



סטנדרטים לעריכת מבחנים בחינוך ובפסיכולוגיה- הצעה לתקנון ישראלי

ברוך נבו
בסיוע: מירי איתן
נעמי שטרך

תוכן העניינים

מבוא

חלק I: סטנדרטים טכניים לבניית מבחן והערכתו

1. פיתוח מבחנים, עריכת תיקונים וחידושים
רקע
סטנדרטים 1.1 - 1.23
2. מהימנות וטעויות מדידה
רקע
סטנדרטים 2.1 - 2.10
1. תקפות
רקע
סטנדרטים 3.1 - 3.22
4. סילוס, יצירת נורמות, יכולת השוואה בין תוצאות, וכיול
רקע
סטנדרטים 4.1 - 4.7
5. פרסומי מבחנים: מדריכים טכניים ומדריכים למשתמשים
רקע
סטנדרטים 5.1 - 5.9

חלק II : סטנדרטים מקצועיים לשימוש במבחנים

6. עקרונות כלליים לשימוש במבחנים

רקע

סטנדרטים 6.1 - 6.13

7. מבחנים קליניים

רקע

סטנדרטים 7.1 - 7.5

8. מבחנים חינוכיים ופסיכולוגיים בבתי-ספר

רקע

סטנדרטים 8.1 - 8.9

9. שימוש במבחנים לצורכי ייעוץ

רקע

סטנדרטים 9.1 - 9.5

10. מבחנים תעסוקתיים

רקע

סטנדרטים 10.1 - 10.7

חלק III: סטנדרטים ליישומים מיוחדים

11. מתן רשיונות ותעודות מקצועיות ותעסוקתיות

רקע

סטנדרטים 11.1 - 11.4

12. עריכת מבחנים למיעוטים לשוניים

רקע

סטנדרטים 12.1 - 12.5

13. עריכת מבחנים לאנשים מוגבלים

רקע

סטנדרטים 13.1 - 13.7

חלק IV: סטנדרטים להליכים מנהלתיים

14. העברת מבחן, ציינון ודיווח

רקע

סטנדרטים 14.1 - 14.10

15. הגנה על זכויות הנבחנים

רקע

סטנדרטים 15.1 - 15.7

מילון

מ ב ר א

מסמך זה מהווה הצעה לתקינה ישראלית בתחום המבחנים. מטרת המסמך היא לשמש בסיס לדיונים אשר יביאו בסופו של דבר ליצירתו של קובץ נהלים מקובל על מועצת הפסיכולוגים והסתדרות הפסיכולוגים.

במקור להשראה וחיקוי להצעה זו שימש המדריך האמריקאי: Standards for Educational and Psychological Testing אשר בגירסתו האחרונה (החמישית) פורסם ב-1985 במשותף על ידי שלושה אירגונים:

American Educational Research Association

American Psychological Association

National Council on Measurement in Education

יו"ר הוועדה המכינה ועורך המסמך האמריקאי היה M.R. Novick.

יעדי פירסום המסמך הנ"ל היו:

(1) לספק בסיס להערכת איכותם של מבחנים ולתת הנחיות לגבי שימוש נכון בהם, כל זאת לאחר שהם הוצאו לאור.

(2) לפרט המלצות הקשורות בפיתוח המבחן טרם פירסומו והוצאתו לאור. אלה הם גם היעדים של ההצעה הישראלית.

המסמך שלפנינו שונה מהמקור האמריקאי, בנקודות הבאות:

א. בשלב זה לא נעשתה אבחנה בין סטנדרטים "ראשוניים", "משניים" ו"מוֹתָנִים" (כפי שנעשה הדבר בארצות הברית). התנאים המיוחדים לישראל אינם מאפשרים העתקה מדויקת של סדר החשיבות והקדמויות האמריקאי. תפקיד זה של אבחנה בין סטנדרטים הכרחיים לרצויים (או מידיים ונדחים) עדיף להשאיר לוועדות הישראליות שתקיימנה ציונים סביב המסמך.

ב. בעת עריכת ההצעה שלפנינו הובאו בחשבון ביקורות שונות שהופיעו בכתבי עת מקצועיים ב-1986 ו-1987 ואשר קראו לתיקונים אחדים במסמך האמריקאי.

ג. ההצעה שלפנינו קצרה בהרבה מ"סטנדרטים" האמריקאי, מתומצתת ומגובשת יותר. מה שאיפשר את השינוי הזה היה אימוצו של עיקרון אשר אמר כי קהל היעד למסמך בישראל יהיה ציבור הפסיכולוגים המוסמכים והיועצים המוסמכים. ציבור זה הוא בעל ידע מקצועי רב ואין לו צורך בהסברים טרומיים, בשינון עקרונות מובנים מאליהם ובחזרות מודגשת.

יחד עם זאת, כאמור, חלקים נרחבים ממסמך זה כמו גם התפיסה הבסיסית, הושאלו מ"סטנדרטים" האמריקאי.

מבחנים ושימושים במבחנים, עליהם חלים סטנדרטים אלה

המונח "מבחנים" כולל כלים מתוקננים להערכת כשרים, יכולת והישגים, אמצעים להערכה ואבחון, רשימות של נטיות, שאלוני אישיות וכלים השלכתיים. המונח "מבחן" מתייחס בדרך כלל למדידות של דגימות התנהגותיות מובנות או מתוכננות, בהם מצופה מהנבחנים, או נאמר להם, לעשות כמיטב יכולתם. כלים לזיהוי נטיות ומאפייני אישיות באמצעות דיווח עצמי, מכונים באורח טיפוסי, ובצדק, "רשימות", "שאלונים" או "רשימות בדיקה", ולא "מבחנים". הוא הדין לגבי דגימות מובנות של התנהגות נצפית. אך כאן, יקראו גם כלים אלה של דיווח עצמי ושל תצפית בשם "מבחנים". הסטנדרטים חלים גם עליהם.

מבנה מסמך ה"תקינה"

חלק מס' 1 של "תקינה": "סטנדרטים טכניים לבניית מבחן והערכתו", כולל סטנדרטים לתקפות, מהימנות, פיתוח מבחנים, דירוג, קביעת נורמות, כיוול ופרסום. חלק מס' 2 מציג "סטנדרטים מקצועיים לשימוש במבחנים". חלק מס' 3, "סטנדרטים ליישומים מיוחדים", כולל סטנדרטים לעריכת מבחנים למיעוטים לשוניים ולאנשים הסובלים מנכות לסוגיה. הסטנדרטים הטכניים והמקצועיים הנידונים בחלקים 1 ו-2, רלוונטים ליישומים המסויימים המופיעים בחלק מס' 3. חלק מס' 4, "סטנדרטים להליכים מינהלתיים", מכיל סטנדרטים הנוגעים להעברת המבחן, לבדיקה ודיווח, וכן סטנדרטים להגנה על זכויות הנבחנים. למרות שכל ארבעת החלקים דנים בנושאים טכניים, הרי שכל אחד מהם מציין כיצד משתלבים הנושאים הטכניים, הביצוע המקצועי והאתיקה המקצועית.

כל פרק פותח בהקדמה, אשר מספקת רקע לסטנדרטים המופיעים בהמשכו. מרבית הסטנדרטים מלווים בהערות. הערות אלה מכוונות להסביר את הסטנדרטים אליהם הן מתלוות. "תקינה" מכיל גם מילון. המילון מספק הגדרות למונחים, כפי שהם מופיעים בכרך זה באופן ספציפי.

חלק ראשון

סטנדרטים טכניים לבניית מבחן והערכתו

Technical Standards for Test Construction and Evaluation

1. פיתוח מבחנים, עריכת תיקונים וחיידושים
(Test development and revision)

רקע

פרק זה דן בנושאים כלליים הנוגעים לפיתוח מבחנים. על מפתחי המבחן מוטלת האחריות לספק עדות למהימנות ותקפות באשר למטרות המבחן המוצהרות, כמו גם לספק מדריכים (manuals) ונורמות בעת הצורך, על מנת לכוון לפירושים נאותים.

על מפתחי המבחנים לערוך מחקר אשר יסייע להבחין בין שימוש נאות לשימוש בלתי נאות ולעצב את המבחנים באורח שיסייע לשימוש נכון במבחן. על המפתח, המו"ל או נותן החסות למבחן, מוטלת האחריות להציג את המידע בצורה נגישה ונחה לשימוש, בלויית תקצירים ופרשנויות, על מנת להקל על בדיקת איכות המבחן והערכתו.

הסטנדרטים בפרק זה דנים במפרטי מבחן ומפרטי פריטים, ניתוח פריטים והליכי מיון, וכמו כן, בהערכת מבנה של מבחנים לשימושים המיועדים להם. מוצגים מספר סטנדרטים מיוחדים, הישימים למבחנים מסוגים מסוימים, לרבות מבחנים ממוחשבים.

סטנדרט 1.1 מבחנים ומערכות מבחנים, יש לפתח על בסיס מדעי איתן ומוצק. על מפתחי מבחנים לנהל כל מחקר שידרש, לצבור עדויות הנוגעות למבחן, לקבוע איזה מידע נדרש טרם פרסום המבחן והפצתו, ואיזה מידע ניתן לספק מאוחר יותר.

סטנדרט 1.2 יש לציין בבירור את המפרטים המשמשים לבניית פריטים או בחירת תצפיות, ובעיצוב המבחן ככלל. יש לתאר את הגדרת התחום המשמש לבניית או בחירת הפריטים. גם כאשר - מסיבות של בטיחות - קשה לשחרר דוגמאות של המבחן לבדיקה, על המידע התיאורי לכלול בכל זאת פריט מייצג אחד לפחות, המזוהה עם כל אחד מהמרכיבים העיקריים לסיווג או הגדרת התוכן. כאשר קושי הפריט הוא אחד ההיבטים של מערכת כזו, יש לספק פריטים המייצגים את רמות הקושי השונות.

הערה: לעיתים, בוני המבחן מציינים כי המבחן הוא תלוי-קריטריון, בניגוד למבחן תלוי-נורמה. נוהג זה גרם לבלבול מה. בפירושים תלוי-נורמה, הציון משווה עם התפלגות ציוניהם של פרטים או קבוצות אחרות. בפירושים תלוי-קריטריון, הציון משקף באורח ישיר רמת כשירות (competency) בתחום-קריטריון מוגדר. יש לזכור כי פירוש ציוני המבחן - ולא פירוש המבחן עצמו - הוא תלוי-הנורמה או תלוי-הקריטריון. על כן, ניתן לפרש כמה מבחנים תלוי-נורמה בדרכים שהן תלויות-קריטריון, ולהיפך. איכותם ויעילותם של פירושים תלוי-קריטריון, תלויים במידה בה הוגדר כדבעי התחום ההתנהגותי המיוצג ע"י המבחן.

סטנדרט 1.3 על הגדרות התחום ומפרטי המבחן להיות ברורים במידה מספקת, כך שמומחים בעלי ידע יוכלו לשפוט את היחס בין הפריטים לבין התחום שהם אמורים לייצג.

סטנדרט 1.4 כאשר פריטי מבחנים מתייחסים לקורס לימודים או קורס הכשרתי, תוכנית לימודים, או ספר לימוד, על המדריך או דו"חות אחרים לכלול את זהותו ותיאורו של הקורס או החומר הנלמד.

סטנדרט 1.5 בעת בחירת סוג הפריטים ותוכנם, עבור מבחנים ושאלונים, על מפתח המבחן לבחון את תוכן וסוג הפריטים ביחס לרקע התרבותי וההתנסויות הקודמות של מגוון הקבוצות האתניות, התרבותיות, קבוצות המין והגיל הקיימות באוכלוסיית הנבחנים לה מיועד המבחן.

סטנדרט 1.6 מקום בו מתאמי תוצאות פריטים עם תוצאות קריטריון - אשר הינם חיצוניים למבחן הנידון - משמשים על מנת לבחור פריטים או לבנות מפתחות לחישוב תוצאות, על תקפות הפריטים להיות מלווה בעדות להצלבת-תקפות של הסולם.

סטנדרט 1.7 מקום בו משמשים אומדים פרמטריים של עקומות תגובה לפריט, בעת פיתוח המבחן, יש לציין בבירור מהו הדגם לתגובה לפריט (item response model), ואת הליכי הכיול. יש להציג מידע לגבי המידה בה אומתו הנחות חשובות של תיאורית התגובה לפריט (לדוגמה - חד-מימדיות, או שיוויון בין פרמטרי שיפוע), על מנת להבהיר את מידת התאמתם של הנתונים לדגם. יש לתאר את המדגמים ששימשו לאמידת פרמטרי פריטים. על המדגמים להיות מגוונים ובגודל נאות.

