפריטים במבחן נתון מודדים בערך אותה תכונה ובאופן דומה בשתי הקבוצות, צפוי להתקבל גרף שבו כל הנקודות מסודרות על קו אחד. אם פריט חורג מקו זה, משמע כי באופן יחסי ליכולת הקבוצתית, קבוצה אחת הצליחה (או לא הצליחה) במיוחד באותו פריט ויש לבדוק אותו.
3. לשם הדגמה מובא כאן שרטוט של ניתוח DIF גרפי של מבחן ידע כללי (בעברית) בין בנים לבנות. אפשר לראות בגרף כי רוב הפריטים ממוקמים על קו אחד למעט שני הפריטים המסומנים בעיגול. בפריטים אלו, בנים הצליחו באופן יחסי יותר מבנות ויש לבדוק אם תוכן הפריטים הקל על הבנים (למשל עסק בעולם תוכן מוכר יותר לבנים) או אולי הקשה על הבנות (למשל בשל תוכן פוגעני או מרתיע שכלול בו). יש לציין כי אין הכרח שהקו עצמו שסביבו "מסתדרות" הנקודות יהיה קו האלכסון שעובר בין הנקודות 0,0 ו-1,1, כי לעיתים אחת הקבוצות יכולה להיות באופן כללי טובה יותר מהקבוצה האחרת ולכן הקו כולו יתמקם מעל או מתחת לאלכסון הראשי. לפי כך, בכל המקרים, הניתוח הגרפי מתייחס לנקודות החורגות מהקו שסביבו ממוקמות מרבית הנקודות.



ד.2.ב. תכנון מסגרת הדגימה והמדגם

מסגרת הדגימה של מחקר החלוץ כללה את כל התלמידים הלומדים במערכת החינוך הישראלית הרשמית, בחינוך הרגיל, במוסדות הגנים (טרום חובה וחובה) ובבתי הספר (כיתות א' עד יא'), מגיל 3.5 ועד גיל 18. המסגרת כללה את כל סוגי הפיקוח. תלמידים עולים חדשים (הנמצאים פחות מארבע שנים בארץ) לא נכללו במדגם. במגזר החרדי, לא נכללו תלמידי ישיבות. בתכנון המוקדם של המדגם נקבע שבמדגם יכללו כ-1000 תלמידים מהמגזר דוברי העברית – בתי ספר בפיקוח ממלכתי, ממלכתי-דתי וחרדי וכ-1000 תלמידים מהמגזר דוברי הערבית – ערבים, דרוזים ובדואים (עוד על המגזר הבדואי, ראו תיבה 5 לעיל).

ד.2.ג. המדגם בפועל

בפועל נדגמו 219 מוסדות חינוך ממגזר דוברי העברית ו-206 מוסדות חינוך ממגזר דוברי הערבית, בהם נדגמו 1018 נבחנים במגזר דוברי העברית ו-894 נבחנים במגזר דוברי הערבית⁹.

- המדגם במחקר היה מדגם שכבות דו-שלבי. בשלב הראשון נדגמו מוסדות חינוך (בתי ספר וגנים), ובכל מוסד נדגם מספר מצומצם של תלמידים.
- נדגמו מוסדות חינוך (גנים ובתי ספר) בכל אחד ממגזרי השפה לפי שכבות גיל ולפי קבוצות האוכלוסייה הרלוונטיות (למשל רמה חברתית כלכלית, סוג פיקוח, פיזור גאוגרפי).
- במחקר החלוץ הוצאו מהמדגם, מטעמי נוחות, יישובים מרוחקים ובתי ספר שהנגישות אליהם נמוכה. המקומות שהוצאו מהמדגם היו אלה: עיריות ומועצות מקומיות – מטולה, קריית שמונה, קצרין, צפת, נהריה, ראש פינה, יסוד המעלה, שלומי, מעלות-תרשיחא, כפר ורדים, חצור הגלילית, מצפה רמון, אילת; מועצות אזוריות – מבואות החרמון, הגליל העליון, מרום הגליל, מטה אשר, מעלה יוסף, חבל אילות והערבה התיכונה. כמו כן הוצאו מהמדגם ערביי ירושלים והיישוב דריג'את, וכל היישובים ביהודה ושומרון, למעט הערים אריאל, מעלה אדומים, מודיעין עילית וביתר עילית.
- בכל בית ספר שנכלל במדגם נדגמו 4-8 נבחנים משכבות גיל שונות. בגני הילדים נדגמו 4 נבחנים מכל גן.
- בין דוברי העברית נדגמו 135 נבחנים במגזר החרדי בקבוצות גיל שונות. בדגימת האוכלוסייה דוברת הערבית הובא בחשבון מגוון של דיאלקטים מאזורים שונים בארץ (עוד על הדיאלקטים ראה בסעיף ג.2.ג).
- תלמידי החינוך המיוחד הלומדים במסגרות של החינוך המיוחד לא נכללו בדגימה של מחקר החלוץ. כמו כן לא נכללו בדגימה תלמידי השילוב – תלמידי החינוך המיוחד הלומדים בבתי ספר רגילים, בכיתות של חינוך מיוחד או בכיתות רגילות.
- לכל נבחן במדגם נדגמו שני נבחנים מחליפים למקרה שהוא לא יוכל להיבחן מסיבה כלשהי.

⁹ לתלמידי המגזר דובר העברית הועברה הגרסה בעברית ולתלמידי המגזר דובר הערבית הועברה הגרסה בערבית בלבד.

- בדגימה לא הובאה בחשבון רמת ההישגים של התלמיד במבחני מיצ"ב או בבית הספר. בשתי הטבלאות שלהלן מוצגת התפלגות הנבחנים במדגם במחקר החלוץ, בכל אחת משתי השפות, לפי משתני הרקע: קבוצות גיל, מגדר, סוג פיקוח (לדוברי עברית) ומגזר (לדוברי ערבית).

טבלה 3: התפלגות הנבחנים במדגם של מחקר החלוץ לפי משתני הרקע- דוברי עברית

סה"כ	קבוצות הגיל					פיקוח מגדר	
	החטיבה העליונה	חטיבת הביניים	ביה"ס היסודי כיתות ד'-ו'	ביה"ס היסודי כיתות א'-ג'	טרומ- חובה וגן חובה		
60	3	16	19	21	1	בנים	חרדי
75	12	12	17	15	19	בנות	
135	15	28	36	36	20	סה"כ	
420	63	99	84	101	73	בנים	ממלכתי וממלכתי-דתי
463	84	145	71	85	78	בנות	
883	147	244	155	186	151	סה"כ	
480	66	115	103	122	74	בנים	סה"כ
538	96	157	88	100	97	בנות	
1018	162	272	191	222	171	סה"כ	

טבלה 4: התפלגות הנבחנים במדגם של מחקר החלוץ לפי משתני הרקע- דוברי ערבית

סה"כ	קבוצות הגיל					אזור/מגדר מגדר	
	החטיבה העליונה	חטיבת הביניים	ביה"ס היסודי כיתות ד'-ו'	ביה"ס היסודי כיתות א'-ג'	טרומ- חובה וגן חובה		
76	21	23	8	11	13	בנים	בדואי - דרום
66	11	15	11	15	14	בנות	
142	32	38	19	26	27	סה"כ	
384	51	90	81	89	73	בנים	ערבי (צפון + המשולש הדרומי והצפוני)
368	49		72	86	71	בנות	
752	100	180	153	175	144	סה"כ	
460	72	113	89	100	86	בנים	סה"כ
434	60	105	83	101	85	בנות	
894	132	218	172	201	171	סה"כ	

החלפת נבחנים

תלמיד שנדגם למחקר ולא התקבל אישור מהוריו לכך שהוא יכול להשתתף במחקר, הוחלף בתלמיד שהוגדר המחליף הראשון שלו. אם המחליף הראשון לא יכול היה להשתתף מסיבה כלשהי, המחליף השני מילא את מקומו. היו מקרים בודדים שבהם תלמידים שנדגמו לא נכללו במחקר מסיבות אחרות,

למשל הם סרבו להשתתף. שיעור החלפת הנדגמים עמד על כ-19% במגזר היהודי ועל כ-22% במגזר הערבי.

ד.2.ד. העברת חלקית של סוללת המבחנים לחלק מן המדגם

בגלל מורכבות התהליך של העברת המבחנים למספר גדול של נבחנים, ובגלל הצורך לייעל את השלב של איסוף הנתונים ולחסוך בזמן ההעברה, הוחלט שלא להעביר את כל המבחנים לכל הנבחנים שעלו במדגם. במקום זאת הוחלט להסתפק במדגם קטן יותר במבחנים מסוימים, למשל במבחנים הצורניים. ההחלטה באילו מבחנים נדרש מדגם גדול ובאילו מבחנים אפשר להסתפק בקבוצה חלקית מתוך המדגם המלא, התקבלה על סמך הידע שהצטבר בהליך התרגום וההתאמה וממחקרי החלוץ המצומצמים בנוגע למורכבות המבחנים ובנוגע לצורך לבדוק מבחינה יסודית מבחנים מסוימים. בעיקרון, מבחנים שבהם היה תהליך הרכבה (פיתוח) הועברו למדגם גדול, בפרט מבחנים שבהם השפה היא גורם מרכזי – הן כמוקד המדידה (למשל מבחני 1), הן כגורם מתווך או דומיננטי. גם מבחן 5 הועבר למדגם גדול, שכן יש בו הוראות מילוליות ואופני תגובה מורכבים. מבחנים אלה ייקראו להלן "מבחנים טעוני שפה". המבחנים הצורניים והמבחנים שאומצו כמעט באופן מלא מהמקור הועברו בהיקף קטן יותר, שכן הם דומים מאוד לגרסת המקור. כמו כן הניחו כי במבחנים אלה, בשתי השפות, לא יעלו ממצאים שונים באופן מובהק מהממצאים של מבחן המקור (מבחינה איכותנית ומבחינה פסיכומטרית); לכן יהיה אפשר לצרף את נתוני המבחנים בשני המגזרים, ולקבל מספר גדול דיו של נתונים כדי לבצע עיבודים פסיכומטריים מספקים.

כדי ליישם את ההחלטה הזאת נבנתה לקראת מחקר החלוץ סוללה חלקית של מבחנים לצד סוללה מלאה. בסוללה החלקית נכללו רק 8 מבחנים טעוני שפה והיא הועברה לכל הנבחנים. שאר המבחנים מהסוללה המלאה (מבחנים שאינם טעוני שפה) הועברו לכ-500 נבחנים בכל מגזר שפה.

ד.2.ה. מועדי המחקר

- בשל אילוצים אובייקטיביים (החופש הגדול היה באמצע מחקר החלוץ) בוצע המחקר בשתי פעימות:
- הפעימה הראשונה: אפריל 2015 – יוני 2015 – בפעימה זו נבחנו כ-250 נבחנים בכל שפה.
 - הפעימה השנייה: נובמבר 2015 – מרץ 2016 – בפעימה זו נבחנו שאר הנבחנים שנדגמו בכל שפה.

ד.2.ו. הכנת חוברת תיעוד ייעודית למחקר החלוץ

לקראת מחקר החלוץ הוכנה חוברת ייעודית לתיעוד ולציינון תשובות הנבחנים. הבוחנים התבקשו לתעד בחוברת את כל התגובות המילוליות הגולמיות של הנבחן במבחני השפה וגם את הציינון שלהן. המטרה הייתה להפיק מידע איכותני שיסייע לעדכן את המחוונים ואת שאלות הביורור ולעמוד על דפוסים חוזרים של טעויות. מידע זה הוקלד במערכת הקלדה ממוחשבת, והצוות נעזר בו בשלב של גיבוש הנוסחים הסופיים. בעזרת חוברת התיעוד נערכו גם תהליכי בקרה על תהליכי ההעברה, הציינון והתיעוד שנעשו על ידי הבוחנים.

ד.2.ז. הכשרת הבוחנים וסדנאות ההכשרה

הבוחנים במחקר החלוץ היו פסיכולוגים חינוכיים מומחים, פסיכולוגים בשלבי התמחות, סטודנטים לתואר שני בפסיכולוגיה באוניברסיטה או במכללה בישראל, ובמקרים מסוימים – סטודנטים מצטיינים לפסיכולוגיה שהיו באותה עת בסוף לימודיהם לתואר ראשון. לעיתים נבחרו גם בוחנים מתחום העבודה הסוציאלית וממקצועות טיפוליים אחרים. המיון הראשוני נעשה על סמך קורות החיים ועל סמך ראיון עם ראש צוות הפיתוח או עם פסיכומטריקאית בכירה.

בראיון שהיה חלק מהליך גיוס הבוחנים הוערכו הנושאים האלה:

- באיזו מידה יש למועמד ניסיון בהעברת אבחונים או מבחנים דומים?
- האם המועמד למד/לומד קורס בתורת המבחנים או בפסיכומטריקה?
- האם למועמד יש ניסיון בעבודה טיפולית או תקשורתית? אם כן – באילו קבוצות גיל?
- באיזו מידה המועמד מגלה רגישות לסביבה וגישה חיובית ונעימה לזולת?
- האם למועמד יש מוטיבציה להשתתף בפרויקט ולהתמיד בו עד סיומו?
- האם המועמד מצטייר כאדם מסודר, מאורגן ושיטתי, המסוגל להעביר את המבחנים לפי ההנחיות ולתעד את העברתם כפי שנדרש?

סדנת ההכשרה

כל הבוחנים השתתפו בסדנה של יומיים (16 שעות הדרכה), שהועברה על ידי אנשים מצוותי הפיתוח. ההדרכה התמקדה בכל המבחנים שהיו כלולים במחקר, וניתנו הסברים הנוגעים לאופן העברת המבחנים, לאופן הצייון שלהם, לנקודות הכניסה והעצירה ולשאלות הבירור. בהדרכה ניתנו הסברים גם לגבי חוברת התיעוד ודפי העבודה לנבחן (במבחנים הרלוונטיים) ונעשו סימולציות של העברה, של תיעוד ושל צייון. ניתנה תשומת לב מיוחדת למבחנים שזוהו כמורכבים להעברה או לצייון במחקרי החלוץ המצומצמים.

בוחנים שלא היו פסיכולוגים או שהיו פסיכולוגים בשלבי ההתמחות שלהם, השתתפו בהרצאה ייעודית בנושא התקשורת עם הנבחן, שניתנה על ידי פסיכולוג חינוכי בכיר מצוות הפיתוח וראש צוות הפיתוח. בהרצאה נדונו דרכים ליצירת קשר חיובי עם הנבחנים לפני תחילת העברת המבחנים ודרכים להתמודדות עם קשיים בזמן העברת המבחנים (קשיים רגשיים, קשיי ריכוז, חוסר עניין). עם סיום הסדנה עברו הבוחנים הדרכה קצרה בנושא הזנת הנתונים והקלדתם במערכת ההקלדה הממוחשבת.

המדריכים עקבו אחר השתתפותם של הבוחנים בסימולציות בסדנה, ובמידת הצורך הדריכו וכיוונו אותם כדי שהעברת המבחנים ותיעודם יהיו מיטביים.

ד.2.ח. היערכות לוגיסטית

- לכל מגזר שפה מונתה רכזת תפעול ומעקב, ושתי הרכזות היוו את "הצוות הלוגיסטי" במטח. רכזות התפעול טיפלו בנושאים האלה: פניות לבתי הספר, תיאום ההעברות, החלפת נבחנים

במקרה הצורך, הדפסת המבחנים וחומרי העזר, הפצתם לבוחנים, איסוף החומרים מהשטח אחרי המבחנים ומעקב אחר הקלדת הנתונים.

- נבנתה מערכת ממוחשבת, שבה הקלידו הבוחנים מרחוק את הנתונים ובה נשמרו הציונים והפרמטרים השונים של כל הנבחנים.

ד.2.ט. תהליך התפעול וההעברה של המחקר בבתי הספר

תהליך התפעול וההעברה של המחקר בבתי הספר התבצע בשלבים האלה:

- פנייה בכתב למנהלי בתי הספר שנדגמו למחקר בליווי שיחת טלפון מהצוות הלוגיסטי במטח (השיחה נערכה לפי פרוטוקול מובנה של שיחה). לפניה צורף מכתב להורים של התלמידים שנדגמו.
- מנהלי בתי הספר התבקשו למנות איש קשר – אדם מצוות בית הספר שמכיר את התלמידים (מורה, יועצת, רכזת שכבה) שיהיה בקשר עם הצוות הלוגיסטי במטח לגבי תפעול המחקר בפועל בבית הספר.
- שליחת מסמך לבתי הספר ובו שמות התלמידים שנדגמו.
- איסוף מידע בנוגע לצרכים מיוחדים של תלמידים שנדגמו בעזרת איש הקשר.
- קבלת אישור מההורים שילדם ישתתף במחקר, העברת האישורים לצוות הלוגיסטי, תיאום מועדי ההיבחות – כל אלה בעזרת איש הקשר בבית הספר.
- הגעת הבוחן לבית הספר ביום הבחינה. בכל בית ספר בוחן אחד העביר את המבחנים לכל התלמידים שנדגמו. בדיקה של התנאים להעברת המבחן: חדר שקט שבו מכשיר השמעה להעברת המבחנים המושמעים.
- העברה בפועל של המבחנים בשניים-שלושה מפגשים, לפי היקף הסוללה (מלאה/מקוצרת), לפי הגיל ולפי צורכי הנבחן.
- הקלדת הציונים של כל המבחנים במערכת הממוחשבת (בדרך כלל על ידי הבוחן מהמחשב הביתי).
- העברת חוברת התיעוד של ההעברה הראשונה של הבוחן (באמצעות שליח) כדי לבצע את תהליך ההערכה הבכירה (ראו להלן בסעיף ד.2.יא. – "הערכה בכירה ובקרת איכות").
- תהליך הערכה בכירה על בסיס חוברת התיעוד הראשונה; מתן משוב לבוחן הכולל נקודות לשימור ונקודות לשיפור.
- המשך העברת המבחנים לתלמידים אחרים באותו בית ספר על ידי אותו בוחן.
- העברת חוברת התיעוד של כל התלמידים של הבוחן (באמצעות שליח) לאחר הקלדת הציונים וביצוע הערכה בכירה לכל החוברות.

ד.2.י. דגשים נוספים בהעברת המבחנים

- הנבחנים, הוריהם, מוריהם או בית ספרם לא קיבלו כל משוב על מידת הצלחתם או אי הצלחתם במבחנים ועל הציונים שלהם במבחנים. בסיום העברת המבחנים קיבל הנבחן שי צנוע על השתתפותו במחקר.
- המטרה העיקרית של מחקר החלוץ הייתה לאסוף מידע סטטיסטי מקיף על כמה שיותר פריטים, ומידע שיסייע לקבוע את נקודות הכניסה והעצירה בכל מבחן. כללי הכניסה והעצירה היו גמישים, שכן המטרה הייתה שכל נבחן ייבחן במספר גדול ככל האפשר של פריטים. נקודת הכניסה הייתה לרוב פריט מבחן 1, ונקודת העצירה הייתה לרוב אחרי שבע תשובות שגויות רצופות (תשובות שגויות בפריטי הלמידה וההדגמה לא נספרו כתשובות שגויות). למרות זאת, אם הבוחן הרגיש שהנבחן מתוסכל ומתקשה להמשיך בביצוע המבחן, הוא היה רשאי להפסיק את העברת המבחן גם אם הנבחן לא צבר שבע שגיאות רצופות.
- במבחנים המוגבלים בזמן הונחו הבוחנים להעביר את המבחנים עד תום הזמן שהוקצה, גם אם הנבחן הפסיק לענות על פריטים.
- במבחנים טעוני שפה התבקשו הבוחנים לתעד ולהקליד את כל תשובות הנבחן, לרבות תשובות בשפה אחרת משפת המבחן. תשובות אלה נבדקו, וחושבו השכיחויות שלהן בכל פריט.

ד.2.יא. הערכה בכירה ובקרת איכות

- שני מעריכים שהיו בעלי ידע ונסיון רב בהעברת המבחן ובעלי רקע פסיכומטרי הוגדרו כמעריכים בכירים. עבור כל בוחן, נבדקה חוברת התיעוד של ההעברה הראשונה שלו. המעריכים הבכירים עברו על החוברת אל מול המערכת הממוחשבת וביצעו בקרה וטיוב של הנתונים (נבדק האם הציונים שהבוחן נתן נכונים והאם הציונים הוקלדו נכון למערכת הממוחשבת). במקרים של אי-בהירות או של קשיים בהערכה, הועברו השאלות לשיפוט של ראש צוות הפיתוח ולעיתים להתיעצות עם אנשי ראמ"ה.
- לכל בוחן ניתן משוב בדף סטנדרטי על חוברת התיעוד הראשונה. המשוב עסק בכל ההיבטים של התיעוד והציון.
- במקרים בודדים ניתן משוב גם על ההעברה השניה של בוחן אם בהעברה הראשונה שלו נמצאו טעויות רבות בתיעוד או בציון או בהקלדת הנתונים.
- במקרים בודדים הופסקה עבודתם של בוחנים מאחר שלא היו טובים דיים. ההחלטה התקבלה מתוך שיקול דעת של המעריכים הבכירים ולאחר התייעצות עם מנהלי הפרויקט במטח ובראמ"ה.
- בנוסף הייתה בקרה של המעריכים הבכירים על כל החוברות (100% בקרה). נבדקו כל הציונים שהבוחנים נתנו לתשובות הנבחנים והאם הציונים הוקלדו נכון למערכת הממוחשבת.

ד.2.יב. אתגרים לוגיסטיים מרכזיים בהעברת מחקר החלוץ

- היה קושי לקבל את אישורי ההורים להיבחנות ילדיהם ונדרשו זמן ומאמץ שכנוע רבים כדי להשיגם. בגלל קושי זה חל עיכוב בביצוע המחקר ושונו לוחות הזמנים שנקבעו.

- בגני ילדים רבים התברר שאין בהם חדר מתאים להעברת המבחנים לילדים ונדרשה גמישות הן מצד הגננות הן מצד הבוחנים כדי למצוא פתרונות יצירתיים לבעיה זו (למשל לבחון את הילד בגן בזמן ששאר הילדים נמצאים בחצר).
- לחלק מהבוחנים היו קשיים בעמידה בלוחות הזמנים שהתחייבו אליהם.

ד.2.יג. עיבוד הנתונים

כללי

מטרת עיבוד הנתונים שנאספו במחקר החלוץ הייתה לבדוק את המאפיינים הפסיכומטריים העיקריים של המבחנים המרכיבים את הסוללות. מטרה נוספת הייתה לתמוך בהחלטות בנוגע להרכב הסופי של הפריטים ולגבש את הגרסאות הסופיות של המבחנים.