הערה: למרות העובדה, כי רבה חשיבותו של גודל המדגם הכללי, חשוב גם שיהיה מספר נאות של מקרים המצויים באיזורים שהם הקריטיים, לקביעת הפרמטרים של פריט.

סטנדרט 1.8 יש לתאר במדריך המבחן של מבחנים אדפטיביים, את הרציונל והעדות התומכת בהליכים ששימשו לבחירת פריטים להפעלה, להפסקת המבחן, ולחישוב תוצאות המבחן.

סטנדרט 1.9 על מפתחי המבחן להיות ערים לצורך בעריכת מחקרים בנוגע להבדלים בביצועי פריטים או מבחנים, עבור קבוצות גיל, מוצא אתני, רקע תרבותי או מין, הנכללות באוכלוסיית הנבחנים.

סטנדרט 1.10 מקום בו נתגלה, כי אסטרטגיות של ביצוע מבחן (כגון ניחושים, דילוג על פריטים קשים וכו'), אשר אינן קשורות למבנים או לתכנים הנמדדים, משפיעות באורח ניכר על ביצועים במבחן, יש להסביר לנבחנים אסטרטגיות אלה לפני הפעלת המבחן, בצורת חוברת מידע, או - במידה וניתן להציג את ההסבר בקצרה - בד בבד עם מתן ההוראות למבחן. יש לעודד את כל הנבחנים להשתמש באסטרטגיות אלה, במידה והשפעותיהן משפרות את הביצועים; מאידך, יש להניאם משימוש באסטרטגיות כאלה, במידה והשפעותיהן מחבלות בביצועים.

סטנדרט 1.11 בעבור מבחנים המציבים מגבלות זמן חמורות, יש לבדוק, בעת פיתוח המבחן, את המידה בה התוצאות כוללות את מרכיב המהירות, ולהעריך את מידת הרלוונטיות של מרכיב כזה, למבנים או לתוכן שהמבחן מיועד למדוד.

סטנדרט 1.12 כחלק מתהליך הפיתוח, יש לבחון את רגישות ביצוע המבחן, לשיפורים החלים בעקבות תרגול, אימון או הדרכה קצרה, בעיקר כאשר מדובר במבחני ביצוע הכוללים אופני תגובה בלתי מוקרים, כגון מבחנים המופעלים באמצעות מחשב. מבחן המעוצב כך שלא יושפע מצורות לימוד אלו - יש להראות כי אכן אינן מושפע מהן. המדריך יסכם עדויות שהופקו ממחקרים כאלה.

סטנדרט 1.13 עבור מדדי נטיות או מדדי אישיות, המיועדים למטרות ברירה או שיבוץ, יש להציג עדות למידה בה התוצאות חשופות לנסיונות של הנבחנים להציג תמונה מזויפת של עצמם, (חיובית או שלילית).

סטנדרט 1.14 על טפסי הדיווח (לפרטים הנבחנים ולמוסדות משתמשים) הדנים בציונים, וחומר הדרכה למבחן, לרבות דו"חות וחומרים ממוחשבים, להיות כאלה שיסייעו לפירוש נאות.

סטנדרט 1.15 במידה ומכינים נוסח מקוצר של המבחן, באמצעות הפחתת מספר הפריטים או אירגון מחדש של חלקי המבחן, יש לספק נתונים אמפיריים או רציונל תיאורטי, על מנת לאמוד את מהימנותו של כל נוסח מקוצר, ואת המתאם בינו לבין הנוסח הסטנדרטי.

סטנדרט 1.16 על מפתחי המבחן מוטלת האחריות להיות ערים לתנאים משתנים; הם צריכים לתקן, לחדש או למנוע שימוש במבחן, כאשר מתקבלים נתוני מחקר חדשים, או הלים שינויים ניכרים בעולם התוכן המיוצג, או נוצרים תנאים חדשים לשימוש במבחן.

סטנדרט 1.17 אין לתת למבחן את הכותרת "מחודש" או לפרסמו ככזה, אלא אם כן עבר המבחן תהליך חידוש עקרוני וחשוב. יש להשתמש במשפט כגון "עם שינויים משניים" כאשר עבר המבחן שינויים קלים ומשניים. יש לערוך התאמות בסולם הציונים בעקבות שינויים אלה.

סטנדרט 1.18 במידה והמבחן, או חלק ממנו, מיועדים למטרות מחקר בלבד, ואינם מופצים לשימוש אופרטיבי, יש להציג עובדה זו בבירור בכל מסמך.

סטנדרט 1.19 יש להציג את ההוראות להעברת המבחן עם ההדגשות הנחוצות ובבהירות מספקת, כך שמשמשים יוכלו להגיע בקירוב לתנאי ההעברה בהם הושגו הנורמות והנתונים לגבי מהימנות ותוקף. יש להדגיש את הצורך בהיצמדות להוראות.

סטנדרט 1.20 על ההוראות הניתנות לנבחן להיות מפורטות במידה נאותה, כך שהנבחן יוכל להגיב למטלות בצורה אליה התכוון מפתח המבחן. בעת הצורך, יש לספק דוגמאות ואפשרויות תרגול.

הערה: לדוגמה, יתכן ובעת שימוש במבחן אישיות, הכוונה תהיה כי הנבחן יספק את התגובה הראשונה שעולה על דעתו. יש להבהיר ציפיה זו בהוראות המתלוות לרשימה. בהוראות לשאלוני נטיות, יתכן ויהא זה חשוב לציין אם הנבחנים אמורים לסמן את הפעילויות שהם מעדיפים באופן אידאלי, או אם הם אמורים לשקול הן את ההזדמנויות והן את היכולת בצורה ריאליסטית.

טבעו וכמותו של חומר התרגול, תלוי ברמות הידע הנצפות מהנבחנים. יתכן, לדוגמה, כי יהיה זה חשוב מאד לספק תרגול לפני שהנבחנים ניגשים למבחן בעל פורמט חדש. בעת הצורך, על ההוראות לתאר בבהירות נושאים כגון ניחוש ומגבלות זמן. במבחני הישג, יש לציין אם הנבחנים יכולים להיעזר ברשימות, ספרים, מחשבי-כיס. על המו"לים לכלול הדרכה להתמודדות עם שאלות טיפוסיות של נבחנים.

סטנדרט 1.21 מקום בו נאספות דוגמאות התנהגותיות מובנות בפורמט מבחן מתוקנן, יש להגדיר בבירור את הסוג הספציפי של ההתנהגות הנצפית. הוראות הניתנות לנבחן, המיועדות להפקת דוגמה התנהגותית מסוימת (לעיתים קרובות קרויה "תגובה מיידית" (prompt) יש להבנות, כפי שיש להבנות כל הוראה עבור כל מבחן. יש לבחון מראש את ההוראות המבחן.

סטנדרט 1.22 על מפתח המבחן להציג הליכים לחישוב ציונים באופן מקומי, (locally), בפירוט משביע רצון ובבהירות, על מנת להביא את הדיוק בחישוב הציונים לרמה מירבית. על ההוראות לדירוג סולמות או הוראות הנוגעות לציונים המתקבלים מקידוד, שקלול או סיווג תגובות חופשיות, להגדיר כל סולם וסולם בבירור.

סטנדרט 1.23 מקום בו נכלל שיפוט בפעולת חישוב ציוני המבחן, על המפתח להציג את הבסיס לחישוב הציונים ולשיפוט ואת ההליכים להכשרת מצייננים, בפירוט משביע רצון, על מנת לאפשר רמה מסוימת של אחדות דעים בקרב המצייננים, שתהא בת השוואה לרמה בה הופקו הנורמות. בעת הצורך, יש להפנות את תשומת הלב לוריאציות בחישוב הציונים, כמקור אפשרי חשוב לטעויות במדידה.

2. מהימנות וטעויות מדידה

(Reliability and errors of measurement)

רקע

המונח מהימנות מתייחס למדה בה תוצאות מבחן משוחררות מהשפעת טעויות מדידה.

הבדלים בתוצאות בין נוסח מבחן אחד למשנהו, או בין שתי העברות של אותו מבחן (לאותם נבחנים) בהפרש זמן ניתן לייחס לפה שמכונה טעויות במדידה. טעויות מדידה מפחיתות את מידת המהימנות (ומכאן את יכולת ההכללה) של התוצאות שהשיג אדם מסוים במדידה יחידה. (אין לייחס הבדלים כגון אלה לטעויות במדידה אם התבגרות, התערבות או אירוע אחר כלשהו, הפכו הבדלים אלה למשמעותיים, או, לחילופין, במידה ואי-העקביות בתשובות רלוונטיות למבנה הנמדד).

הערכה נאותה של מבחן צריכה להתבסס על זיהוי המקורות העיקריים לטעויות מדידה, גודל השגיאות הנובעות ממקורות אלה, המדד לרמת המהימנות שיש לצפות לה בין זוגות של תוצאות תחת נסיבות מסוימות, ומידת היכולת לערוך הכללה של התוצאות.

על מפתחי מבחנים ומו"לים מוטלת אחריות רבה בכל הנוגע לאיסוף נתונים ודיווח על עדויות הנוגעות למהימנות וטעויות במדידה. המשתמש הטיפוסי במבחנים אלה לא יבצע בדרך כלל מחקרי מהימנות משלו, אך מוטלת עליו האחריות לקבוע כי המידע שבידו הנוגע למהימנות וטעויות מדידה, רלוונטי לבדיקת אפשרויות בשימושים המתוכננים ולפרשנויות. בהיעדר מידע כזה, עליו לספק בעצמו את העדויות הנדרשות.

מקדם המהימנות הינו מונח רחב. ניתן לבסס מקדמי מהימנות שונים ואומדנים שונים למרכיבי טעויות מדידה, על סוגים שונים של עדויות; כל סוג של עדות מורה על משמעות אחרת. מקדם המהימנות המבוסס על הקשר בין תוצאות נוסחי מבחן חלופיים המשמשים בשני אירועים נפרדים, מושפע ממספר מקורות לטעות, הכוללים השתנות אקראית של התגובה, שינויים שחלו בנבחנים, הבדלים בתכני נוסחי המבחנים, והבדלים בצורת העברת המבחן. מאידך מקדם מהימנות המבוסס על השוואות תוצאות של חלקי מבחנים (למשל, מהימנות מבחן חצוי) מושפע ממקורות טעות הכוללים בעיקר השתנות אקראית של התגובה והבדלים תכניים.

חיוני הדבר, כי השיטה המשמשת להערכת מידת המהימנות, תביא בחשבון את מקורות הטעות המשמעותיים ביותר לשימושים ופירושים מסוימים של המבחן. לא כל מקורות הטעות צפויים להיות רלוונטיים למבחן נתון. על כן, אמידת מרכיבים מוגדרים היטב של שונות נצפית ושונות הטעויות, היא תוצאה שימושית במיוחד של מחקר מהימנות. דיווח על טעויות תקן, רווחי סמך או מדדים אחרים של אי-דיוק באמידה, אף הם מועילים. דיווח על מקדם המהימנות לבדו, מספק מידע מצומצם מדי.

אמדני מהימנותו של מבחן צריכים להתייחס לא רק למקורות לטעות, אלא, בנוסף, לסוג ההחלטות אותן יש לקבל על בסיס ציוני מבחנים ועל פי רמות המצרף (aggregation) הצפויות (פרטים לעומת קבוצות של נבחנים). אומדנים לעקביות ההחלטות נדרשים כל אימת שכללי ההכרעה מסווגים אנשים לקטגוריות על פי ציוני מבחן. הערכה של טעות התקן סביב נקודת החתך מועילה אף היא.

ציונים מורכבים (composite) חלקם צפויים להיות יותר מהימנים, וחלקם פחות מהימנים, מהציונים החלקיים. אולם, ציונים המייצגים הבדלים בין ציונים שהושגו בשני מבחנים או מביצוע חוזר של אותו המבחן (קרויים ציוני "שיפור"), הינם בדרך כלל פחות מהימנים מכל אחד מהחלקים. ציון מורכב נוסף, אשר בדרך כלל הינו פחות מהימן מחלקיו, הוא ציון יחס.

סטנדרט 2.1 לכל ציון כללי, ציון של תת-מבחן או שילוב של ציונים, יש לספק אומדנים למהימנות הרלוונטיות וטעויות התקן במדידה, במידה ראויה של פירוט.

סטנדרט 2.2 יש לתאר את המדגמים של פרטים, קבוצות או תצפיות, המשמשות למטרת אמידת מהימנות וטעויות מדידה.

סטנדרט 2.3 יש להגדיר בבהירות כל שיטה מדווחת לאמידת מהימנות, ולבטאה במונחים של מרכיבי שונות, מקדמי מתאם, טעויות תקן במדידה, אחוזי ההכרעות הנכונים, או סטטיסטיים שווי ערך. יש להסביר בבהירות את התנאים והמדגמים בהם התקבל האומדן למהימנות, ואת המצבים אשר עליהם ניתן לישמו.

סטנדרט 2.4 במידה ומקדמי המהימנות מתוקנים לקיצוץ תחום, יש לדווח על המקדם המתוקן והבלתי-מתוקן כאחד, יחד עם סטיות התקן של הקבוצה אשר נבחנו בפועל, וסטיות התקן של הקבוצה אשר אומדניה המתוקנים מוצגים.

סטנדרט 2.5 על אומדני מהימנות, המבוססים על נוסחים חלופיים של מבחן, אשר הופעלו על אותו מדגם של פרטים (נבחנים) בשני אירועים נפרדים, לציין את הסדר בו הופעלו שני הנוסחים, מרווח הזמן בין ההעברות, והרציונל לבחירת מרווח זמן זה. יש לספק ממוצעים וסטיות תקן שהתקבלו משני הנוסחים, כמו גם את טעויות התקן במדידה והאומדן למהימנות המקבילה.

הערות:

- א. במספר מקרים, יתכן ויהא זה רצוי לקבל תוצאות ביותר משתי נקודות זמן, בעיקר אם קיימת ציפיה לאי-יציבות ניכרת.
 - ב. מקום בו משמשים נוסחים מקבילים לבדיקת יציבות, יש להכיר בכך שהבדלי תוכן בין הנוסחים, בנוסף לאי היציבות, תורמים לשונות.
 - ג. אומדים ליציבות, המבוססים על העברה חוזרת של אותו הנוסח, עלולים להיות מוגדלים באורח המעוות את המציאות, בשל השפעות הזכרון.
-

סטנדרט 2.6 מקדמים המבוססים על ניתוח פנימי (internal analysis) אינם מהווים תחליף למהימנות מקבילה (alternate form), או לאומדני יציבות (stability).

סטנדרט 2.8 כאשר מעורבים תהליכי שיפוט (הערכה) בחישוב תוצאותיו של מבחן, יש לספק עדות לגבי מידת ההתאמה בין ההערכות הבלתי תלויות (מהימנות בין-מעריכים). במידה ולא סופקה עדות כזו, יש להפנות את תשומת הלב להבדלי שיפוט כמקור אפשרי חשוב לטעויות מדידה.

הערה: ניתוחי מרכיבי שונות הינם מועילים במיוחד לציוני מבחנים שעברו הליך של שיפוט; ניתוחים אלה יכולים לספק אומדי שונות נפרדים לשאלות, שיפוטים, משקלות המשמשים בתהליך השיפוט, ופרק הזמן שהוקצב, למשל. התאמה בין מעריכים או בין צופים, עשויה להיות חשובה במיוחד מקום בו נאספים נתוני תצפית הכרוכים בהבחנות דקות.

סטנדרט 2.9 מקום בו קיימות סיבות תיאורטיות או אמפיריות מקובלות, לצפות כי המהימנויות או טעויות התקן במדידה ישתנו באורח ניכר בין אוכלוסיות שונות, יש להציג אומדנים עבור כל אוכלוסיה מרכזית שלגביה מומלץ המבחן.

סטנדרט 2.10 יש לדווח על טעות התקן של המדידה באיזורי הציונים הקריטיים. כאשר מפורטים ציוני חתך לצורך ברירה או סיווג, יש לדווח על טעות התקן בנקודות החתך או בסמוך להם עבור החלטות דיכוטומיות, יש לספק אומדנים לאחוז הנבחנים המסווגים באותו האופן בשני אירועים או בנוסחים הלופיים של המבחן.

הערה: בנסיבות מסוימות, לא יהא זה בר ביצוע לדווח על טעויות תקן במדידה לגבי כל רמת ציונים, אולם יש לדווח עליהן ברמות או מרווחים מתאימים, המופרדים היטב זה מזה.

3. תקפות

(Validity)

רקע

התקפות היא אחד השיקולים המרכזיים בהערכת מבחנים. מתוצאות שהושגו במבחן נתון, ניתן לגזור מגוון של מסקנות, וקיימות דרכים רבות לצבירת עדויות לתמיכה בכל מסקנה ומסקנה. התקפות מתייחסת למידה בה העדות תומכת במסקנות הנגזרות מהתוצאות. המסקנות הנוגעות לשימושים הספציפיים במבחן הן החייבות בתיקוף, ולא המבחן עצמו.

עדויות התקפות קובצו, באופן מסורתי, לשלוש קטגוריות עקריות: תקפות תוכן, תקפות קריטריון ותקפות מבנה; משמעות כל קטגוריה תוהר להלן. ההבחנה בין הקטגוריות השונות אינה תמיד חד משמעית, וכמו כן, לא ניתן לומר שלתיקוף מסקנה מסוימת עדיפה קטגוריה מסוימת דווקא. תהליך אידאלי לתיקוף כולל מספר סוגים של עדויות. רצוי להגיע לצבירת תשלובת של עדויות, המשקפת באופן אופטימלי את איכותו של מבחן לאור המטרה המיועדת לו. בנסיבות מסוימות, עדות הקשורה לתוכנו של מבחן, הינה עדות קריטית; בנסיבות אחרות, עדות לגבי הקריטריון היא העדות הקריטית.

עדות הנוגעת למשמעותו הפסיכולוגית של מבנה (construct), הינה כמעט תמיד רלוונטית.

צבירת עדויות עשויה לעיתים להיות כרוכה לא רק בבדיקת הכלי הקיים במצב הנוכחי, אלא גם בבדיקת העדות הקיימת על שימוש בכלי זה או כלים הדומים לו, במצבים תהליך זה קרוי הכללה (generalization) - והוא מתבצע על בסיס אלמנטים משותפים (כמו במקרה של תקפות משותפת) או על בסיס של דמיון תפקיד כולל.

עדות לתקפות-מבנה (construct related evidence)

העדות הקשורה לתקפות-מבנה, מתמקדת בעיקר בציון המבחן כמדד למבנים פסיכולוגיים מבוקשים. יכולת תימצות, תפיסה מרחבית, הבנת הנשמע - כל אלה הינם מבנים. כך גם תכונות אישיות כגון דומיננטיות, סקרנות, צורך בהישג. על המבנה המבוקש, להיכלל במסגרת מושגית המגדירה את משמעות המבנה, המבחינה בינו לבין מבנים אחרים, והמציינת כיצד יש לקשור בין המדדים למבנה זה, לבין משתנים אחרים.

עדות על פרשנות מבנה המבחן, ניתן להשיג ממספר מקורות. כך, למשל, ניתן להשתמש בהתאמה פנימית בין פריטי מבחן (הומוגניות), כדי לתמוך בטענה, כי המבחן מודד בעיקר מבנה יחיד. עוצמת הקשר בין המבחן הנידון למדדים אחרים, אשר אמורים להשתייך לאותו המבנה, וחולשת הקשר בין המבחן למדדים אחרים, אשר לכאורה משתייכים למבנים אחרים, תומכים הן בזיהוי המבנים, הן בהבחנה ביניהם.

עדות לתקפות תוכן (content related evidence)

ככלל, עדות לתקפות-תוכן משקפת את המידה בה מייצגות דגימות פריטים, מטלות או שאלות, את תחום התוכן המוגדר. בדרך כלל, מסתמכות השיטות על שיפוט של מומחים, בעת נסיון להעריך את היחס בין חלקי המבחן והתחום המוגדר. לדוגמה, את שטחותיו (facets) של תחום לימודי, ניתן להגדיר, וניתן לבקש ממומחים בתחום זה לסווג סדרות של פריטים לקטגוריות המוגדרות באמצעות שטחות אלה. דוגמה נוספת: ניתן לשלב תצפיות שיטתיות על התנהגות במסגרת עיסוק מסוים, עם שיפוט של מומחים, במגמה ליצור התנהגויות מדגם מייצג או מדגם הכרחי (קרי תחום העיסוק).

עדות לתקפות-תוכן (יותר מאשר סוגי תקפיות אחרים) מהווה גורם בעל חשיבות מרכזית במהלך פיתוח המבחן, בין אם פיתוח כזה מתבצע במסגרת מחקרית, או בהקשר של פעילות יישומית. על שיפוט מקצועי של מומחים להוות חלק אינטגרלי ממלאכת הגדרת הדברים הנמדדים, כגון תיאור תחום התוכן, הפקת או בחירת דגימת תוכן, והגדרת פורמט הפריטים והשיטה לחישוב התוצאות. מכאן, שהמסקנות לגבי תקפות התוכן, קשורות לתהליך בניית המבחן, כמו גם לביסוס עדות לתקפות, לאחר שפוח המבחן ונבחר לשימוש.

עדות לתקפות-קריטריון (criterion related evidence)

עדות לתקפות קריטריון מראה כי ציוני המבחן קשורים אמפירית לקריטריון אחד, או למספר קריטריונים.

את היחס בין ציוני המבחן ומדדי הקריטריון, ניתן לבטא במספר צורות, אולם השאלה הבסיסית היא תמיד: "באיזו מידה של דיוק ניתן לחזות ביצוע קריטריון על פי ציוני מבחן?" הקביעה, האם רמת דיוק חיזוי מסוימת הינה גבוהה או נמוכה, מועילה או בלתי מועילה, תלויה בהקשר בו מתקבלת ההכרעה.

בחירת קריטריון והליכי המדידה המשמשים לקבלת מדדי-קריטריון, הם בעלי חשיבות מרכזית.

הכללת תקפות (validity generalization)

נושא חשוב במסגרות חינוכיות ותעסוקתיות, הוא המידה בה ניתן להכליל עדות לתקפות-קריטריון, אשר התקבלה במצב אחד למצב אחר, מבלי לערוך מחקר נוסף באשר לתקפות המצב החדש. במידה ויכולת ההכללה מוגבלת, אזי יהיה צורך בעדות לתקפות-קריטריון ספציפית למצב החדש. במידה וניתן להגיע להכללה רחבה, יתכן כי לא יהיה צורך בעדות חדשה. כמה מההיבטים העיקריים של שאלת הכללת התקפות הם:

- (א) מידת הדמיון בין החזאים של שני המצבים.
- (ב) מידת הדמיון של הקריטריונים.
- (ג) מידת הדמיון של סיטואציות המבחן.
- (ד) מידת הדמיון של סוג הנבחנים להם מיועד המבחן.
- (ה) התקופה בה נערך המחקר (בעית ההתיישנות).

אחת השאלות הקשות, שאיננה ניתנת לפתרון חד משמעי היא, מתי בדיוק השתנות ההיבטים הללו היא מרחיקת לכת מכדי לאפשר הכללה.

הטייה בשימוש במבחנים

נתונים אמפיריים בדבר קשר בין חזאי לקריטריון יכולים להצביע על מצב של משוואות חיזוי שונות עבור קבוצות דמוגרפיות שונות, קבוצות בעלות התנסויות שונות, או עבור קבוצות אשר עברו טיפול שונה או שהינן מעורבות בתוכניות לימודיות שונות.

אם נמצא כי חיזוי נתון גורר חיזויים שונים של רמת הקריטריון לגבי אנשים המשתיכים לקבוצות שונות, אזי, אם מתעלמים מכך ונותנים אותה פרשנות לתוצאה נתונה, ללא קשר לקבוצה אליה משתייך הפרט, הכרעת הברירה תהיה מוטא.

אפשרות אחת להתמודד עם הבעיה היא לבנות אלגוריתם נפרד (לדוגמה, קו רגרסיה) לכל קבוצה. אפשרות נוספת היא לחפש חזאים אשר מפחיתים את החיזוי הדיפרנציאלי, מבלי לצמצם באורח ניכר את מידת הדיוק הכללית של החיזוי.

קיימות מספר דרכים להערכת הטיית ברירה ולטיפול בהם, והן מתבססות על הגדרות שונות של הוגנותו של הליך הברירה. שלא כמו הטיית ברירה, הוגנות אינה מונח פסיכומטרי טכני; זהו מונח הנתון להגדרות שונות בנסיבות חברתיות ופוליטיות שונות.