קבוצות הניתוח

- כל מבחן בכל שפה נותח בנפרד.
 - כל מבחן בכל שפה נותח בחמש קבוצות גיל. הקבוצות הוגדרו לפי דרגת הכיתה של הנבחנים בעת ההעברה. המטרה הייתה לבדוק את תפקוד המבחנים בתוך קבוצות הומוגניות מבחינת הגיל, ולהשוות את תפקוד המבחנים בין קבוצות הגיל השונות.
- קבוצות הגיל שהוגדרו היו אלה:
- גן: טרום-חובה וחובה
- יסודי צעיר: כיתות א'-ג'
- יסודי בוגר: כיתות ד'-ו'
- חטיבת ביניים: כיתות ז'-ט'
- חטיבה עליונה: כיתות י'-י"א

הקבוצות הוגדרו לפי הכיתה שבה התלמיד למד, בלי קשר למוסד שבו הוא למד. למשל אם תלמיד למד בכיתה ט' בתיכון ארבע שנתי (כיתות ט'-יב'), הוא שויך לקבוצה של חטיבת הביניים (כיתות ז'-ט'); אם תלמיד גן חובה למד בבית ספר יסודי במסגרת "החטיבה הצעירה" הוא שויך לקבוצת הגן.

הניתוחים העיקריים

- ניתוח כל אחד מהמבחנים נבדקה התפלגות ציוני המבחן וחושבה המהימנות הפנימית של המבחן (מדד קרונברך α).
- ניתוח פריטים (Item Analysis) של כל מבחן בוצע בנפרד לפי התורה הקלאסית של המבחנים. לכל פריט חושבו הפרמטרים האלה: התפלגות הציונים, ממוצע הציונים (דרגת הקושי של הפריט), שיעור אי-ההשבה, מדד ההבחנה (המתאם בין ציון הפריט ובין הציון הכולל של המבחן - point bi-serial) והמהימנות הפנימית של המבחן ללא הפריט.
- ניתוח תפקוד דיפרנציאלי (DIF- Differential Item Functioning) בוצע לפי שיעור ההצלחה בכל פריט במבחן נתון, בקבוצות הניתוח האלה: נבחנים דוברי עברית לעומת נבחנים דוברי ערבית (במבחנים הצורניים), בנים לעומת בנות, נבחני המגזר החרדי לעומת שאר דוברי העברית ונבחני המגזר הבדואי בדרום לעומת שאר דוברי הערבית.

- נעשה ניתוח איכותני של הפרוטוקולים של העברת המבחנים ושל התשובות שניתנו על ידי הנבחנים בפריטים הפתוחים (לרבות ניתוח של שיעור התשובות שניתנו בשפות שונות משפת המבחן וניתוח של שיעור התשובות שניתנו בסלנג). בנוסף מופו סוגי הבעיות שהתעוררו בהעברת המבחנים (למשל אי בהירות של ההוראות לנבחן).
- בנספח 3 בדוח זה מוצגת דוגמה לניתוח התוצאות של הפריטים במבחן T1 (הבנה מילולית- אנלוגיות מילוליות), בכל שפה בנפרד, בחמש קבוצות הגיל. כמו כן מוצגות בדיקות התפקוד הדיפרנציאלי של פריטי המבחן, בקבוצות המוקד שהוגדרו בכל שפה.

ד.3. גיבוש הנוסחים הסופיים

גיבוש הנוסחים הסופיים של המבחנים נעשה על בסיס המידע והנתונים שנאספו ועובדו במחקר החלוץ. צוותי הפיתוח ומנהלי הפרויקט בחנו את המידע שנאסף על כל מבחן, בכל אחת מהשפות. הצוותים דנו בממצאים וקיבלו החלטות הנוגעות להרכבים הסופיים של המבחנים, למשל אילו פריטים יישארו ואילו פריטים יושמטו.

ד.3.א. שיטת העבודה

- צוותי הפיתוח בעברית ובערבית עבדו בשלבים מסוימים במשותף ובשלבים מסוימים בנפרד.
- כל צוות דן בכל אחד מהמבחנים בנפרד, והתייחס לעיבוד הנתונים ולמידע שנאסף לגבי תפקוד הפריטים בכל מבחן.
- לאחר גיבוש ההמלצות לנוסחים הסופיים בשתי השפות, הושוו הנוסחים בעברית לנוסחים בערבית ונעשו התאמות סופיות.
- ההמלצות לנוסחים הסופיים שגובשו בצוותים עברו לבדיקה ולאישור של מנהלי הפרויקט בראמ"ה שבחנו את ההמלצות גם מתוך הסתכלות הוליסטית על כלל הפריטים בכל המבחנים.
- נעשה עיבוד נתונים חוזר לכל המבחנים והפריטים שנכללו בנוסחים הסופיים שלהם. עיבוד זה נעשה כדי לאשש את תקינות הנוסחים ואת התפקוד הפסיכומטרי שלהם. בשלב זה נעשו עדכונים קלים סופיים נוספים (למשל שינוי מיקום של פריט).
- הנוסחים הסופיים אושרו בישיבה שהשתתפו בה ראשי הצוותים, הפסיכומטריקאים ומנהלי הפרויקט בראמ"ה.

ד.3.ב. השיקולים בנוגע לפסילת פריטים ולחידוד מחוונים

שיקולים בהליך פסילת פריטים

- דרגת הקושי של הפריט: מדד קושי נמוך מדי או גבוה מדי (פריט קל מדי או קשה מדי), אצל רוב הנבחנים או בקבוצת גיל מסוימת.
- רמת ההבחנה של הפריט: מדד הבחנה נמוך מ-0.20 אצל כל הנבחנים או בקבוצת גיל מסוימת.
- התפקוד הדיפרנציאלי של הפריט – תפקוד המצביע על הטיה נגד מגזר כלשהו (ממלכתי וממ"ד לעומת חרדי; בדואים בדרום לעומת שאר הנבחנים; בנים לעומת בנות).

- תוקף הפריט לפי טיב הקשר בין רמת הקושי של הפריט ובין הגיל. הציפיה הייתה שדרגת הקושי של פריט תלך ותתרד ככל שעולים בגיל, כלומר הפריט נעשה קל יותר ככל שהנבחנים מבוגרים יותר.
- מידע בפרוטוקולים של ההעברות בנוגע לשיעור גבוה של בעיות שהתעוררו בהעברת הפריט. למשל אי-בהירות בהוראות שאי אפשר לתקנה.
- פריטים שהיו בהם הרבה תגובות שצוות הפיתוח התקשה להחליט האם הן נכונות או לא (ההנחה הייתה שיהיו בעיות בציון התגובות לפריט גם בהמשך).

עדכון המחווניים ושאלות הבירור

בעקבות התגובות המילוליות של הנבחנים שנאספו במחקר החלוץ, עלו סוגיות הנוגעות למחווניים ולשאלות הבירור. למשל, בפריטים רבים התקבלו תשובות נבחנים שלא נצפו מראש ולכן לא הייתה התייחסות אליהם במחווני. לגבי תגובות אלו היה צריך לשקול האם להוסיף אותן למחווני ואם כן – האם להחשיבן כתגובה נכונה, כתגובה שגויה או כתגובה שיש לשאול לגביה שאלת בירור. גם בנוגע לתגובות שהיו צפויות, נעשה בעקבות מחקר החלוץ דיון מחודש כיצד להתייחס אליהן. למשל, במקרים בהם ניתנו תגובות (נכונות) בשפה שאינה שפת המבחן. האם לקבל תגובות אלו כנכונות (מכיוון שהן נכונות תכנית), כשגויות (מכיוון שהמילה המדויקת בשפת המבחן לא נאמרה), או לשאול לגביהן שאלה או מספר שאלות בירור¹⁰. במגזר דוברי העברית, מקרים אלו אירעו בעיקר בקרב נבחנים עולים חדשים או נבחנים מבתיים שעברית אינה השפה הדבורה העיקרית בהם. במגזר דוברי הערבית, רוב המקרים האלו ארעו כאשר נבחנים השיבו לשאלות מסוימות בעברית או נתנו את שם המונח בעברית. ההתלבטות הייתה גדולה יותר במקרים שבהם מילה מסוימת בעברית מוכרת ומשמשת במגזר הערבי יותר מהמונח בשפה הערבית שפחות מוכר או נהוג, וכשנרשם שיעור גבוה יחסית של תגובות כאלו.

ד.3.ג. שיקולים בנוגע לקביעת הרכב הפריטים וסידורם בכל אחד מהמבחנים

- פיזור ברמות הקושי: הייתה חשיבות לכך שבכל מבחן יהיו פריטים במנעד רחב ככל האפשר של דרגות קושי, מן הקלים ביותר עד הקשים ביותר.
- כמות הפריטים: הקו המנחה היה לשמור על כמות הפריטים במבחנים כמו שקיים בגרסאות המקור. ברוב המבחנים מספר הפריטים שנקבע בסופו של דבר היה זהה או דומה למספר הפריטים במבחנים המקוריים. במקרים מסוימים המבחנים בעברית ובערבית הוארכו או קוצרו לעומת המקור משיקולים שונים. למשל התווספו פריטים ברמת קושי מסוימת אם התברר שחסרים. קוצרו מבחנים שהכבידו על הנבחן, כמו חלק ממבחני העיבוד השמיעתי שקוצרו בלי לפגוע בתפקוד הפסיכומטרי שלהם. במקרים בודדים הוגדל מספר הפריטים הקלים המתאימים

¹⁰ במקרה כזה, נשקל גם מספר שאלות הבירור שכבר ישנן במבחן והעומס על הבוחן והנבחן הכרוך בבירורים.

לנבחנים הצעירים ולנבחני החינוך המיוחד, כדי לשפר את רמת ההבחנה באוכלוסיות אלה. למשל תת-מבחן 1א – "אוצר מלים לפי תמונות" הוארך בשלושה פריטים לעומת המקור.

- **קביעת סדר הפריטים:** בעזרת הנתונים על דרגות הקושי של הפריטים בקבוצות הגיל השונות ומעבר לקבוצות הגיל נקבע סדר הפריטים בכל מבחן. במבחנים שבהם הפריטים מוצגים בסדר עולה לפי דרגות הקושי, סודרו הפריטים הראשונים לפי רמות הקושי שהובחנו בגיל הגן, פריטי המבחן האמצעיים לפי רמות הקושי שהובחנו בכל קבוצות הגיל והפריטים האחרונים סודרו לפי דרגות הקושי שהובחנו בגיל תיכון.
- **התמקדות מול פיזור:** ניתנה תשומת לב לשאלות הבאות: (1) האם פריטי המבחנים לקוחים יתר על המידה מעולם תוכן מצומצם או שואלים על היבט לשוני מצומצם? (2) האם פריטי המבחנים לקוחים מעולמות תוכן רבים מדי, ואינם מודדים את מה שרצינו למדוד? שאלות אלה עלו לדיון והשפיעו על קביעת הפריטים במבחנים.
- **תלות בין פריטים:** (1) האם יש תלות מבחינת התוכן בין פריט במבחן כלשהו ובין פריט אחר באותו מבחן? (2) האם יש תלות מבחינת התוכן בין פריטים במבחן כלשהו ובין פריטים במבחן אחר? שאלות אלה עלו לדיון והשפיעו על קביעת הפריטים במבחנים.
- **דמיון בנוסחים בין שתי השפות:** בכל מקום שבו זה היה אפשרי נעשה ניסיון לשמור על דמיון בין הנוסח בעברית ובין הנוסח בערבית מבחינת תוכני הפריטים ומבחינת הסדר שלהם.
- **המבחנים מול גרסת המקור:** בחלק מהמבחנים הצורניים נשמרה זהות הפריטים לעומת גרסת המקור, ובמידת האפשר גם הסדר שלהם.