חיזוי דיפרנציאלי

בהקשר של סיווג או שיבוץ, נדרשת עדות על מנת לשפוט את מידת התאמתו של מבחן לסיווג או הכוונת פרט למשרה מסוימת ולא לאחרת, לטיפול מסוים ולא לאחר. אפשר שמבחנים יהיו בעלי יכולת חיזוי גבוהה לגבי ביצועים בתוכניות לימודים שונות או משרות שונות, מבלי שיספקו את המידע הדיפרנציאלי הנחוץ לשם שיפוט השוואתי של יעילות ההכוונה או הטיפול.

סטנדרט 3.1 יש להציג עדות לתקפותן של המסקנות העיקריות אשר לגביהן מומלץ השימוש במבחן.

הערה: הכמות והאיכות של העדויות הנדרשות לתקפות, תלויות בשאלה הספציפית הנשאלת, וכן בקיום עדויות קודמות.

סטנדרט 3.2 אם לא נבדקה תקפותו של פירוש מקובל, יש להבהיר עובדה זו, ויש להזהיר משתמשים בכוח לגבי מתן פירושים כאלה.

הערה: לא קיים מבחן התקף לכל המטרות או לכל המצבים, כפי שלא קיימים שני מצבים זהים. לכן המשתמש נדרש תמיד לערוך הכללה כלשהי. על מפתחי המבחן להציג את עדותם המפורטת לתקפות, כך שהיא תוכל לסייע לעריכת הכללה כזו.

סטנדרט 3.3 כל אימת שמוצע פירוש לתוצאות של תת-מבחן, הפרשי תוצאות או פרופילים, יש להציג עדות המצדיקה באופן ברור וישיר פירוש כזה. כאשר מתשבים ציונים מורכבים, יש לתת את הבסיס והרציונל למתן משקלות לתוצאות תת-המבחנים. קיום עדויות התומכות בכל אחד מהמרכיבים בנפרד (למשל, לגבי פירוש פרופיל) אינו מספיק.

סטנדרט 3.4 את הרכבו של מדגם התיקוף והתנאים בהם בוצע התיקוף יש לתאר בפירוט מירבי.

סטנדרט 3.5 כאשר עדות לתקפות-תוכן מהווה אלמנט מרכזי בהקשר לשימוש במבחן מסוים, יש לספק הגדרה ברורה של עולם התוכן הנידון, מידת הרלוונטיות שלו לשימוש המוצע למבחן, ולתאר את ההליכים ששימשו לבניית המבחן, אשר אמור לייצג עולם תוכן זה. כאשר דגימת התוכן מיועדת לשקף קריטיות ולא ייצוגיות, אזי יש לתאר בפירוט אף את הרציונל לדגש היחסי שהושם על גורמים קריטיים בתחום.

סטנדרט 3.6 במידה ומומחים לנושא מסוים נתבקשו לשפוט, האם פריטים מהווים מדגם מתאים לתחום, הרי שיש לספק תיאור של הכשרתם הרלוונטית, נסיונם וכישוריהם של המומחים. יש לתאר בפירוט כל תהליך השיפוט.

סטנדרט 3.7 כאשר מוצע מבחן כמדד למבנה (פסיכולוגי), יש להבחין בין מבנה זה למבנים אחרים; יש להצהיר בבחירות מהו הפירוש המוצע לציוני המבחן. כמו כן, יש להציג עדות לתקפות-מבנה, כתמיכה במסקנות כאלה.

סטנדרט 3.8 כאשר מוצע מבחן כמדד למבנה, יש להציג עדות לכך שהקשר בין הציון למבנה זה - כאשר הוא נמדד באמצעות שיטות (מבחנים) שונות - הדוק יותר מאשר הקשר בין הציון הזה למדדי מבנים הנבדלים ממנו באורח בולט. בנוסף, כאשר מתקבלים מספר ציונים ממבחן יחיד, כאשר כל אחד מהם אמור למדוד מבנה אחר, הרי שיש לדווח על הקשרים הפנימיים (למשל intercorrelations) בין הציונים, בתוך מדגם אחד או יותר על עדות לתקפות-מבנה ההראות, כי הקשר בין ציוני המבחן הנידון לציונים במבחנים מגוונים אחרים מתנהגים על-פי תיאוריה.

סטנדרט 3.9 בדו"ח על מחקר תקפות-קריטריון (criterion related study), יש לכלול את תיאור המדגם והניתוח הסטטיסטי ששימש לאמוד הדיוק בחיזוי. הסטטיסטיקות הבסיסיות תכלולנה את מספר המקרים (N), מדדים לנטייה מרכזית והשתנות, קשרים סטטיסטיים, ותיאור של כל נטייה בולטת לעבר התפלגות בלתי נורמלית.

סטנדרט 3.10 יש לתאר במדויק כל מדד קריטריוני; כמו כן, יש להסביר את הרציונל לבחירתו כקריטריון רלוונטי.

סטנדרט 3.11 יש לבחון בזהירות את האיכות הטכנית של כל מערכת קריטריונים. את הקריטריונים יש למדוד באורח בלתי תלוי בציוני חיזוי של מבחן (סכנת "זיהום"). אם העדות מצביעה על כך שהמדד הקריטריוני מושפע במידה ניכרת מגורמים בלתי רלוונטיים, אזי יש לדווח על עדות זו. כאשר ננקטים צעדים מיוחדים על מנת לצמצם את השפעותיהם של גורמים בלתי רלוונטיים, יש לתאר צעדים אלה בפירוט ראוי.

סטנדרט 3.12 מקום בו מבוססת מערכת קריטריונים על שיפוטיהם של מעריכים, יש לדווח על רמת הידע של המעריכים, הקשורה לביצועי הערכה. במידת האפשר, יש לציין את רמת הנסיון וההכשרה של המעריכים.

סטנדרט 3.13 במידה ומתקבלים מספר מדדים קריטריוניים, אולם לצרכי מחקר מסוים, נעשה שימוש בתוצאת קריטריון מורכבת (composite) אחת בלבד, הרי שיש לתאר את הכללים לשילוב הקריטריונים.

סטנדרט 3.14 כאשר לא ניתן להשיג עדות ייחודית וישירה לתקפות, ניתן לבסס שימוש מסויים על הכללת תקפות שמקורה במחקרים קודמים, בתנאי שהמצב הספציפי בו נעשה שימוש במבחן, דומה במידה רבה לקבוצת המצבים עליה התבססה הכללת התקפות.

סטנדרט 3.15 כאשר נערכות התאמות סטטיסטיות, כגון אלו הנעשות לצורך קיצוץ תחום (restriction of range) או הקלשה (attenuation), למשל, יש לדווח על המקדמים המתוקנים והבלתי מתוקנים כאחד; כמו כן, יש לדווח על הטכניקות אשר שימשו להתאמה ולתיקון.

סטנדרט 3.16 יש לדווח על מרווח הזמן בין הפעלת המבחן לאיסוף נתוני קריטריון. יש לפרט את לוח הזמנים ולציין את פרק הזמן שהוקדש לאיסוף הנתונים.

סטנדרט 3.17 כאשר קיימת עדות לתקפות-קריטריון, יש להציג את הרציונל לבחירה בין גישת החיזוי (predictive) לבין הגישה המקבילה (concurrent).

סטנדרט 3.18 במחקרים העוסקים בתקפות-קריטריון עבור מצבי ברירה, יש לכלול, במידת האפשר, מחקר לגבי עוצמת ההטיה בחיזוי עקב חיזוי דיפרנציאלי באשר לקבוצות אשר לגביהן קיימת סבירות מקדמית גבוהה לחיזוי דיפרנציאלי, כאשר הן נחשפות למבחן המסוים הנידון.

סטנדרט 3.19 כאשר נערכים מחקרים בנושא חיזוי דיפרנציאלי, על הדו"חות לכלול משוואות רגרסיה (או מקבילה נאותה) נפרדות עבור כל קבוצה, עיסוק או טיפול. מקדמי מתאם כשלעצמם אינם מספיקים.

סטנדרט 3.20 כאשר מבחן מתוכנן או משמש בפועל למטרת סיווג אנשים לקבוצות טיפול חלופיות מוגדרות (כגון תוכניות חלופיות לתעסוקה, תיראפיה או חינוך), יש לספק עדות לחיזוי דיפרנציאלי של מבחן, לצורך מטרה זו.

סטנדרט 3.21 במידה ומומלץ לעשות שימוש בנקודות חתך ספציפיות לצורך קבלת הכרעות סווגיות (לדוגמה - לצורך אבחון פתולוגי דיפרנציאלי) על המדריך למשתמש (test manual) לכלול אזהרה בדבר שיעורי הסווג המוטעה הצפויים, המשתנים בהתאם לאחוז הפרטים הנבחנים, אשר משתייך למעשה לכל קטגוריה.

סטנדרט 3.22 כאשר נבחר מספר מצומצם של חזאים מתוך מאגר גדול, והמשקלות נקבעים בו-זמנית, יש להביא בחשבון חוסר יציבות אפשרי של מקדמי המשקלות והתקפות (טעויות דגימה), ולכן, יש צורך לערוך הצלבת תקפות למקדמים הנ"ל (cross-validation).

4. סילום, יצירת נורמות, יכולת השוואה בין ציונים וכיול (Scaling, norming, score comparability, and equating)

רקע

פעמים רבות יש להסתייע, בעת פירוש ציוני מבחן, בנורמות, אשר מורות על היחס בין תוצאות גולמיות לבין התפלגויות תיאורטיות או אמפיריות מוגדרות. בעת השוואת ציונים מנוסחים שונים של אותו מבחן, נוסחים מחודשים, ואפילו מבחנים שונים, המשמעות נגזרת מכיול התוצאות או התאמת משקלות שבלעדיהם אין אפשרות להשוואה.

סולמות ונורמות

בשל העובדה, כי ציונים גולמיים של מבחן מבוטאים ביחידות שמקורן במאפיינים השרירותיים של המבחן, פותח מגוון של אמצעי סילום לצורך דיווח על הציונים. לכל שימוש נתון במבחן, ניתן למצוא מסגרות התייחסות אפשריות רבות של נורמות: נורמות אוכלוסיה ספציפית; נורמות ארציות; נורמות המבוססות על סיווגים תעסוקתיים וכד'

מלאכת יצירת הנורמות היא מלאכה קשה ועלוהה גבוהה ביותר; אין צורך לבצעה באופן אוטומטי או להציגה כדרישה לכל המבחנים המתפרסמים. מספר מבחנים, למשל, כוללים רק "נורמות למשתמש" (user norms) או "נורמות תוכנית" (program norms), אשר מורכבות מסטטיסטיקות תיאוריות המבוססות על כל אוכלוסית הנבחנים בפרק זמן נתון, ולא על נורמות שהתקבלו באמצעות שיטות דגימה פורמליות. חרף העובדה כי סטטיסטיקות תיאוריות כגון אלה המתבססות על מדגמים הזדמנותיים, הינן מועילות לעיתים קרובות, הרי שאין לבלבל בינן לבין נורמות המייצגות קבוצות מוגדרות ופורמליות יותר.

השוואתיות וכיול

במצבים רבים נדרשים מספר נוסחים מקבילים של מבחן. האידיאל הוא, כי נוסחים חלופיים של מבחן יוכלו לשמש כנוסחים שווי-ערך לחלוטין; אולם לא ניתן לממש אידיאל זה במישור המעשה. גם שינויים משניים ביותר בתוכנם של שני נוסחים, עשויים למנוע מהנוסחים להיות חלופיים במובן האידיאלי של המילה.

אם מדווחות תוצאות גולמיות בלבד, הרי שהבדל בלתי מכוון במידת הקושי בין נוסחים, יעניק יתרון לאנשים שניגשו למבחן כאשר הופיע הנוסח הקל יותר. על כן, קיים בדרך כלל הצורך בהמרת התוצאות שהתקבלו מנוסח אחד, ליחידות של נוסח אחר - תהליך המכונה בד"כ כיול (equating). ההמרה עשויה לכלול משוואות או טבלת ציונים מומרים, הקשורים לכל תוצאה נצפית (לדוגמה - מספר התשובות הנכונות בנוסח של מבחן). בכל אחד משני המקרים, תהיה ההמרה נתונה לשגיאה, בשל הליך ההמרה והדגימה.

נוסחי מבחנים השונים זה מזה מבחינת מהימנותם, לא ניתנים לכיול במובנה המדויק של המילה. מועמד אשר ציונו צפוי להיות נמוך מציון חתך ידוע, יהנה מסיכוי גבוה יותר לעבור את המבחן, באמצעות היבחנות במבחן מהימן פחות, וזאת משום שיש לו תקווה שטעות המדידה תפעל לטובתו. מאידך, אנשים המצפים לציונים גבוהים מציון החתך, אמורים להעדיף מבחן יותר מהימן, על מנת למזער את הסיכוי לטעות מדידה אשר תפעל לרעתם.

בשל העובדה, כי מבחנים בעלי רמת מהימנות שונה אינם ניתנים לכיול במובן המדויק של המילה, הרי שלעיתים קרובות רצוי להמיר את הציונים כך שיהיו ברי השוואה. מספר שיטות להמרה נחשבות לתהליך של כיול, אולם כינוי זה הינו מטעה. תיאור הולם יותר הוא כי התוצאות סולמו על מנת להפכן לבנות-השוואה. לומר שהפכנו ציונים לבני השוואה - זוהי טענה חלשה יותר מאשר לומר שהם כויילו. ציונים מכוילים אמורים להיות חלופיים זה לזה, בעוד שציונים ברי-השוואה אמורים להידמות זה לזה במובן מסוים. לדוגמה, ציונים ברי-השוואה עשויים להיות מקבילים לדירוג אחוזי (פרצנטילה) הזחה באוכלוסיה מסוימת; אולם לא בהכרח ישתמע מכך, כי לאנשים בעלי יכולת שווה יהיה אותו הסיכוי לקבל ציון נתון, ואף לא כי ציונים ברי-השוואה יתאימו לאותו האחוזון באוכלוסיות אחרות. לעומת זאת, פרשנויות כאלה יהיו נאמנות למציאות במידה ויערך לציונים כיול קפדני.

משם כפי שלא ניתן לכייל ציונים כאשר המבחנים שונים זה מזה ברמת מהימנותם, כך לא ניתן לכייל ציונים של מבחנים המודדים מאפיינים שונים זה מזה.

בעיות קשות מתעוררות כאשר המבחנים מודדים מאפיינים דומים, אך לא זהים. כדוגמאות לכך ישמשו שלושה מצבים נפוצים:

(א) כיול מבחנים בעלי שמות ומטרות דומים, אולם הם בעלי מפרטי תוכן שונים, והוצאו לאור ע"י מו"לי מבחנים שונים;

(ב) כיול נוסח חדש לנוסח ישן בעקבות שינוי מהותי במפרטי המבחן;

(ג) כיול אנכי בין מבחנים שעוצבו למטרת התאמה מירבית לרמות התפתחות או חינוך שונות (לדוגמה - כיול מבחן שעוצב עבור כיתות ב' ו-ג' עם מבחן שעוצב עבור כיתות ד' ו-ה').

בכל אחד מהמקרים הללו, יתכנו סיבות טובות לרצות כי הציונים ישוקללו כך שיהיו, במובן מסוים, ברי השוואה. ההליכים המשמשים לקבלת ציונים ברי-השוואה עשויים להיות זהים להליכים המשמשים לכיול, אולם לא תהיה עמידה בדרישות הקפדניות הנלוות לתהליך כיול מבחנים; על כן, התוצאות המתקבלות יכוננו תוצאות מסולמות (scaled) או בנות-השוואה (comparable), אך לא מכוילות (equated).

סטנדרט 4.1 יש לתאר את הסולמות המשמשים לדווח ציונים, ואת הרציונל לבחירתם, בפרסומי המבחן ובמדריך, על מנת לאפשר פירוש מדויק של הציונים, הן על ידי המשתמש, הן על ידי הנבחן. יש לפרט כיצד נגזרו הציונים המסולמים מהציונים הגולמיים.

סטנדרט 4.3 על הנורמות המוצגות להתייחס לקבוצות שתוארו בהירות. על קבוצות אלה להיות מהסוג הרצוי העתידי למשתמשים במבחן. על מו"לי המבחנים אף לעודד משתמשים לפתח נורמות מקומיות/ספציפיות, כאשר הנורמות שפורסמו אינן מספקות לגבי משתמשים מסוימים. הערה: יש להביא לכך שמשמשים יהיו ערים למצבים בהם נורמות מתאימות במידה שונה לקבוצות שונות.

סטנדרט 4.4 דו"חות המתארים נורמות צריכים: לציין את השנה בה נאספו הנתונים הנורמטיביים, לספק סטטיסטיקות תיאוריות ולתאר את שיטת הדיגום ושיעורי שיתוף הפעולה, כל זאת בפירוט מספק, כך שניתן יהיה להעריך את איכותו של המחקר.

סטנדרט 4.5 כאשר קיימת כוונה להשתמש בתוצאות שהתקבלו מנוסחים שונים של מבחן - לרבות מבחנים ממוחשבים או מבחנים אדפטיביים ממוחשבים - באופן חלופי, יש לספק מידע לגבי המידה בה נוסחים אלה מקבילים זה לזה. יש לספק מידע על ההקבלה, ובכלל זה, מידע ספציפי על שיטת הכיול, צורת ההעברה, הליכים סטטיסטיים בהם נעשה שימוש, הדגימה וגודל הדגימה. יש לדווח על בדיקות תקופתיות באשר לאיכות הכיול.

סטנדרט 4.6 כאשר נעשה שימוש במערך של מבחן עוגן (anchor test design), בעת כיוול נוסחי מבחנים, יש לתאר את מאפייניו של מבחן העוגן, ובעיקר, את היחס בינו לבין הנוסחים המכויילים. עבור כל דגימת נבחנים, בכל נוסח, יש לספק בנפרד מפרטי תוכן למבחן העוגן, ומידע סטטיסטי לגבי היחס בין מבחן העוגן לכל נוסח.

סטנדרט 4.7 על תוכניות מבחנים מתמשכות, המנסות לשמור על סולם ציונים משותף לאורך זמן, לנהל בדיקות תקופתיות באשר ליציבות הסולם.

5. פרטומי מבחנים: מדריכים טכניים ומדריכים למשתמשים

(Test publications: technical manuals and users' guides)

רקע

על מו"לים לספק מידע, אשר יאפשר את הערכת מידת התאמתו ואיכותו הטכנית של המבחן. יש לספק מקורות עזר נפרדים עבור כל צורת הצגה של המבחן, לדוגמה - מבחן בכתב (נייר ועפרון), מבחן ממוחשב וכד'. יש לספק אף דוגמאות לפירושי ציונים הקשורים ליישומים המתוכננים ע"י מפתח המבחן, כך שיתאפשר למשתמש לבצע שיפוט מקצועי באשר למסקנות שהוא מבקש להסיק על בסיס תוצאות המבחנים. בדרך כלל, אין המדריך אמור לכלול יותר מתקצירים של המחקרים, אולם יש לאפשר לכל המעוניין להשיג את המקורות המוזכרים במדריך.

יש להעריך מדריך על פי מידת שלמותו, דיוקו ובהירותו. על המדריך לספק מידע הנוגע לקבוצות נבחנים רבות ושונות. יתכן וחלקים מהמדריך, יחולקו עבור קטגוריות מוגדרות של משתמשים (לדוגמה - יועצים, פסיכולוגים בבתי ספר, מנהלים, מורים, חוקרים או נבחנים).

על מפתחי המבחנים והמו"לים לספק במדריכים מידע אשר יסייע לצמצם את הסיכוי לאי-הבנה או שימוש לא נכון בתוצאות המבחנים.

גם כאשר מפותח מבחן (או סוללת מבחנים) לצורך שימוש פנימי במסגרת ארגון יחיד, תופק תועלת רבה ממדריך קצר. הכנת המדריך מסייעת למפתח המבחן לארגן את מחשבותיו, להסדיר הליכים, להביא את רעיונותיו וכוונותיו לידיעת המשתמשים.

סטנדרט 5.1 יש לספק מדריך טכני למשתמשים העתידיים במבחן בעת הבאת המבחן לדפוס או שרורו לשימוש מעשי.

סטנדרט 5.2 במדריכי המבחן יש לכלול תיאור יסודי של הרציונל למבחן, הצהרות לגבי השימושים המומלצים למבחן, ותקציר העובדות התומכות בשימושים כגון אלה. כאשר ניתן לצפות שיעשה שימוש בלתי נאות במבחן, על מדריך המבחן לכלול אזהרות ספציפיות מפני שימושים בלתי ראויים כאלה.

סטנדרט 5.3 יש להקפיד ולוודא כי החומר המשמש לפרסום והרחבת תפוצתו של מבחן יהיה מדויק. על מו"לים להימנע משיטות פרסום המרמזות כי המבחן משיג דבר מה, אם אין לכך גיבוי או תמיכה בבסיס המחקרי של המבחן.

סטנדרט 5.4 על מדריכים טכניים לאזכר סדרה מאוזנת ומייצגת של מחקרים הנוגעים לשימושים ספציפיים וכלליים במבחן. את המחקרים המאוזכרים עצמם - על המו"ל להיות מסוגל לספק, על פי דרישה, למשתמש העתידי.

סטנדרט 5.5 על מדריכי המבחנים לציין כל הכשרה (qualification) מיוחדת בה צריך לעמוד האדם המעביר את המבחן והאדם המפרש את המבחן.

סטנדרט 5.6 בנוסף לפרסום מדריכי מבחנים הכוללים תיקונים, השלמות ועריכה מחודשת, על מפתחי המבחנים והמו"לים להגיב בחיוב לבקשות סבירות לקבלת מידע נוסף, אשר עשוי להידרש על מנת לקבוע את מידת התאמתו או אי התאמתו של השימוש המיועד למבחן.

סטנדרט 5.7 דיווחים המופיעים במדריך המבחן, על היחסים בין ציוני המבחן לבין מערכת קריטריונים מסוימת, הינם, כפי המשתמע, כמותיים, ויש להציגם בצורה זו.

הערה: מדריך עשוי לציין, לדוגמה, כי יכולת מרחבית היא תכונה הנדרשת לתחומי האדריכלות וההנדסה. הצהרה זו אינה משביעת רצון מבחינת הדרישה לטיעונים כמותיים, מאחר ואינה מציינת באיזו מידה הודגם אמפירית כי ההצלחה בתחום האדריכלות תלויה ביכולת מרחבית.

סטנדרט 5.8 אין לכלול במדריך למבחן טענה המרמזת כי המבחן נתון לפירוש פשוט בידי הדיוטות, אלא אם כן קיימת עדות התומכת בתקפות טענה זו. בהיעדר עדות כזו, על המדריך לציין מהו המידע שיש לספק על תוצאות המבחן, לאנשים החסרים את ההכשרה הנדרשת בד"כ לצורך מתן פירושים לתוצאות. לצד מבחנים שתוכננו לאפשר לנבחנים לחשב את תוצאותיהם, יש לספק עזרים לפרשנות.

סטנדרט 5.9 על ארגונים המציעים פירוש מבחנים באמצעים אוטומטיים (automated test interpretation), לאפשר קבלת מידע לגבי רצינות המבחן ותקציר העדויות התומכות בפירוש שניתן. על מידע זה לכלול את מידת תקפותם של ציוני החתך או הכללים התבניתיים (configural rules), המשמשים לתהליך זה, ותיאור המדגמים מהם התקבלו.

הערה: פרסום נקודות החתך או הכלליים התבניתיים, הינו פעולה רצויה מאד. מקום בו אינטרסים מסחריים גוררים אי מתן אלגוריתמים כאלה, ניתן לערוך הסדרים לסקירה חיצונית של אלגוריתמים אלה בחסות גורם (שלישי) מקצועי, אשר יבטיח הגנה נאותה על האינטרס המסחרי.

חלק שני

סטנדרטים מקצועיים לשימוש במבחנים

Professional Standards for Test Use

6. עקרונות כלליים לשימוש במבחנים
(General principles of test use)

רקע

ההבחנה בין פיתוח מבחן לשימוש בו אינה תמיד ברורה. על המשתמשים להיות מודעים לסטנדרטים החלים על פיתוח מבחן, כך שיוכלו להעריך את המבחנים בהם הם שוקלים להשתמש. פרק זה כולל סטנדרטים כלליים, הישימים לשימושים במבחנים בכל המסגרות פעילות מקצועית. הפרקים הבאים, מציגים סטנדרטים ספציפיים יותר לתחומי מומחיות.

על המשתמש במבחן, כאשר הוא בוחר או מפרש מבחן, לדעת מהי מטרת המבחן ותוצאותיו הצפויות. על המשתמש להכיר את ההליכים הנדרשים לצורך יעילות, ולצורך מיזעור ההטייה בשימוש במבחנים. האחריות על שימוש נאות במבחן מוטלת על המשתמש. על המשתמש להכיר את המבחן ואת השימושים הנאותים שלו, ועליו לספק מידע זה לאחרים, כפי הנדרש.