ד.3.ד. ניתוח של נתוני המבחנים בגרסתם הסופית

לאחר קביעת סדר הפריטים בכל מבחן ולאחר קביעת התצורה הסופית של המבחנים, נעשה עיבוד חוזר של הנתונים שהתבסס רק על הפריטים שנשארו במבחנים בגרסתם הסופית. המטרה הייתה לאשש את ההחלטות לגבי הרכב הפריטים והסדר שלהם כפי שיהיו בנוסחים הסופיים. ההנחה הייתה שכיוון שמספר הפריטים שהושמטו היה קטן מאוד, ולא הייתה מגבלה של זמן במבחן, השמטתם לא תשנה את תפקוד הפריטים שנשארו. אכן הממצאים מעיבוד הנתונים החוזר היו דומים מאד לממצאים מעיבוד הנתונים הראשון.

אחרי עיבוד הנתונים החוזר של המבחנים ושל הפריטים התקבלו כמה החלטות נוספות כמתואר להלן.

ד.3.ה. גיבוש כללי הכניסה והיציאה במבחנים

גיבוש כללי הכניסה והיציאה במבחנים, בקבוצות הגיל השונות, התבסס על הניתוח החוזר של הנתונים בגרסאות המוצעות.

- לשם קביעת נקודות הכניסה נבחנו נקודות הכניסה במבחן המקור באנגלית וכמו כן שיעורי ההצלחה בפריטים הראשונים בקבוצות הגיל השונות. תחילה הוחלט כמה נקודות כניסה צריכות להיות בכל מבחן (לפי קבוצות גיל) ולאחר מכן נקבעו נקודות הכניסה. (עוד על כללי כניסה ויציאה, ראו לעיל בסעיף ג.2.יד. – "כללי כניסה ויציאה").
- הוחלט שנקודת הכניסה תהיה תמיד בתחילת עמוד או בתחילת טור, ובהתאם לכך עומדו הפריטים בדפי המבחן.

- בכל המבחנים שבהם כללי היציאה היו פשוטים, למשל "הפסק אחרי שלוש טעויות רצופות" – נבדקה התקפות של כלל היציאה במבחן כלשהו באמצעות בדיקת אחוזי ההצלחה בפריטים שהיו אחרי רצף של שלוש טעויות.
- גם במבחנים שבהם כללי היציאה נקבעו לפי מספר הנקודות שצבר הנבחן עד לנקודת העצירה (למשל במבחנים 2, 5, 16), נקבעו כללי יציאה בהתאם לממצאים. התקפות של כללי היציאה במבחן כלשהו נבדקה באמצעות בדיקת אחוזי ההצלחה בפריטים שלאחר נקודות העצירה.
- במבחנים שבהם כלל היציאה היה "כל הפריטים במקבץ שגויים", הובאו בחשבון סדר הפריטים בעמוד וגם החלוקה למקבצים.
- בנוסף הוחלט כי בדומה למקור, במבחנים שבהם הפריטים מוצגים בעמוד הפונה לנבחן, ימשיך הנבחן בעבודתו עד סוף העמוד שאותו התחיל, גם אם לפי כלל היציאה היה עליו לסיים את המבחן באמצע העמוד.

טבלה 5: דוגמה – ריכוז מידע לקראת דיון בנושא גיבוש הנוסח הסופי של

תת-מבחן 1א בעברית¹¹

מספר הפריט (במחקר החלוץ)	מדדי הבחנה (α קרונבארך)	DIF מגזרים	DIF בנים או בנות	קשיים שהתגלו בהעברה	תלות בין פריטים; גיוון תכנים	שאלות בירור
פריט 1	100% ענו נכון					
פריט 5	100% ענו נכון; המתאם -0.01					
פריט 6	96% ענו נכון בקרב נבחני גן; המתאם 0.03			"מפתחות" - התשובה היחידה בלשון רבים בכל פריטי המבחן - נתון חריג		
פריט 12	מתאם נמוך בחט"ב: 0.18			"מברגה" - חוזר על עצמו		
פריט 16		המגזר החרדי -0.8				

¹¹ בטבלה זו מופיעים מספרי הפריטים בלבד כדי לשמור על סודיות התוכן של הפריטים.

מספר הפריט (במחקר החלוץ)	מדדי הבחנה (α קרונבאך)	DIF מגזרים	DIF בנים או בנות	קשיים שהתגלו בהעברה	תלות בין פריטים; גיוון תכנים	שאלות בירור
פריט 18		המגזר החרדי -0.5			תכנים דומים בפריטים 23, 22	
פריט 19		המגזר החרדי -0.7				
פריט 22		המגזר החרדי -0.5			תכנים דומים בפריטים 23, 18	19.4%
פריט 23		המגזר החרדי -0.5		"מגדלור", - כיצד לציין?	תכנים דומים בפריטים 22, 18	10.3%
פריט 24						15%
פריט 25						19.3%
פריט 26						33.1%
פריט 28			בנים +0.3			54.5%
פריט 29		המגזר החרדי -0.5				23.3%
פריט 30						45.3%
פריט 31		המגזר החרדי +0.5		magel פער בין אופן ההגייה התקני ובין אופן ההגייה השגור	תוכן דומה בפריטים 34, 32	12.3%

מספר הפריט (במחקר החלוץ)	מדדי הבחנה (α) (קרונבאך)	DIF מגזרים	DIF בנים או בנות	קשיים שהתגלו בהעברה	תלות בין פריטים; גיוון תכנים	שאלות בירור
פריט 32						36.5%
פריט 33						19.5%
פריט 34		המגזר החרדי +0.4				21%
פריט 35	0% ענו נכון; המתאם 0.08					27.1%

הסבר הכותרות בטבלה

מדדי הבחנה (α קרונבאך) – פריטים בעלי רמת הבחנה חריגה המחייבת בדיקה.

DIF מגזרים – פריטים שהיו מוטים נגד מגזר כלשהו באופן חריג שחייב בדיקה.

DIF בנים או בנות – פריטים שהיו מוטים נגד בנים או נגד בנות באופן חריג שחייב בדיקה.

קשיים שהתגלו בהעברה – קשיים שחשוב להביא בחשבון בבניית הנוסח הסופי, למשל אי-בהירות בפריט (האם יש לפסול את הפריט? האם נדרש תיקון מחוון?)

תלות בין פריטים; גיוון תכנים – תלות מבחינת התוכן: דמיון בין תוכן של פריט כלשהו

ובין תכנים של פריטים אחרים במבחן, או בין תוכן של פריט כלשהו ובין פריטים

במבחנים אחרים. כמו כן בעיות הנוגעות למידת הגיוון שהפריט מאפשר.

שאלות בירור – פריטים שבהם היה שיעור גבוה של שאלות בירור.

ד.4. היקף השינויים ואפיונם במבחנים השונים

טבלה 6: היקפי הפיתוח של הפריטים בכל שלבי המחקר
(בעמודות הימניות כתובים בסוגריים מספרי המבחנים במקור באנגלית)

מספר המבחן	תת-המבחן	מספר הפריטים במקור	מספר הפריטים במחקרי החלוץ המצומצמים		מספר הפריטים במחקר החלוץ		מספר הפריטים בנוסחים הסופיים		זהות בין עברית וערבית
			ערבית	ערבית	ערבית	ערבית	ערבית	ערבית	
(1) הבנה מילולית	א1 (א1) אוצר מילים על-פי תמונות	23	36	35	35	35	26	26	לא
	ב1 (ב1) מילים נרדפות	15	25	32	25	25	18	18	לא
	ג1 (ג1) הפכים	18	24	36	24	26	20	20	לא
	ד1 (ד1) אנלוגיות מילוליות	15	23	31	24	29	20	20	לא
(2) למידה חזותית שמיעתית		משפטים 17 סיפורים 7	משפטים 19 סיפורים 8	משפטים 22 סיפורים 8	משפטים 19 סיפורים 8	משפטים 19 סיפורים 8	משפטים 19 סיפורים 8	משפטים 19 סיפורים 8	לא
(3) יחסים מרחביים		33	45	45	42	42	33	33	כן
(4) מיזוג צלילים		33	43	41	39	41	32	32	לא
(5) יצירת כללים		40	48	48	47	47	40	40	כן
(6) התאמה חזותית		התאמה חזותית 1: 26 התאמה חזותית 2: 60	התאמה חזותית 1: 26 התאמה חזותית 2: 60	התאמה חזותית 1: 26 התאמה חזותית 2: 60	לא היה במחקר החלוץ	לא היה במחקר החלוץ	התאמה חזותית 1: 26 התאמה חזותית 2: 60	התאמה חזותית 1: 26 התאמה חזותית 2: 60	כן
(7) זכירת מילים		24	29	29	29	29	24	24	לא ובנוסף, בעברית מועבר במכשיר השמעה ובערבית בעל-פה

לא	26	26	32	32	34	38	26	8א (11א) "איפה אפשר למצוא...?"	8 (11) ידיעות כלליות
לא	22	22	27	27	33	32	22	8ב (11ב) "בשביל מה משתמשים ב...?"	
כן אבל זהו מבחן תלוי שפה	3	3	3	3	4	4	3		9 (12) שטף סמנטי
כן	23	23	27	28	27	28	24		10 (13) זיהוי ציורים
לא	30	30	50	48	60	51	44		11 (8) מילים לא שלמות
כן	35	35	35	35	35	35	35		12 (15) אנליזה- סינתזה
כן	40	40	40	40	40	40	40		13 (16) מהירות החלטה
הפריטים זהים, אך בעברית הם מועברים במכשיר השמעה ובערבית בעל-פה	30	30	30	30	30	30	30		14 (7) זכירת ספרות לאחור
כן	72	72	72	72	72	72	72		15 (21) זכירת שמות
כן	21	21	27	27	27	27	21		16 (19) תכנון
לא	50	50	50	50	52	50	50		17 (14) קשב שמיעתי
כן	47	47	55	55	47	47	47		18 (24) סדרות מספרים
כן	30	30	לא היה במחקר החלוץ	לא היה במחקר החלוץ	30	30	30		19 (26) איתור ציורים מהיר

לא ובנוסף, בעברית מועבר במכשיר השמעה ובערבית בעל-פה	21	21	24	24	24	24	21		20 (9) זיכרון עבודה שמיעתי
כן	120	120	120	120	120	120	120		21 (18) שיום תמונות מהיר
כן	24	24	29	29	29	29	24		22 (28) סיבוב עצמים
כן	36	36	לא היה במחקר החלוץ	לא היה במחקר החלוץ	36	36	36		23 (30) זכירת שמות- מושהית

4. נספחים

נספח 1 - שמות המבחנים בגרסאות בישראל (בעברית ובערבית)

ובגרסת המקור

אוגדן 2 - הסוללה המורחבת - ישראל				אוגדן 1 - הסוללה הבסיסית - ישראל			
שם המבחן בגרסת המקור	מספר המבחן בגרסת המקור	שם המבחן	מספר המבחן החדש	שם המבחן בגרסת המקור	מספר המבחן בגרסת המקור	שם המבחן	מספר המבחן החדש
Memory for Names	21	זכירת שמות تذكر أسماء	15	Verbal Comprehension	1	הבנה מילולית الفهم الكلامي	1
Planning	19	תכנון تخطيط	16	Visual-Auditory Learning	2	למידה חזותית-שמיעתית تعلم بصري-سمعي	2
Auditory Attention	14	קשב שמיעתי الانتباه السمعي	17	Spatial Relations	3	עיבוד מרחבי إدراك المدى	3
Number Series	24	סדרות מספרים متواليات عددية	18	Sound Blending	4	מיזוג צלילים دمج الأصوات	4
Cross Out	26	איתור ציורים מהיר تمييز سريع لرسومات	19	Concept Formation	5	יצירת כללים استنتاج قواعد	5
Auditory Working Memory	9	זיכרון עבודה שמיעתי الذاكرة الفعالة السمعية	20	Visual Matching	6	התאמה חזותית ملاءمة بصرية	6
Rapid Picture Naming	18	שיום תמונות מהיר تسمية سريعة لرسومات	21	Memory for Words	17	זכירת מילים تذكر كلمات	7