6.1 סטנדרט על המשתמשים במבחנים להעריך את התיעוד הקיים באשר לתקפותם ומהימנותם של מבחנים בהקשר לשימוש הספציפי המיועד להם.

יתכנו מצבים בהם המשתמש יאלץ לבסס באופן חלקי את שיפוטי, הכרעותיו ומסקנותיו המקצועיות, על מבחנים אשר לגבי מהימנותם או תקפותם, קיימת עדות מועטה בלבד, בהקשר לשימוש המיועד להם. במצבים אלה, על המשתמשים להימנע מלרמוז על יומרות גבוהות יותר.

6.2 סטנדרט כאשר עורך המשתמש שינוי ניכר בפורמט המבחן, בשיטת ההפעלה, ההוראות, השפה או התוכן, עליו לברר את מידת תקפותו של השימוש במבחן על השינויים שנערכו בו, או להציג נימוקים התומכים בטענה, כי אין צורך, או לא קיימת אפשרות לבסס תקפות נוספת.

6.3 סטנדרט מקום בו מיועד המבחן לשימוש למטרה שלא בעבורה בוססה תקפותו, או כאשר לא קיימת תמיכה בטענה לתקפות, אזי על המשתמש האחריות לספק עדות לתקפות, או לוותר על השימוש במבחן.

6.4 סטנדרט על המשתמש במבחן לציין באורח מדויק, מהי מידת הרלוונטיות של מבחן לתחליף ההערכה וקבלת ההכרעות, ואין הוא רשאי להשתמש בציון מבחן על מנת להצדיק הערכות, המלצות או הכרעות, אשר התקבלו בעיקר על בסיס מקור אחר.

סטנדרט 6.5 על משתמשים להיות ערים להשלכות בלתי מכוונות של השימוש במבחנים, ועליהם להשתדל להימנע מפעילות הגוררת השלכות שליליות בלתי מכוונות.
הערה: לדוגמה: על המשתמשים במבחנים לפעול נגד נטייתם של אנשים לייחס משמעות מופרזת ונטולת בסיס לציוני מבחנים.

סטנדרט 6.6 את האחריות לשימוש במבחנים רשאים ליטול על עצמם, אך ורק אנשים אשר להם ההכשרה והנסיון הנדרשים לנשיאה באחריות זו באורח מקצועי וטכני נאות. יש לעמוד בכל דרישה המופיעה במדריך, הנוגעת לכישורים הדרושים לצורך העברת מבחן ומתן פירוש לתוצאותיו.
הערה: לסטנדרט זה משמעות מיוחדת במצבים בהם רבה ההשפעה הפוטנציאלית של הבחנים על הנבחנים.

סטנדרט 6.7 על המשתמשים במבחנים לוודא, באמצעות בדיקות תקופתיות, כי ההליכים הנוכחיים בהם הם עושים שימוש, לא הפכו בלתי מתאימים בשל שינויים שחלו באוכלוסיית הנבחנים, במטרות המבחנים, או בשיטות הקיימות.

סטנדרט 6.8 כאשר משוחררות תוצאות מבחנים לפרסום, על האחראים לשחרור התוצאות לספק מידע אשר ימנע פרשנות מוטעית של ציוני המבחן.

סטנדרט 6.9 כאשר נעשה שימוש בנקודות חתך על מנת לברור, לסווג או להעניק אישורים/תעודות לנבחנים, יציג המשתמש את השיטה והרציונל ששימשו לקבלת נקודת חתך זו, לרבות הניתוחים הטכניים.
מקום בו נקודות החתך מבוססות בעיקר על שיפוט מקצועי (הערכה), יש לתעד את כישוריהם של השופטים.

סטנדרט 6.10 בעת השימוש במבחנים אל למעבירי המבחנים והמשתמשים לנסות להעריך נבחנים אשר מאפייניהם המיוחדים - גיל, מוגבלות, רקע לשוני או תרבותי - חורגים מתחומי הכשרתם האקדמית או נסיונם של המשתמשים.

סטנדרט 6.11 בעת השימוש במבחנים, אין להתייחס לציונו של נבחן כמשקף יכולת ירודה, ללא שקילת הסברים הלופיים לאי יכולתו של הנבחן לבצע בהצלחה את מטלות המבחן בהפעלה זו.

הערה: דוגמאות להסברים הלופיים שיש לשקול הן: רקע תרבותי, שפה, גיל, מין, תנאים בריאותיים מגבילים וכד'.

סטנדרט 6.12 יש להשתמש במבחנים שפותחו לסינון ראשוני אך ורק לצורך זיהוי נבחנים אשר עשויים להיזדקק להערכה נוספת. אין להשתמש בתוצאות מבחנים אלה על מנת לאפיין אדם או לקבל הכרעות לגביו.

סטנדרט 6.13 המשתמשים במבחנים אינם רשאים להשתמש בפירושים לציוני מבחן, לרבות כאלה שפורשו ע"י מחשב, אלא אם כן מצוי בידם מדריך למבחן זה, הכולל מידע לגבי תקפות הפירושים לצורך היישומים המיועדים, ומידע על המדגמים עליהן בוססה תקפות זו.

הערה: המשתמש בשירותי מחשב, מחוייב להתמצא באורח יסודי בעקרונות על פיהם בוצעו הפירושים, ועליו להיות מסוגל להעריך את פירושי הביצועים במבחנים המבוססים על מחשב, לאור עדויות נוספות. קיים צורך בשיפוט מקצועי ניכר, על מנת להשתמש כראוי בפירושים המבוססים על מחשב.

7. מבחנים קליניים
(Clinical testing)

רקע

השימוש במבחנים כחלק מתהליך הערכה קלינית, מאופיין בעיקר בעובדה, שהמבחנים נבחרים באופן אינדיבידואלי עבור כל פרט, ובדרך כלל נבחן רק אדם אחד בעת ובעונה אחת. מגוון רחב של מבחנים והליכים מצוי בשימוש, לרבות מבחני כושר שכלי ויכולת, מבחני אישיות, תצפיות התנהגותיות ואחרים. המבחנים משמשים לקבלת הכרעות לגבי סיווג אבחוני, איתור נזקים נוירולוגיים, זיהוי של תכונות אישיות והערכת תוצאות טיפול.

מבחנים קליניים עשויים להיות גורם בהכרעות אשר להן השפעה רבת עוצמה ומתמשכת על חייהם של אנשים (לדוגמה - שחרור על תנאי, מתן פסק דין, אשפוז, כשירות לעמוד למשפט וכשירות לאחזקת ילד).

שימוש הולך וגובר נעשה במחשבים, על מנת להפעיל, לחשב תוצאות ולפרש מבחנים. מכל מקום, בין אם הקלינאים מבצעים את המבחנים בעצמם, בין אם הם משתמשים בשירותיו של מחשב או עוזר, הרי שעליהם מוטלת האחריות לבחירת המבחנים המתאימים לפרט ספציפי ולוודא כי תנאי המבחן נאותים. בנוסף, על הקלינאי מוטלת האחריות להבטיח סודיות, להעריך את מידת התאמתן של תוצאות המבחן ופירושי המחשב, ולהביאן, במונחים מתאימים, לידיעת הלקוח וליתר הנוגעים בדבר.

רצוי כי הפירושים הקליניים יתקבלו באמצעות בדיקת מספר מקורות שהתקבלו באופן בלתי תלוי.

תחום נוסף הדורש תשומת לב מיוחדת הוא תחום המבחנים ה"פתוחים". למשל, מבחני השלמת משפטים, ראיונות ותצפיות על התנהגות בסביבה טבעית או מלאכותית. יש להבטיח כי הכלים מופעלים באווירה תומכת, אך ללא מתן הצעות או רמזים אשר ישפיעו על התגובה הניתנת, אלא אם כן עידוד כזה תוכנן באורח ספציפי כחלק מהמבחן. כמו כן, את הקטגוריות המשמשות להערכת התגובות יש להגדיר באורח ברור, ויתכן ויהא צורך בהכשרה מיוחדת לקידוד התגובות.

סטנדרט 7.1 אל לקלינאים לטעון, כי פירושי תגובות נבחנים מבוססים על עדות אמפירית לתקפות, אלא אם כן אמנם קיימת עדות כזו.

7.2 סטנדרט כאשר נאמדת תקפות באמצעות מדידת ההתאמה בין תוצאות המבחן והאבחנות הקליניות, יש להגדיר או לזהות בשימת לב את המונחים והקטגוריות האבחוניים המשמשים בהקשר זה, ויש לציין מה היתה השיטה ששימשה לאבחון. במידה והאבחון התבצע בהתבסס על שיפוט, יש לכלול מידע לגבי ההכשרה, הנסיון והמעמד המקצועי של השופט, ומידע לגבי טיבם ועוצמתם של קשרי השופט עם הנבחנים.

7.3 סטנדרט כאשר קיים צורך באבחון דיפרנציאלי, על המשתמש לבחור - במידת האפשר - מבחן אשר לגביו קיימת עדות ליכולתו להבחין בין קבוצות האבחון הרלוונטיות. הערה: הבדלים בין ממצעי קבוצות, ומשמעותם הסטטיסטית, מספקים מידע מועיל בנוגע לתקפות, גם למטרות אבחון פרטים. מידע נוסף עשוי לכלול טבלה, המורה על מידת החפיפה בין התפלגויות חיזוי בקרב קבוצות קריטריונים שונות.

7.4 סטנדרט על הקלינאים לחלוק עם לקוחותיהם את המידע לגבי תוצאות המבחן ופירושיהן, כמו גם את המידע לגבי טווח הטעות של פירושים אלה, כל זאת במידה והמידע מועיל ללקוח. מידע כזה יש להביא לידיעת הלקוח (או לבא-כוחו) בשפה שהוא יהיה מסוגל להבין.

7.5 סטנדרט קיים צורך בעדות לתקפות-קריטריון לגבי אוכלוסיות הדומות לאוכלוסיה שעליה יופעל המבחן, כאשר ההמלצות או ההכרעות מוצגות כבעלות בסיס אקטואלי ובסיס קליני.

8. מבחנים חינוכיים ופסיכולוגיים בבתי ספר

(Educational testing and psychological testing in the schools)

רקע

מבחנים במערכת החינוך זוכים לתפוצה נרחבת. פרק זה דן בארבעה תחומי יישום: תוכניות מבחנים בבתי ספר, מבחנים לצורך מתן תעודות השכלה ולימוד, ברירה לצרכים של מערכת החינוך, וחינוך מיוחד.

תוכניות למבחנים בבתי ספר School Testing Programs

תוכניות מבחן בבתי ספר הינן חלק אינטגרלי ממערכת החינוך היסודית והתיכונית. בכל שנה, ניגשים אלפי תלמידים, במאות בתי ספר, למבחנים קבוצתיים סטנדרטיים - מבחני יכולת ומבחני הישג. תוכניות מבחן אלה מספקות מידע לגבי רמות היכולת וההישגים של התלמידים כפרטים, ולגבי קבוצות של תלמידים המשתייכות לרמות שונות. קבוצת המשתמשים הקשורה לתוכניות המבחנים בבתי הספר, היא אולי הגדולה והמגוונת ביותר מכל קבוצת המשתמשים במבחנים - החל מתלמידים ומורים וכלה בבוני תכניות לימודים ומעצבי מדיניות. למשתמשים אלה לסוגיהם, יש לספק סיוע בפירוש תוצאות המבחן.

מבחנים לצורך מתן תעודות, בבתי ספר יסודיים ותיכוניים

ציוניהם של תלמידים במבחנים לקבלת תעודות חינוכיות, משמשים באופן בלעדי או בשיתוף עם קריטריונים אחרים, לקבלת הכרעות הנוגעות לסיום הלימודים בבית הספר התיכון או למעבר תלמידים משנה לשנה, לסיווג תלמידים לצורך שיפור הישגיהם, להערכת מידת היעילות של הלימודים בבית הספר, להקצאת משאבים ולהערכת תפקוד המורים.

מרבית התוכניות למבחנים המשמשים לצורך מתן תעודות חינוכיות, משמשות בציוני תוך מוגדרים מראש, על מנת להבחין בין ציוני "עובר" לציוני "נכשל". ציון החתך הופך לציר מרכזי בתהליך קבלת ההכרעות. מחקרים הורו, כי יתכנו אי-התאמות ניכרות מאד בין ציוני החתך המתקבלים משיטות שונות לחישוב נקודות תוך. לפיכך, יש לבחון בזהירות את מהימנותן ותקפותן של הכרעות ומסקנות, המבוססות על ציוני תוך של מבחנים לצורך מתן תעודות חינוכיות.