Block Rotation	28	סיבוב עצמים إدارة أشكال	22	General Information	11	ידע כללי معلومات عامة	8
Memory for Named-Delayed	30	זכירת שמות מושהית تذكر أسماء مؤجل	23	Retrieval Fluency	12	שטף סמנטי الطلاقة الدلالية	9

Picture Recognition	13	זיהוי ציורים تمييز رسومات	10
Incomplete Words	8	מלים לא שלמות كلمات غير كاملة	11
Analysis-Synthesis	15	אנליזה- סינתזה تحليل تركيب	12
Decision Speed	16	מהירות החלטה سرعة اتخاذ القرار	13
Number Reversed	7	זכירת ספרות לאחור تذكر أرقام بترتيب عكسي	14

נספח 2 – המבחנים שלא הותאמו לגרסאות בישראל

שם המבחן	מספר המבחן בגרסת המקור	שם המבחן בגרסת המקור
למידה חזותית-שמיעתית מושהית	10	Visual-Auditory Learning - Delayed
סימון זוגות	20	Pair Cancellation
השלמת גשטלט	22	Visual Closure
הבחנה בין צירופי קולות	23	Sound Patterns - Voice
מטריצות מספרים	25	Number Matrices
זכירת משפטים	27	Memory for Sentences
הבחנה בין צירופי צלילים מוזיקליים	29	Sound Patterns - Music

נספח 3 – עיבוד נתוני תת-מבחן 1ד'

במסגרת עיבוד הנתונים של מחקר החלוץ בוצעו ניתוחי פריטים לכל המבחנים בכל אחת מהשפות. כל מבחן נותח כמה פעמים – עבור כל אחת מחמש קבוצות הגיל בנפרד ועבור כלל הנבחנים במדגם מעבר לקבוצות הגיל. במבחנים שהוצאו מהם פריטים בעקבות עיבוד הנתונים, נערכו ניתוחים נוספים על המבחן המקוצר, כלומר המבחן בגרסתו הסופית ללא הפריטים שנפסלו. לשם הדגמה, בנספח זה מוצגות תוצאות ניתוח הפריטים של תת-מבחן 1ד' בעברית. התוצאות מוצגות הן עבור כלל הנבחנים והן עבור כל קבוצת גיל בנפרד, הן לפני קיצור המבחן והן על נוסח המבחן הסופי, לאחר הקיצור.

תת-מבחן ד' – ממוצעים וסטיות תקן של הפריטים בנוסח המקורי (24 פריטים)												
תלמידים בכיתות ז-ט n= 161		תלמידים בכיתות ד-ו n= 190		תלמידים בכיתות א-ג n= 221		תלמידים בגני ילדים n= 167		כל הנבחים n= 1011				
מהימנות פנימית (קרונברך אלפא) $\alpha=0.783$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.746$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.745$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.815$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.785$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.894$		
ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	מספר פריט במחקר החלוצי
0.00	1.00	0.00	1.00	0.10	0.99	0.21	0.95	0.36	0.84	0.19	0.96	פריט 1
0.34	0.86	0.28	0.91	0.32	0.88	0.45	0.73	0.50	0.43	0.41	0.78	פריט 2
0.17	0.97	0.24	0.94	0.30	0.90	0.37	0.84	0.46	0.71	0.33	0.88	פריט 3
0.08	0.99	0.00	1.00	0.12	0.98	0.26	0.93	0.48	0.65	0.27	0.92	פריט 4
0.19	0.96	0.19	0.96	0.19	0.96	0.35	0.86	0.50	0.51	0.34	0.86	פריט 5
0.25	0.93	0.28	0.92	0.33	0.88	0.44	0.74	0.50	0.50	0.40	0.80	פריט 6
0.19	0.96	0.19	0.96	0.21	0.95	0.44	0.75	0.42	0.22	0.41	0.79	פריט 7
0.14	0.98	0.19	0.96	0.30	0.90	0.42	0.77	0.49	0.38	0.39	0.82	פריט 8
0.28	0.91	0.29	0.91	0.39	0.82	0.49	0.59	0.40	0.20	0.46	0.70	פריט 9
0.38	0.83	0.42	0.78	0.44	0.74	0.50	0.53	0.30	0.10	0.49	0.61	פריט 10
0.46	0.71	0.49	0.59	0.49	0.39	0.40	0.20	0.20	0.04	0.49	0.40	פריט 11
0.30	0.90	0.34	0.87	0.38	0.83	0.47	0.67	0.36	0.16	0.46	0.70	פריט 12
0.32	0.89	0.42	0.77	0.47	0.67	0.50	0.43	0.24	0.06	0.49	0.58	פריט 13
0.50	0.55	0.50	0.54	0.47	0.33	0.28	0.09	0.00	0.00	0.46	0.31	פריט 14
0.49	0.60	0.50	0.47	0.50	0.46	0.46	0.30	0.24	0.06	0.49	0.38	פריט 15
0.47	0.68	0.49	0.58	0.48	0.36	0.36	0.15	0.08	0.01	0.48	0.37	פריט 16
0.39	0.18	0.33	0.13	0.26	0.07	0.07	0.00	0.00	0.00	0.27	0.08	פריט 17
0.48	0.37	0.46	0.31	0.43	0.25	0.32	0.12	0.19	0.04	0.41	0.22	פריט 18
0.46	0.70	0.49	0.58	0.48	0.36	0.26	0.07	0.00	0.00	0.48	0.35	פריט 19
0.48	0.63	0.48	0.64	0.50	0.57	0.50	0.44	0.20	0.04	0.50	0.49	פריט 20
0.40	0.19	0.30	0.10	0.16	0.03	0.07	0.00	0.00	0.00	0.25	0.06	פריט 21
0.47	0.34	0.40	0.20	0.22	0.05	0.13	0.02	0.00	0.00	0.33	0.12	פריט 22
0.11	0.01	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	פריט 23
0.16	0.02	0.15	0.02	0.12	0.02	0.00	0.00	0.00	0.00	0.11	0.01	פריט 24

תת-מבחן 1ד' – מתאמים בין הפריטים לציון במבחן בנוסח המקורי (24 פריטים) ¹²												
תלמידים בכיתות ז-ט n= 161		תלמידים בכיתות ד-ו n= 190		תלמידים בכיתות א-ג n= 221		תלמידים בגני ילדים n= 167		כל הנבחים n= 1011				
מהימנות פנימית (קרונברך אלפא) $\alpha=0.783$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.746$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.745$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.815$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.785$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.894$		
מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מספר פריט במחקר החלוץ
0.78		0.75		0.74	0.19	0.81	0.31	0.79	0.16	0.89	0.31	פריט 1
0.78	0.23	0.75	0.08	0.74	0.24	0.81	0.35	0.78	0.39	0.89	0.45	פריט 2
0.78	0.28	0.74	0.17	0.75	0.13	0.81	0.29	0.78	0.38	0.89	0.35	פריט 3
0.78	0.33	0.75		0.74	0.10	0.81	0.46	0.77	0.44	0.89	0.48	פריט 4
0.78	0.22	0.74	0.24	0.74	0.16	0.81	0.37	0.77	0.47	0.89	0.51	פריט 5
0.77	0.39	0.75	0.08	0.74	0.18	0.81	0.41	0.76	0.56	0.89	0.48	פריט 6
0.78	0.34	0.74	0.24	0.74	0.29	0.80	0.58	0.77	0.51	0.89	0.67	פריט 7
0.78	0.27	0.74	0.25	0.74	0.30	0.80	0.50	0.77	0.47	0.89	0.59	פריט 8
0.77	0.45	0.74	0.28	0.73	0.36	0.80	0.48	0.77	0.49	0.89	0.64	פריט 9
0.76	0.51	0.72	0.51	0.72	0.43	0.80	0.48	0.78	0.33	0.89	0.64	פריט 10
0.77	0.46	0.72	0.46	0.72	0.43	0.81	0.35	0.78	0.25	0.89	0.57	פריט 11
0.78	0.28	0.74	0.23	0.74	0.30	0.80	0.44	0.77	0.53	0.89	0.60	פריט 12
0.77	0.40	0.74	0.26	0.74	0.27	0.81	0.41	0.78	0.36	0.89	0.59	פריט 13
0.78	0.31	0.73	0.39	0.73	0.35	0.81	0.30	0.79		0.89	0.51	פריט 14
0.77	0.45	0.73	0.39	0.73	0.35	0.81	0.38	0.77	0.46	0.89	0.50	פריט 15
0.76	0.53	0.72	0.44	0.72	0.51	0.81	0.40	0.78	0.23	0.89	0.60	פריט 16
0.78	0.22	0.74	0.29	0.74	0.27	0.82	0.16	0.79		0.89	0.30	פריט 17
0.77	0.40	0.73	0.42	0.73	0.32	0.81	0.32	0.78	0.37	0.89	0.43	פריט 18
0.78	0.34	0.74	0.31	0.73	0.40	0.81	0.29	0.79		0.89	0.54	פריט 19
0.78	0.29	0.73	0.39	0.73	0.35	0.80	0.49	0.78	0.38	0.89	0.54	פריט 20

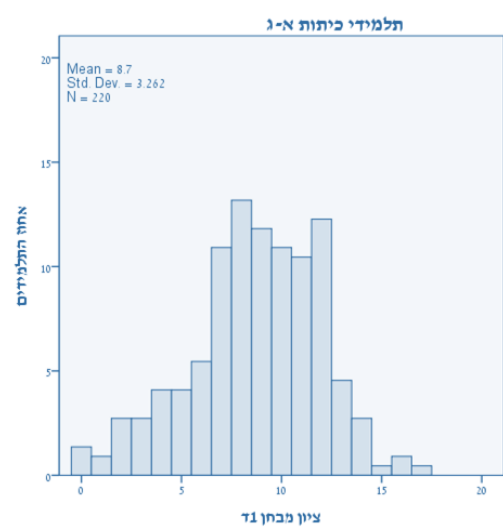
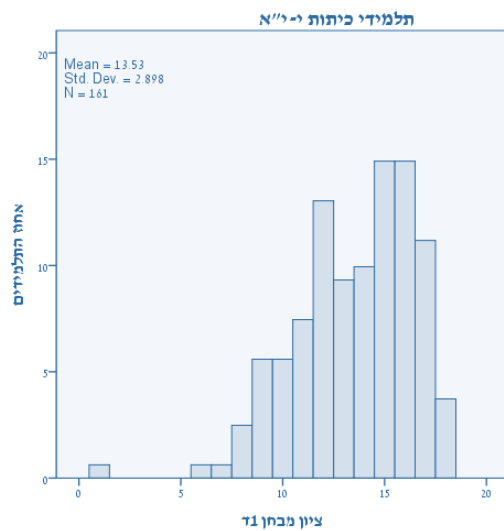
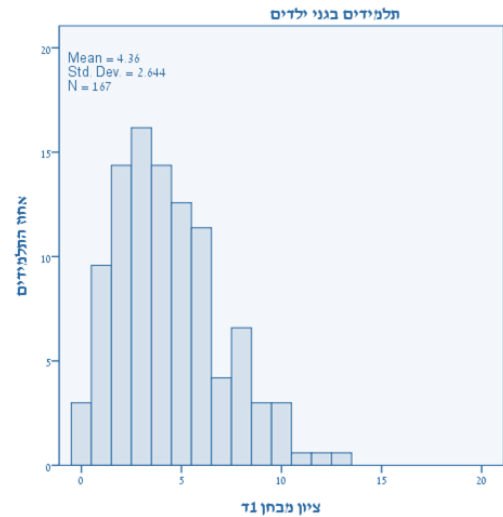
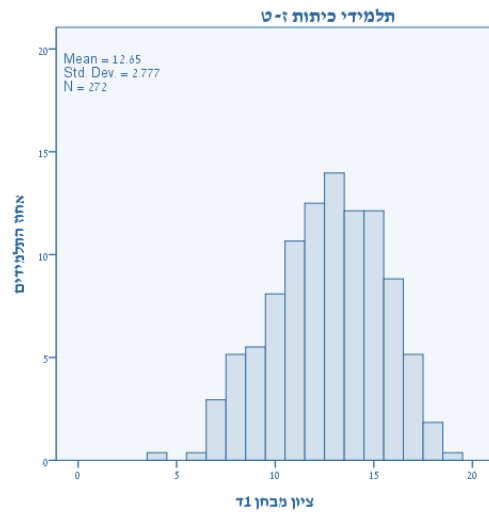
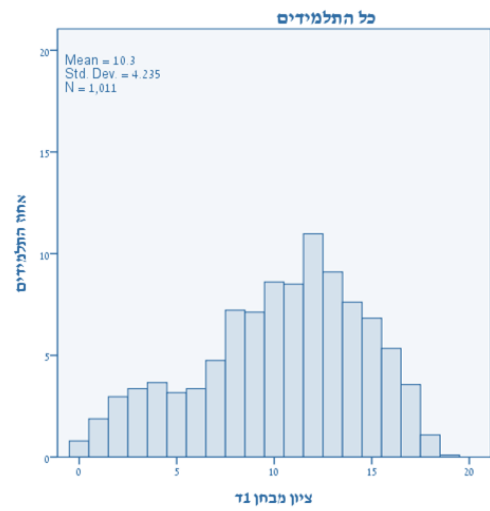
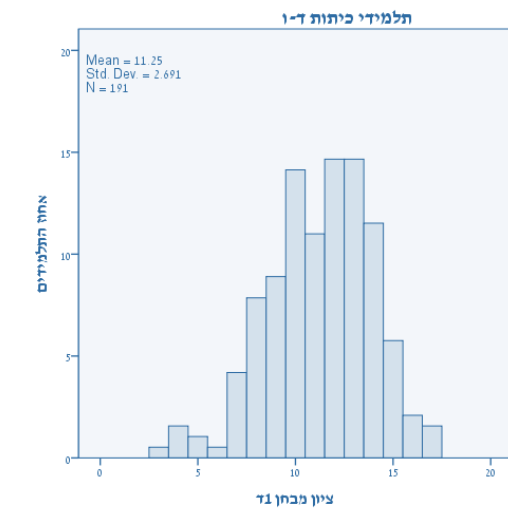
¹² כאשר לא מופיע ערך בתא מסוים הסיבה היא היעדר מתאם בגלל שונות אפס בפריט.

פריט 21	0.29	0.89		0.79	0.03	0.82	0.19	0.74	0.32	0.74	0.28	0.78
פריט 22	0.40	0.89		0.79	0.13	0.81	0.27	0.74	0.37	0.73	0.48	0.77
פריט 23	0.08	0.90		0.79		0.82		0.75	0.12	0.75	0.06	0.78
פריט 24	0.12	0.90		0.79		0.82	0.09	0.75	0.11	0.75	0.12	0.78

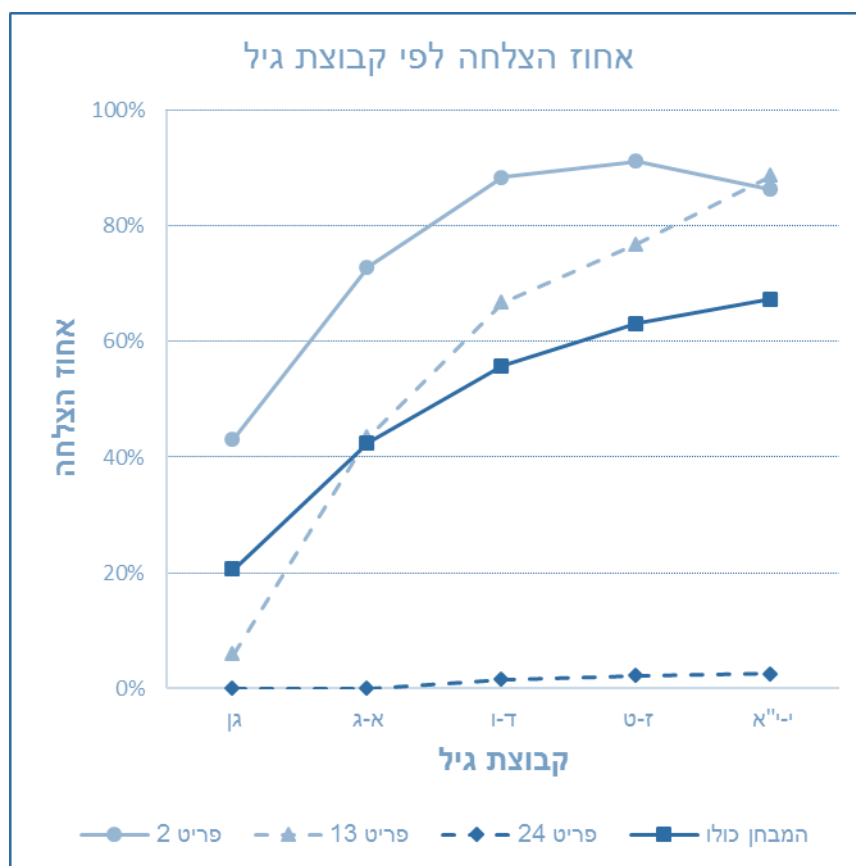
תת-מבחן 1ד' – ממוצעים וסטיות תקן של הפריטים בנוסח המקוצר (20 פריטים)													
		כל הנבחנים n= 1011		תלמידים בגני ילדים n= 167		תלמידים בכיתות א-ג n= 221		תלמידים בכיתות ד-ו n= 190		תלמידים בכיתות ז-ט n= 272		תלמידים בכיתות י-י"א n= 161	
		מהימנות פנימית (קרונברך אלפא) $\alpha=0.875$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.756$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.793$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.700$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.726$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.762$	
מספר הפריט בנוסח הסופי	מספר פריט במחקר החלוץ	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן	ממוצע	סטיית תקן
פריט 1	פריט 1	0.96	0.19	0.84	0.36	0.95	0.21	0.99	0.10	1.00	0.00	1.00	0.00
פריט 2	פריט 3	0.88	0.33	0.71	0.46	0.84	0.37	0.90	0.30	0.94	0.24	0.97	0.17
פריט 3	פריט 4	0.92	0.27	0.65	0.48	0.93	0.26	0.98	0.12	1.00	0.00	0.99	0.08
פריט 4	פריט 5	0.86	0.34	0.51	0.50	0.85	0.35	0.96	0.19	0.96	0.19	0.96	0.19
פריט 5	פריט 6	0.80	0.40	0.50	0.50	0.74	0.44	0.88	0.33	0.92	0.28	0.93	0.25
פריט 6	פריט 8	0.82	0.39	0.38	0.49	0.77	0.42	0.90	0.30	0.96	0.19	0.98	0.14
פריט 7	פריט 7	0.79	0.41	0.22	0.42	0.75	0.44	0.95	0.21	0.96	0.19	0.96	0.19
פריט 8	פריט 9	0.70	0.46	0.20	0.40	0.59	0.49	0.81	0.39	0.91	0.29	0.91	0.28
פריט 9	פריט 10	0.61	0.49	0.10	0.30	0.53	0.50	0.74	0.44	0.78	0.42	0.83	0.38
פריט 10	פריט 13	0.58	0.49	0.06	0.24	0.43	0.50	0.67	0.47	0.77	0.42	0.89	0.32
פריט 11	פריט 20	0.49	0.50	0.04	0.20	0.44	0.50	0.58	0.50	0.64	0.48	0.63	0.48
פריט 12	פריט 11	0.40	0.49	0.04	0.20	0.20	0.40	0.39	0.49	0.59	0.49	0.71	0.46
פריט 13	פריט 16	0.37	0.48	0.01	0.08	0.15	0.36	0.37	0.48	0.58	0.49	0.68	0.47
פריט 14	פריט 15	0.38	0.49	0.06	0.24	0.30	0.46	0.47	0.50	0.47	0.50	0.60	0.49
פריט 15	פריט 14	0.31	0.46	0.00	0.00	0.09	0.28	0.32	0.47	0.54	0.50	0.55	0.50
פריט 16	פריט 18	0.22	0.41	0.04	0.19	0.12	0.32	0.25	0.43	0.31	0.46	0.37	0.48
פריט 17	פריט 22	0.12	0.33	0.00	0.00	0.02	0.13	0.05	0.22	0.20	0.40	0.34	0.47
פריט 18	פריט 21	0.06	0.25	0.00	0.00	0.00	0.07	0.03	0.16	0.10	0.30	0.19	0.40
פריט 19	פריט 24	0.01	0.11	0.00	0.00	0.00	0.00	0.02	0.12	0.02	0.15	0.02	0.16
פריט 20	פריט 23	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.09	0.01	0.11

תת-מבחן 1ד' – מתאמים בין הפריטים לציון במבחן בנוסח המקוצר (20 פריטים)													
תלמידים בכיתות י"א n= 161		תלמידים בכיתות ז-ט n= 272		תלמידים בכיתות ד-ו n= 190		תלמידים בכיתות א-ג n= 221		תלמידים בגני ילדים n= 167		כל הנבחרים n= 1011			
מהימנות פנימית (קרונברך אלפא) $\alpha=0.762$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.726$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.700$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.793$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.756$		מהימנות פנימית (קרונברך אלפא) $\alpha=0.875$			
מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מהימנות המבחן ללא הפריט	מתאם מנוכה בין הפריט לציון במבחן	מספר פריט במחקר החלוץ	מספר הפריט בנוסח הסופי
0.76		0.73		0.70	0.18	0.79	0.32	0.76	0.18	0.87	0.31	פריט 1	פריט 1
0.76	0.26	0.73	0.15	0.70	0.16	0.79	0.27	0.74	0.38	0.87	0.35	פריט 3	פריט 2
0.76	0.32	0.73		0.70	0.09	0.78	0.46	0.74	0.42	0.87	0.48	פריט 4	פריט 3
0.76	0.22	0.72	0.24	0.70	0.18	0.78	0.36	0.74	0.42	0.87	0.50	פריט 5	פריט 4
0.75	0.39	0.73	0.08	0.70	0.18	0.78	0.42	0.72	0.58	0.87	0.48	פריט 6	פריט 5
0.76	0.24	0.72	0.27	0.69	0.27	0.78	0.48	0.73	0.45	0.87	0.58	פריט 8	פריט 6
0.76	0.32	0.72	0.22	0.69	0.31	0.77	0.58	0.73	0.50	0.86	0.66	פריט 7	פריט 7
0.75	0.45	0.72	0.28	0.68	0.33	0.77	0.50	0.73	0.48	0.86	0.63	פריט 9	פריט 8
0.74	0.50	0.69	0.50	0.67	0.42	0.78	0.47	0.75	0.32	0.86	0.63	פריט 10	פריט 9
0.75	0.41	0.72	0.25	0.69	0.30	0.78	0.39	0.74	0.38	0.87	0.59	פריט 13	פריט 10
0.76	0.28	0.71	0.39	0.68	0.34	0.77	0.49	0.75	0.37	0.87	0.53	פריט 20	פריט 11
0.74	0.46	0.70	0.43	0.67	0.44	0.79	0.35	0.75	0.22	0.87	0.56	פריט 11	פריט 12
0.74	0.50	0.70	0.42	0.66	0.48	0.78	0.39	0.75	0.23	0.87	0.59	פריט 16	פריט 13
0.74	0.45	0.70	0.41	0.68	0.34	0.78	0.36	0.74	0.47	0.87	0.50	פריט 15	פריט 14
0.76	0.30	0.71	0.37	0.69	0.29	0.79	0.31	0.76		0.87	0.50	פריט 14	פריט 15
0.75	0.40	0.70	0.43	0.68	0.31	0.79	0.32	0.75	0.34	0.87	0.43	פריט 18	פריט 16
0.74	0.46	0.71	0.39	0.69	0.23	0.79	0.14	0.76		0.87	0.39	פריט 22	פריט 17
0.76	0.27	0.71	0.31	0.70	0.17	0.80	0.05	0.76		0.87	0.29	פריט 21	פריט 18
0.76	0.10	0.73	0.10	0.70	0.13	0.80		0.76		0.88	0.12	פריט 24	פריט 19
0.76	0.06	0.73	0.14	0.70		0.80		0.76		0.88	0.09	פריט 23	פריט 20