מיון לצרכי חינוך

תנאי הקבלה משתנים במידה ניכרת בין מוסדות שונים להשכלה, כגון אוניברסיטאות, מכללות ובתי ספר מקצועיים. הדרישות שונות לא רק בין מוסד למוסד, אלא גם בין תחום התמחות אחד למשנהו בתוך מוסד אחד. קבלה על בסיס תחרותי, הכרוכה לרוב במבחנים, נפוצה יותר ויותר גם בבתי ספר תיכוניים.

כאשר מתבצע מיון במערכת החינוך הרי בדרך כלל הוא מתבסס על מספר סוגים של מידע. למשל: עותקי תעודות, ממוצע ציונים, דירוג בכיתה מבחינת הישגים ותוצאות מבחנים.

עדות תלוית-קריטריון היא הגישה הנפוצה ביותר לתיקוף בהקשר למערכת מיון מועמדים, חרף העובדה, כי עדויות תלויות-מבנה ותלויות-תוכן, חשובות אף הן. יש לציין כי יש מוסדות הממיינים תלמידים בכוונה לקדם מטרות אחרות מלבד השגת הצטיינות אקדמית, ויש להביא עובדה זו בחשבון כאשר מעריכים תקפות המבחן.

חינוך מיוחד

במערכת החינוך המיוחד, מסייעים מבחנים בזיהוי אופיים, מקורם ומידת עוצמתם של קשייו של נבחן יחיד בלימודים או בהסתגלות לבית הספר. בסופו של דבר, משמשות תוצאות המבחן לעיצוב תוכנית לימודים אישית לתלמיד, אשר כוללת לעיתים את העברתו לבית ספר או כיתה מיוחדת. במערכת החינוך המיוחד נעשה שימוש במגוון רחב של מבחנים, הכוללים מבחני יכולת למידה, מבחני הישגים, מבחני מיומנות ספציפית כגון מיומנויות מוטוריות-ויזואליות), מבחני דיבור ושפה, מבחני שמיעה וראייה, שאלוני אישיות, תצפיות התנהגותיות, וטכניקות השלכה - ורשימה זו אינה ממצה.

במערכת החינוך המיוחד, המבחנים נבחרים, מופעלים ומתפרשים על ידי פסיכולוגים של בתי הספר, מורים, מורים לחינוך מיוחד, ואנשי מקצוע אחרים, כגון מומחים לליקויי דיבור ומומחים לתיראפיה גופנית. קבוצה מגוונת זו של המשתמשים במבחנים, כוללת אנשי מקצוע ברמות שונות של הכשרה, בתחום המדידה וההערכה, וברמות שונות של מומחיות טכנית בעריכת מבחנים. לאור זאת חשובה ביותר עמידה בדרישות הסטנדרטים המקצועיים למבחנים במערכת החינוך המיוחד. עמידה קפדנית בדרישות נחוצה במצבים בהם הכרעות המבוססות על המבחנים, יהיו בעלות השפעה מכרעת על חינוכו ומצבו של הילד, מקום בו עומד על הפרק שיבוצו בכיתה או בית ספר מיוחדים.

מתן שירותים למערכת החינוך הרגילה

פסיכולוגים בבתי הספר, עושים שימוש במגוון רחב של מבחנים המופעלים על פרטים, במסגרת מתן שירות לתלמידים בכיתות לימוד רגילות. שירותים אלה נועדו לתלמידים אשר יתכן ואינם זקוקים לחינוך מיוחד, אולם סובלים מבעיות התנהגותיות, נפשיות ו/או לימודיות. תוצאות מבחנים הינן מקור חשוב לנתוני הערכה לגבי תלמידים אלה.

את הקשרים בין מדדי חיזוי לבין קריטריונים המשמשים לצורך ברירה, במסגרת קבלת תלמידים למוסד חינוכי, יש לתאר באמצעות משוואות רגרסיה וטעויות תקן של אומדים, או באמצעות טבלת ציפיות, בנוסף למקדמי מתאם.

אל למשתמשים במבחנים לטעון כי קיימת עדות אמפירית לגבי הקשרים בין ציוני מבחנים מסוימים, תוכניות חינוכיות והישגי תלמידים, אלא אם כן קיימת עדות כזו.

סטנדרט 8.1 על האחראים על תוכנית מבחנים בבתי ספר, להבטיח כי הבוחנים המעבירים את המבחנים יקבלו הדרכה טובה בהליכי העברה נאותה של המבחן, וכי יבינו את החשיבות שבמילוי צמוד של הוראות הבוחן.

סטנדרט 8.2 על האחראים על תוכניות מבחנים בבתי הספר, להבטיח כי אנשי הסגל המשתמשים בתוצאות המבחנים במסגרת בית הספר, יקבלו הדרכה מתאימה באשר לשיטות הנאותות לפירוש תוצאות המבחנים.

סטנדרט 8.3 במידה וציוני מבחן משמשים לבדיקת הבדלים בין הכושר השכלי וההישגים של תלמיד מסוים, יש להביא בחשבון את רמת המהימנות (או טעות התקן) של הפרש הציונים.

סטנדרט 8.4 כאשר המבחן מיועד לאשר גמר לימודים ברמה מסוימת, מעבר משנה לשנה או סיום הלימודים בבית הספר, יש לתאר בפירוט הראוי הן את תחום המבחן, הן את תחום ההוראה ברמה הנתונה, כך שניתן יהיה להעריך את מידת ההתאמה בין תחום המבחן ועולם התוכן ההוראתי.

סטנדרט 8.5 עבור מבחן המפותח ע"י המדינה, ומיועד לאישור מעבר תלמידים משנת לימודים אחת לשניה, גמר הלימודים או סיווג, יש לפתח ולהפיץ מדריכים למשתמשים או דו"חות טכניים, למען המשתמשים בו.

סטנדרט 8.6 על ציוני מבחנים הקשורים במתן תעודות לימוד יש לדווח מיד לכל הצדדים הנוגעים בדבר, לרבות תלמידים, הורים ומורים. על הדיווח לכלול את תיאור המבחן, מה נמדד בו, המסקנות וההכרעות המתבססות על ציוני המבחן, הציון שהתקבל, מידע על אופן פירוש הציון, וכל ציון חתך המשמש לסיווג.

סטנדרט 8.7 כאשר מבחן משמש לקבלת הכרעות לגבי מעבר תלמיד משנה לשנה או סיום לימודיו, נדרשת עדות לכך כי המבחן מכסה אך ורק את המידע הספציפי או הכללי, המיומנויות ותחומי היכולת אשר לתלמיד ניתנה אפשרות ללמד.

סטנדרט 8.8 לתלמידים אשר נדרשים להפגין שליטה במיומנויות מסוימות או ידע, (באמצעות מבחנים), יש לספק כמה הזדמנויות להפגין את מיומנותם. ("מועד ב").

סטנדרט 8.9 במערכת בתי הספר היסודיים או התיכוניים, אין לקבל אוטומטית הכרעה או אפיון, אשר להם תהיה השפעה מרכזית על הנבחן, על בסיס תוצאת מבחן יחידה. אנשי המקצוע שבידיהם ההכרעה, צריכים להביא בחשבון מידע רלוונטי נוסף. לדוגמה, אין לשבח תלמיד בכיתות או בתי ספר מיוחדים, אך ורק על בסיס תוצאות מבחן יכולת. בהכרעות אלה יש להעניק משקל אף למידע אחר, הקשור ליכולת הלמידה של התלמיד.

9. שימוש במבחנים לצרכי ייעוץ
(Test use in counselling)

רקע

כאשר נעשה שימוש במבחן לצורך ייעוץ, הנבחן הינו גם המשתמש העיקרי בתוצאות המבחן, אולם מאחר ובדרך כלל הנבחנים אינם בעלי ידע או מיומנות בעריכת מבחנים הם נזקקים לסיוע והדרכה מיועצים. יועצים עוסקים בהערכת מגוון מאפיינים של הפרט, במגמה לסייע בהכוונה בתהליכי קבלת החלטות המתבצעות על-ידיו. לשם כך, מסתייעים היועצים באמצעים רבים ושונים, כגון מבחנים, ראיונות, תצפיות וכד'. כל האמצעים הללו נחשבים למבחנים ודורשים עדות לתקפות ומהימנות.

מטרתו העיקרית של מבחן לצרכי ייעוץ היא לסייע ברכישת מידע רלוונטי על הנועץ, ודיווח על מידע זה, בלוויית פירושים נאותים, כך שנועצים בעלי סוגי רקע שונים, יוכלו להיעזר במידע בעת קבלת הכרעות חשובות בתחום החינוך, הקריירה, והתחום האישי.

יועצים עובדים במסגרות שונות - מיעוץ פרטי ועד יעוץ למוסדות גדולים כגון משרדי ממשלה ובתי ספר. על מפתחי המבחנים להביא בחשבון שקיימים מצבים בהם הנבחנים עצמם יקבלו מעט מאד עזרה מקצועית בפירוש הציונים, ולפיכך, עליהם לספק, יחד עם המבחנים חומר פרשני והוראות, אשר יהיו מקיפים וקלים להבנה. קיימות מערכות הכוון המסתייעות במחשבים; גם כאן - כמו בכל המבחנים ה"קלאסיים" - על מפתחי המבחן לפרט את הליכי פיתוח המבחן ועדויות תומכות לתוקף ולמהימנות.

סטנדרט 9.1 על יועצים לבחון נתונים טכניים, ולפתח רציונל להחלטה, האם להשתמש בנורמות משולבות או מופרדות, בנוגע לנשים וגברים, ולפרט את הנ"ל בדו"חות שיתנו לנבחנים.

סטנדרט 9.2 במידה ומו"ל מאגד מבחנים האמורים לשמש במשולב לצרכי ייעוץ, על היועץ לערוך ביקורת על המדריך ולבדוק את הרציונל לשילוב המסוים של המבחנים, ואת ההצדקה לפירוש היחסים בין הציונים.

סטנדרט 9.3 על יועצים לבדוק אם קיים במדריכי המבחן מידע על האופן בו מתפלגות האופציות המוצעות או המרומזות לקריירה (לדוגמה - התוכנית המקצועית או העיסוק המוצע על פי הציון הגבוה ביותר במבחן), לגבי מדגמים של נבחנים טיפוסיים, מכל מין, ומדגמים של קבוצות של גזע או מוצא אתני.

סטנדרט 9.4 על יועצים לבדוק את המדריך למבחן הניתן לנבחנים, ולהזהיר כראוי את הנבחנים, כי אל לחס לחסתמך אך ורק על תוצאות המבחן בעת קבלת החלטות לגבי תכנון מהלכים בחייהם. על היועצים לעודד הפקת מספר רב של הערכות לגבי תחומי יכולתו, מיומנותיו החברתיות והנטיות של פרט (עקרון ההצלחה של עדויות), על היועצים לעודד נבחנים לשקול מידע רלוונטי אחר, לגבי מיומנות אישיות וחברתיות, ערכים, תחומי ענין, הישגים, התנסויות, ואף לשקול תוצאות ממבחנים אחרים, כמו גם תצפיות.

סטנדרט 9.5 על היועצים לבדוק את המדגמים עליהם בוססו הנורמות של מדדי יכולת ונטיות, וכן מבחנים אחרים רלוונטיים, ולברר מהי מידת התאמתם לנועצים שלהם. על יועץ להביא בחשבון את גילו, התנסויותיו ורקעו של הלקוח, כפי שהם משווים עם מאפייניהן של קבוצות של נורמות, אשר עליהן מבוססות התוצאות.

10. מבחנים תעסוקתיים
(Employment testing)

רקע

בתחום התעסוקה, ניתן להשתמש במבחנים בצירוף עם מידע נוסף, על מנת לערוך תחזיות או לקבל החלטות לגבי עובדים. נושאים רלוונטיים למודלים של תורת קבלת ההחלטות, הינם ישימים באורח ישיר לתחום זה. לדוגמה - השיעור היחסי של טעויות החלטה שונות, תועלת בהכרעה נכונה (וחשיפוטים הערכיים אשר כרוכות הכרעות תועלת אלה), השוואות בין אסטרטגיות חלופיות, ומידת היכולת להשיג מידע מוקדם ולהשתמש בו כדי להפחית אי-וודאות - כל אלה הינם שיקולים רלוונטיים. סוגי ההכרעות העיקריים אשר המבחן עשוי לאפשר במסגרת זו הם ברירה, סיווג, קידום עובדים וכד'.