תת-מבחן 1ד' בעברית – נוסח סופי (20 פריטים)
התפלגות הציונים* בקרב כל התלמידים ובקרב כל קבוצת גיל



תת-מבחן 1ד' בעברית – ממוצעים בכל קבוצת גיל
במבחן השלם ובשלושה פריטים לדוגמה



הנתונים מתייחסים לתוצאות המבחן המקורי, לפני פסילת פריטים.
פריטים 2 ו-24 המופיעים בגרף נפסלו ואינם מופיעים במבחן הסופי.

תת-מבחן 1ד' בעברית – נוסח סופי (20 פריטים)
הבדלים לפי מגדר בגרסה לדוברי עברית

מספר פריט בנוסח הסופי	מספר פריט במחקר החלוץ	בנים (N=475)		בנות (N=536)		סה"כ (N=1011)		בנים פחות בנות (מתוקנן)**
		ס"ת	ממוצע*	ס"ת	ממוצע	ס"ת	ממוצע	
סה"כ	סה"כ	4.31	10.07	4.16	10.50	4.24	10.30	-0.10
פריט 1	פריט 1	0.21	0.95	0.17	0.97	0.19	0.96	-0.11
פריט 2	פריט 3	0.34	0.87	0.32	0.88	0.33	0.88	-0.05
פריט 3	פריט 4	0.27	0.92	0.27	0.92	0.27	0.92	-0.01
פריט 4	פריט 5	0.35	0.86	0.34	0.87	0.34	0.86	-0.04
פריט 5	פריט 6	0.39	0.81	0.40	0.80	0.40	0.80	0.02
פריט 6	פריט 8	0.41	0.79	0.36	0.85	0.39	0.82	-0.15
פריט 7	פריט 7	0.42	0.77	0.39	0.81	0.41	0.79	-0.11
פריט 8	פריט 9	0.46	0.69	0.45	0.72	0.46	0.70	-0.07
פריט 9	פריט 10	0.50	0.57	0.48	0.65	0.49	0.61	-0.17
פריט 10	פריט 13	0.49	0.59	0.50	0.57	0.49	0.58	0.03
פריט 11	פריט 20	0.50	0.45	0.50	0.51	0.50	0.49	-0.12
פריט 12	פריט 11	0.49	0.40	0.49	0.40	0.49	0.40	0.00
פריט 13	פריט 16	0.47	0.33	0.49	0.40	0.48	0.37	-0.16
פריט 14	פריט 15	0.48	0.37	0.49	0.40	0.49	0.38	-0.06
פריט 15	פריט 14	0.45	0.28	0.47	0.34	0.46	0.31	-0.12
פריט 16	פריט 18	0.41	0.21	0.42	0.23	0.41	0.22	-0.05
פריט 17	פריט 22	0.34	0.13	0.32	0.11	0.33	0.12	0.06
פריט 18	פריט 21	0.27	0.08	0.22	0.05	0.25	0.06	0.12
פריט 19	פריט 24	0.11	0.01	0.11	0.01	0.11	0.01	0.00
פריט 20	פריט 23	0.06	0.00	0.06	0.00	0.06	0.00	0.01

* הניקוד לכל פריט: 1 לתשובה נכונה, 0 לתשובה שגויה או לאי-השבה.

הניקוד לכל המבחן: 20 נקודות – נקודה לכל תשובה נכונה

** הציון המתוקנן הוא הפער בין הממוצעים בפריט לבנים לעומת בנות, מחולק בסטיית התקן הכללית בפריט.

ראו גרף להלן, להצגה חזותית של נתונים אלה.

תת-מבחן 1ד' בעברית – נוסח סופי (20 פריטים)
הבדלים לפי מגזר

מספר פריט בנוסח הסופי	מספר פריט במחקר החלוצי	חרדי (N=132)		ממלכתי וממ"ד (N=879)		סה"כ (N=1011)		חרדי פחות ממלכתי וממ"ד (מתוקנן)**
		ס"ת	ממוצע*	ס"ת	ממוצע	ס"ת	ממוצע	
סה"כ	סה"כ	4.11	9.40	4.24	10.43	4.24	10.30	-0.24
פריט 1	פריט 1	0.21	0.95	0.19	0.96	0.19	0.96	-0.05
פריט 2	פריט 3	0.35	0.86	0.33	0.88	0.33	0.88	-0.07
פריט 3	פריט 4	0.30	0.90	0.26	0.93	0.27	0.92	-0.09
פריט 4	פריט 5	0.35	0.86	0.34	0.87	0.34	0.86	-0.03
פריט 5	פריט 6	0.41	0.79	0.40	0.81	0.40	0.80	-0.04
פריט 6	פריט 8	0.42	0.77	0.38	0.82	0.39	0.82	-0.13
פריט 7	פריט 7	0.44	0.73	0.40	0.80	0.41	0.79	-0.16
פריט 8	פריט 9	0.47	0.68	0.46	0.71	0.46	0.70	-0.06
פריט 9	פריט 10	0.50	0.52	0.48	0.63	0.49	0.61	-0.21
פריט 10	פריט 13	0.50	0.50	0.49	0.59	0.49	0.58	-0.18
פריט 11	פריט 20	0.47	0.33	0.50	0.51	0.50	0.49	-0.35
פריט 12	פריט 11	0.46	0.30	0.49	0.41	0.49	0.40	-0.24
פריט 13	פריט 16	0.45	0.28	0.49	0.38	0.48	0.37	-0.21
פריט 14	פריט 15	0.47	0.32	0.49	0.39	0.49	0.38	-0.16
פריט 15	פריט 14	0.46	0.30	0.46	0.31	0.46	0.31	-0.04
פריט 16	פריט 18	0.39	0.19	0.42	0.22	0.41	0.22	-0.08
פריט 17	פריט 22	0.28	0.08	0.33	0.13	0.33	0.12	-0.13
פריט 18	פריט 21	0.15	0.02	0.26	0.07	0.25	0.06	-0.19
פריט 19	פריט 24	0.12	0.02	0.11	0.01	0.11	0.01	0.02
פריט 20	פריט 23	0.00	0.00	0.07	0.00	0.06	0.00	-0.07

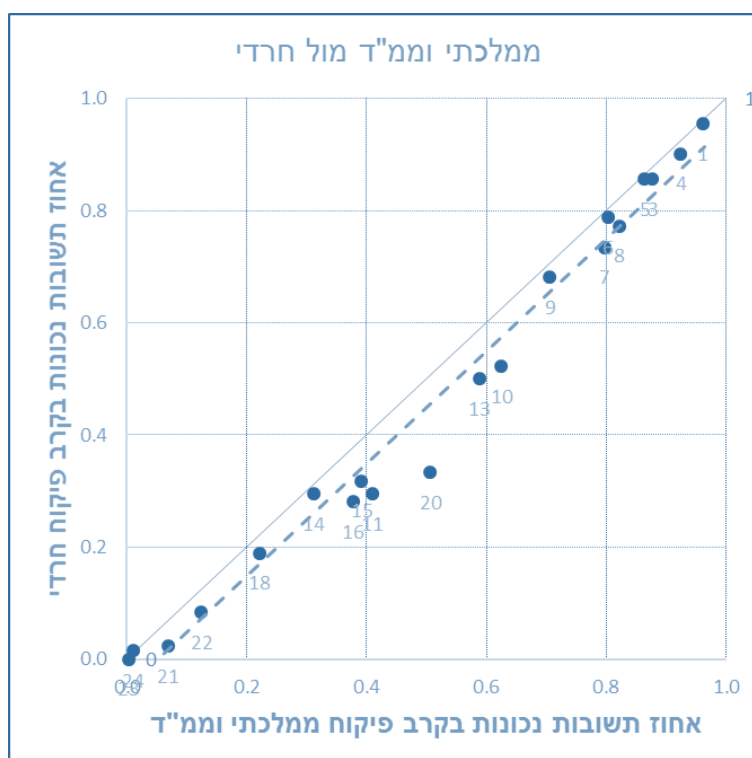
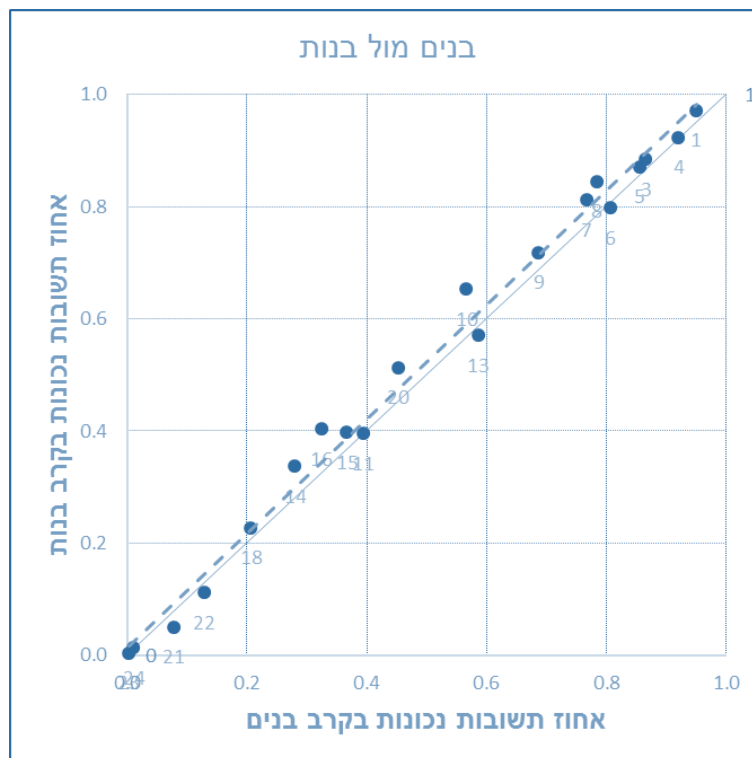
* הניקוד לכל פריט: 1 לתשובה נכונה, 0 לתשובה שגויה או לא-השבה.

הניקוד לכל המבחן: 20 נקודות – נקודה לכל תשובה נכונה

** הציון המתוקנן הוא הפער בין הממוצעים בפריט לממלכתי וממ"ד לעומת חרדי, מחולק בסטיית התקן הכללית בפריט.

ראו גרף להלן, להצגה חזותית של נתונים אלה.

תת-מבחן 1ד' בעברית – נוסח סופי (20 פריטים)
הבדלים לפי מגדר ולפי מגזר



כל פריט מיוצג בנקודה ומתחתיה מספר הפריט. המספור הוא לפי מספר הפריט במחקר החלוץ.

נספח 4 – קרדיטים: רשימת השותפים המלאה

ראמ"ה

הנהלה – אחריות כוללת

פרופ' מיכל בלר

ד"ר חגית גליקמן

ד"ר יואל רפ

ניהול הפרויקט

ד"ר עינת נוטע-קורן

אימאן עואדיה

תרגום והתאמה

אלה שפירא

ד"ר פאדי כנעאן

הדיל שחברי

בועז רוזנבאום

דגימה ועיבוד נתונים

נתנאל גולדשמידט

ד"ר פרנקלין זרומב

ריכוז, תכנון וניהול לוגיסטי וכלכלי

עופר טל שחר

מרינה אינדזליכין

כרמית ויינשטיין

משרד החינוך

ד"ר חוה פרידמן

ועדה מתודולוגית מיעצת (לפי סדר א'-ב')

פרופ' רות בייט מרום

ד"ר ענת בן סימון

ד"ר יואב כהן

ד"ר שרונה מי-טל

פרופ' פאדיה נאסר

פרופ' ברוך נבו

פרופ' יוסי צלגוב

מאשה שינמן

מנכ"לית ראמ"ה (עד 2013)

מנכ"לית ראמ"ה (מ-2013 עד 2022)

מנהל אגף המבחנים, ראמ"ה (עד 2018)

מנהלת תחום מבחנים ארציים, ניהול הפרויקט עד 2017

מנהלת תחום מבחנים בערבית, ניהול הפרויקט מ-2017

מנהלת גף פסיכומטריקה (מ-2015)

מנהל גף פיתוח כלי מדידה בערבית (עד 2018)

מרכזת בכירה פיתוח כלי מדידה בערבית (מ-2019)

מנהל תחום מבחנים ממוחשבים

מנהל תחום סטטיסטיקה

מנהל תחום סטטיסטיקה

מנהל אגף א' (משאבי אנוש, תקציב ולוגיסטיקה)

כלכלנית

רכזת תחום מבחנים ארציים

מנהלת אגף פסיכולוגיה (שפ"י), משרד החינוך

האוניברסיטה הפתוחה

המרכז הארצי לבחינות ולהערכה

המרכז הארצי לבחינות ולהערכה

שפ"י, משרד החינוך

אוניברסיטת תל אביב

אוניברסיטת חיפה

אוניברסיטת בן-גוריון

שפ"י, משרד החינוך

ועדת מומחים מלווה למבחנים השפתיים (לפי סדר א'-ב')

ד"ר ענת בן סימון	המרכז הארצי לבחינות ולהערכה
ד"ר עמליה בראון	אוניברסיטת תל אביב
כמיל נג'אר	אגף שירות פסיכולוגי ייעוצי, משרד החינוך
פרופ' אלינור סאיג חדד	אוניברסיטת בר-אילן
חגית שניידר	יועצת, ראמ"ה

צוות מטח (המרכז לטכנולוגיה חינוכית)

תרגום והתאמה של סוללת המבחנים לעברית ולערבית, ביצוע מחקר החלוץ וגיבוש המבחנים הסופיים

הנהלה – אחריות כוללת

ד"ר טלי פרוינד	מנהלת מינהלת הערכה ומדידה (עד 2017)
עוזי סגיר	מנהל פיתוח מבחנים

ניהול

ד"ר שם שמי	מנהל הפרויקט וראש צוות עברית
------------	------------------------------

צוות פיתוח ערבית

ד"ר פאדי כנעאן	ראש צוות ערבית ופסיכומטריקאי (עד 2013)
הדיל שחברי	ראשת צוות ערבית ופסיכומטריקאית (מ-2013)
ד"ר נרימאן הינדי	פסיכו-לינגוויסטיקאית
מסבאח טאטור	פסיכולוג חינוכי
ראדי מוסא	פסיכולוג חינוכי
יארם כריים	רכזת ופסיכומטריקאית
אמירה סעיד	רכזת

צוות פיתוח עברית

ד"ר גלית בן-צבי	פסיכו-לינגוויסטיקאית
סמדר ספיר-יוגב	פסיכולוגית חינוכית
רשי קור	פסיכולוג חינוכי
אלה שפירא	פסיכומטריקאית (עד 2015)
נדב כספי	פסיכומטריקאי (מ-2015)

פסיכומטריקה ועיבוד נתונים

אסטלה מלמד	פסיכומטריקאית ראשית
ד"ר בעז צור	חוקר ופסיכומטריקאי

עריכת לשון

מיכל אילון	עורכת לשון (עברית)
עלאא' שלעביה	עורך לשון (ערבית)

עיצוב גרפי

לילך חזאי

מנהלת הפקה ועיצוב גרפי

צוות ההפקה: רהאדה ערפאת, גלדיס חיפאוו, שני גואטה, אולה נסטרובה, אורנה אזולאי, אנג'ליק נעים ז"ל, ירמי לז'ה (איורים)

שותפים למחקר החלוץ:

דרור הראל (מנהל לוגיסטי במחקר החלוץ), אוהד מנשרוב (מעריך בכיר), פרנסין עיסא (רכזת ופסיכומטריקאית), יארא כריים (רכזת), חן יפרח (רכזת), מאיה לוז (רכזת)

צוות טלדור

תפעול מחקר הנורמות

הנהלה – אחריות כוללת

שי גואטה

סגן מנהל חטיבת BPO

ניהול

אלינור טבע

מנהלת פרויקט

עידו לב

מנהל מרכז תיאום

ד"ר לירון גרינברג

פסיכולוגיה חינוכית מומחית – מנהלת תוכן

שותפים למחקר הנורמות – הדרכה והערכה בכירה (לפי סדר הא'-ב')

שירלי אהרונובסקי, סירין בראנסה, תמר ברליון, נופר גבע, יארא ג'ירייס, דוניה דאהר, נימרה וטרמן, שיימא חאג' יחיא, ויליט חורי, ג'נאן טנוס, ראדי מוסא, שרה מילר, טניה עזאם, נועה קוסובסקי, דניאל שמש

צוות מאל"ו (המרכז הארצי לבחינות ולהערכה)

פיתוח סולמות הציונים והנורמות

ד"ר ענת בן סימון

מנכ"לית מאל"ו (עד 2023)

ד"ר נעמי גפני

מנהלת פרויקט (עד 2019)

ד"ר יואל רפ

מנהל פרויקט (מ-2019), מנכ"ל מאלו (מ-2023)

ד"ר דביר קלפר

מתמטיקאי, חוקר בכיר

אסטלה מלמד

פסיכומטריקאית

נתנאל סילברשטיין

חוקר בכיר ומנהל היחידה

נועם קשת

חוקר

איתמר כהן

חוקר

תודה למדריכים בסדנאות ההכשרה לבוחנים. תודה למאות הפסיכולוגים, המתמחים והסטודנטים שעמלו והעבירו במסירות ובתחושת שליחות את המבחנים לתלמידים במחקר החלוץ ובמחקר הנורמות. תודה מיוחדת לאלפי התלמידים שהשתתפו במחקר החלוץ ובמחקר הנורמות.

הכנת הדוח:

כתיבה

ד"ר שם שמי

אלה שפירא

ד"ר יואל רפ

ד"ר פאדי כנעאן

הכנת נתונים להצגה בדוח

אסטלה מלמד

ד"ר בועז צור

ריוויז

ד"ר עינת נוטע-קורן

אימאן עואדיה

הדיל שחברי

עריכת לשון

מיכל אילון

מנהל פרויקט וודקוק ג'ונסון, מטח

מנהלת גף פסיכומטריקה, ראמ"ה

מנהל אגף המבחנים, ראמ"ה (עד 2018)

מנהל גף פיתוח כלי מדידה בערבית, ראמ"ה (עד

2018)

פסיכומטריקאית ראשית, מטח

חוקר ופסיכומטריקאי, מטח

מנהלת תחום מבחנים ארציים, ראמ"ה (עד 2017)

מנהלת תחום מבחנים בערבית, ראמ"ה

מרכזת בכירה פיתוח כלי מדידה בערבית, ראמ"ה

עורכת לשון, מטח

- Carroll, J. B. (1993). Human cognitive abilities: A survey of factoranalytic studies. Cambridge,UK: Cambridge University Press.
- Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 122–130). New York, NY: Guilford Press.
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In: H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–22). San Diego: Pergamon.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 69–76). New York, NY: Guilford Press.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- Cattell, R. B. (1957). *Personality and motivation structure and measurement*. New York, NY: World Book.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology*, 54(1), 1.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.
- Ferguson, C. (1959). *Diglossia* Word, 15, 325–340.
- Gudmundsson, E. (2009) Guidelines for translating and adapting psychological instruments. *Nordic Psychology*, 61:2, 29–45.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3–38). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R.K. and Lee, M. K., (2013). Methods for translating and adapting tests to increase Cross-Language Validity. In Saklofske, D. H, Reynolds, C.R. and Schwean, V. (Eds.), *The Oxford Handbook of Child Psychological Assessment*, (pp. 171–182). Oxford University Press.

- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45, 153-171.
- Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology*, 1, 1-30.
- Holes, C. (2004). Modern arabic: structures, functions, and varieties. *Washington, DC: Georgetown University Press*.
- Horn, J. L. (1965). Fluid and crystallized intelligence: A factor analytic and developmental study of the structure among primary mental abilities. Unpublished doctoral dissertation, University of Illinois, Champaign.
- Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 75, 242-259.
- Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (Rev. ed., pp. 645-685). New York, NY: Academic Press.
- Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual* (pp. 197-232). Rolling Meadows, IL: Riverside Publishing.
- International Test Commission (2005). International Guidelines on Test Adaptation. [www.intestcom.org]
- International Test Commission (2010). International Test Commission Guidelines for Translating and Adapting Tests. [<http://www.intestcom.org>]
- Malda, M., Van de Vijver, F. J. R., Srinivasan, K., Transler, C., & Sukumar, P. (2008). Adapting a cognitive test for a different culture: An illustration of qualitative procedures.
- Maamouri, M. (1998). Language education and human development: Arabic diglossia and its impact on the quality of education in the arab region. *Discussion Paper prepared for the World Bank. The Mediterranean Development Forum*.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136-181). New York: The Guilford Press.
- McGrew, K. S., & Woodcock, R. W. (2001). Technical Manual: *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). Technical Manual. *Woodcock-Johnson III Normative Update*. Rolling Meadows, IL: Riverside Publishing.

Mother, N., & Woodcock, R. W. (2001). Examiner's Manual. *Woodcock-Johnson III Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside Publishing.

Saiegh-Haddad, E. (2003). Linguistic distance and initial reading acquisition: The case of Arabic diglossia. *Applied Psycholinguistics*, 24, 431–451.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

לקריאה נוספת בנושא ניתוח גורמים

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999) Psychological research. *Psychological Methods*, 4 (3), 272–299.

Stapleton, C.D. (1997). Basic concepts in exploratory factor analysis (EFA) as a tool to evaluate score validity. *Texas A&M University*.

Child, D. (2006). *Essentials of factor analysis* (3rd ed). Continuum, New York.