הכרעות לגבי קידום נבדלות מהכרעות ברירה בעיקר בכך שבמקרה של הכרעת קידום, כבר מצוי מידע בארגון לגבי תפקודו בעבר של הפרט, אשר קידומו נשקל. על מעסיק יהיה להחליט באיזו מידה יש לשלב מידע לגבי רמת התיפקוד במסגרת המשרה, עם המידע המתקבל מהמבחן.

שימוש יעיל במבחן, יכול לתרום רבות להצלחתם של פרטים וליחס הוגן כלפיהם במסגרות תעסוקתיות. מקרב החלופות הקיימות, המבחנים הם עזרי החלטות הברירה התקפים ביותר, והגורמים למידה הפחותה ביותר של אפליה.

סטנדרט 10.1 על עדות לתקפות-קריטריון במצב נתון מותר להסיק ממחקר תקפות קודם יחיד, אך ורק אם המחקר הקודם הזה בוצע בתנאים טובים (מדגם גדול, קריטריון טוב) ובתנאי שלמצב הנוכחי יש קוים משותפים רבים עם המצב הקודם.

הערה: קוים משותפים רבים משמעו, כי מטלות המשרה ודרישותיה, או המבנים הפסיכולוגיים המונחים בבסיסה, מראים דמיון רב (כפי שנקבע הדבר באמצעות ניתוח עיסוקים), כי החזאי זהה (לדוגמה - המבחן זהה) וכי רמת הביצועים הנדרשת זהה.

סטנדרט 10.2 במידה ונעשה שימוש במבחנים על מנת לקבל הכרעות שיבוץ (לדוגמה - דפוס תוצאות החזאי ישמש לאיוש דיפרנציאלי של משרות), יש לתעד עדות לחיזוי דיפרנציאלי בקרב עיסוקים או קבוצות עיסוקים.

סטנדרט 10.3 יש להסביר היטב את הרציונל לרלוונטיות של קריטריון. על החסבר לכלול את תיאור המשרה הנידונה, ואת הדרך בה נקבעה מידת הרלוונטיות. **הערה:** כללית, הרלוונטיות הנשפטת של הקריטריון, אמורה לנבוע ישירות מהקשר בין צורות ההתנהגות או אלמנטים אחרים אותם מעריך הקריטריון, לבין מטרות הארגון. מטרת מדידת הקריטריון לביצוע העבודה היא לייצג את הגורמים העיקריים המשקפים את תרומתו של פרט לארגון. מכל מקום, יש לשקול את המידה בה המבחן הינו מקיף (comprehensive) כנגד טיב המדידה.

סטנדרט 10.4 מקום בו עדות לתקפות-תוכן אמורה להיות גורם התומך בשימוש במבחנים לצורך ברירה או קידום, יש להדגים קשר חזק בין תוכן המבחן ותוכן העיסוק. תיקוף תוכן יש לבסס על הגדרה יסודית וברורה של תחום התוכן הנידון. לצורך ברירה, סיווג וקידום במסגרת תעסוקתית, יש לבסס את אפיון התחום על ניתוח עיסוקים.

סטנדרט 10.5 במידה ותקפותו של מבחן ברירה לצרכי עיסוק מסוים, מבוססת על עדות תלוית-תוכן, מותר להכליל לגבי המבחן במצב חדש, אך ורק במידה והגורמים הקריטיים בתוכן העיסוק הינם זהים עקרונית (כפי שיקבע באמצעות ניתוח עיסוקים), ובמידה ורמתו של חומר הקריאה של המבחן אינה עולה על הרמה הנדרשת מקבוצת המועמדים החדשה והעיסוק החדש, ובמידה ולא קיימים מאפיינים בולטים במצב החדש, אשר עשויים לשנות באורח ניכר את משמעותו המקורית של חומר המבחן.

סטנדרט 10.6 במידה ועדות תלוית-מבנה אמורה להיות הגורם התומך העיקרי בתקפות לצורך ברירה, אזי יש לבסס שני קשרים. ראשית, יש צורך בעדות לתקפותו של המבחן כמדד למבנה; שנית, יש צורך בעדות לתקפותו של המבנה כאמצעי המגדיר את הגורמים העיקריים בביצוע העיסוק. קיים צורך ברציונל מושגי (קונצפטואלי) ברור לקשר זה. יש להגדיר בזהירות הן את המבנה, הן את גורמי העיסוק שאליהם הוא אמור להיות קשור. דפוס עקבי של התוצאות, אמור להצביע על היחסים המשוערים הללו. אין להשתמש אך ורק בשיפוט של מומחים, על מנת לבסס את הטענה לעדות תלוית-מבנה.

סטנדרט 10.7 יש לספק הסבר ברור לכל בסיס טכני עבור כל ציון תתן ששימש לקבלת הכרעות אישיות. אין לקבוע ציוני תתן אך ורק על בסיס ההמלצות המופיעות במדריך המבחן.

חלק שלישי

סטנדרטים ליישומים מיוחדים

Standards for Particular Applications

11. מתן רישיונות ותעודות מקצועיות ותעסוקתיות
(Professional and occupational licensure and certification)

רקע

המטרה העיקרית של קיום מסגרת למתן רישיונות ותעודות, היא להגן על הציבור. דרישות רישוי נכפות על מנת להבטיח כי לבעלי הרישיונות יהיו הידע והמיומנות הנדרשים במידה מספקת על מנת לבצע פעולות מקצועיות חשובות בבטיחות וביעילות. מטרתה של מערכת מתן התעודות היא לספק לציבור (לרבות מעסיקים וסוכנויות ממשלתיות) מנגנון מהימן לזיהוי בעלי מקצוע אשר עומדים בסטנדרטים מסוימים.

הבחינות המשמשות למתן תעודות או רישיונות, מפותחות בדרך כלל ע"י ועדות המורכבות מאנשי המקצוע. קיים מגוון רחב של סוגי בחינות, ועליהן נמנים מבחנים מסורתיים הכוללים שאלות בחירה (multiple-choice), בחינות בכתב, בחינות בע"פ ומבחני ביצוע (לדוגמה, הטסת מטוס או סתימת שן).

נושאים של תקפות, הנידונים בחלקים אחרים של "סטנדרטים", הינם רלוונטיים גם למבחנים לצורך מתן רישיונות ותעודות. אך יש לערוך מספר הבחנות חשובות. כך, למשל, בדיקות של תקפות תלוית-קריטריון הינן בעייתיות יותר בהקשר של מתן תעודות ורישיונות, מאשר במסגרות תעסוקתיות רבות. לא כל בעלי הרישיונות והתעודות מתקבלים בהכרח לעבודה; אלו שהתקבלו עשויים לתפקד במגוון רחב של משרות ולעבוד עם מעסיקים רבים ושונים, והלקם עשויים אף להיות עצמאים. לעיתים קרובות, הופכים גורמים אלה את המתקרים המסורתיים הנערכים לשם איסוף עדויות לתקפות-קריטריון, לבלתי אפשריים לביצוע.

הקושי בעריכת מחקרים לתקפות תלוית-קריטריון אינו מפחית, מכל מקום, מחשיבותה של התקפות, המהווה גם בתנאים אלה את הנושא המרכזי. על המשתמשים במבחנים לפתח את הבסיס העדוטי לתמיכה בשימוש מסוים. מכל מקום, לגבי מתן רישיונות ותעודות, יש להסתמך בעיקר על עדות תוכן, קרי, עדות להתאמתו של המבחן לעולם התוכן.

תהליך קביעת ציון חתך לבחינות לצורך מתן תעודות ורישיונות, שונה מזה המשמש למיון תעסוקתי וברירת סטודנטים (למשל). נקודת החתך נקבעת לפי רמת הביצועים הנדרשת, ולא לפי נתוני היצע וביקוש.

בעוד שמבחנים תעסוקתיים עשויים למדוד בצורה נאותה את כושרו של הפרט ללמוד עיסוק מסוים, אנשים הניגשים למבחנים לצורך מתן תעודות ורישיונות הם בדרך כלל אנשים שהשלימו את הכשרתם, והינם מבקשים להיחשב כבעלי כישורים בתחום עיסוק, ולא בתפקיד ספציפי. להבחנה זו השלכות חשובות על התוכן שעל מבחנים למתן רישיונות ותעודות לכסות.

סטנדרט 11.1 יש להגדיר בבחירות את תחום התוכן (content domain) האמור להיות מכוסה ע"י מבחן לצורך מתן תעודות ורשיונות, ולהסבירו במונחים של השיבות התוכן לביצוע יעיל וכשיר (קומפוטנטי) של מטלות עיסוק. יש לספק רציונל לתמיכה בטענה כי הידע או המיומנות המוערכים, אכן נדרשים לביצוע כשיר של העיסוק, וכי הם עקיבים עם המטרות שבעבורן נוצרה התוכנית למתן רשיונות או תעודות.

הערה: ניתוח עיסוקים עשוי לספק את הבסיס העיקרי להגדרת תחום התוכן. במידה ונעשה שימוש בבחינה יחידה לצורך מתן תעודות או רשיונות, לאנשים המועסקים במגוון מסגרות והתמחויות, הרי שיש לנתח מספר עיסוקים, כשהדגש מושם על ידע ומיומנות הנדרשים על מנת להגן על הציבור.

אין לכלול בבחינה לצורך מתן רשיון, מיומנות אשר אינה קשורה ישירות למטרת הרישוי. אם כישורים מסויימים חשובים להצטיינות בעיסוק, אך היעדרם אינו מהווה הכרה או סכנה לציבור - יש להשמיטם ממבחן הרישוי.

סטנדרט 11.2 יש לספק אומדים למהימנותן של הכרעות במסגרת מתן רשיונות ותעודות. (ראה פרק המהימנות).

סטנדרט 11.3 במידה ונבחנים אשר נכשלו במבחן מבקשים זאת, יש לציין בפניהם את הציון שקיבלו במבחן, ואת הציון המינימלי הנדרש למעבר. יש לספק לנבחנים מידע לגבי ביצועיהם בחלקים של המבחן, אשר בעבורם מופקים ציונים או דו"חות נפרדים, ומשמשים בתהליך קבלת ההחלטות.

סטנדרט 11.4 יש לדווח לנבחנים על הכללים וההליכים המשמשים לשילוב ציונים עם הערכות אחרות, על מנת לקבוע את התוצאה הכללית. עדיף לעשות זאת לפני הפעלת המבחן.

12. עריכת מבחנים למיעוטים לשוניים
(Testing linguistic minorities)

רקע

לגבי אדם שהעברית אינה שפת אמו, כל מבחן הניתן בשפה העברית מהווה בחלקו מבחן שפה או מבחן של ידיעת קרוא וכתוב. על כן, עריכת מבחנים לפרטים אשר לא נחשפו במידה מספקת לעברית, מהווה אתגר מקצועי מיוחד. תוצאות המבחנים עלולות שלא לשקף במדויק את תחומי היכולת והכשרים הנמדדים, במידה וביצוע המבחן תלוי במידת בקיאותם של הנבחנים הללו בעברית. לפיכך, יש להקדיש תשומת לב מיוחדת להיבטים רבים של פיתוח המבחן, הפעלתו, פירושו, וקבלת הכרעות על פיו. כן חשוב להביא בחשבון את רקעו של הפרט באשר לשפה, בעת פיתויו, בחירה והפעלת מבחנים, ובעת מתן פירוש לביצועים במבחן.

מבחנים בשפות זרות

לעיתים, יאה מן הראוי לערוך את המבחן בשפתם של הנבחנים. מכל מקום, קיימות מספר מכשולות מהן יש להימנע במבחנים דו-לשוניים. לא ניתן להניח כי תרגום ישיר מניב גרסה של המבחן, המקבילה למבחן המקורי בתכונות הסטטיסטיות או הפסיכומטריות. כך, למשל, מילים בשתי השפות, אשר דומה שהן קרובות זו לזו מבחינת משמעותן, עלולות להיות שונות מאד זו מזו מבחינות אחרות, החשובות למטרות המבחן.

מבחני בקיאות בשפה

מבחני בקיאות בשפה נדרשים על-מנת לסייע בקליטתם של עולים מרקעים שונים וגילים שונים. צורך זה מודגש בחינוכם של ילדים, אולם הוא חשוב גם בחינוך מבוגרים. בשל העובדה, כי תלמידים אמורים לרכוש בקיאות בעברית, ברמה הנדרשת לגילם ולשנת הלימוד שלהם, הרי שנדרשים מבחנים המתאימים להערכת התקדמותם. המבחנים עשויים לבדוק מיומנויות שונות כגון כתיבה, דיבור, האזנה. ככלל, קיימים קשיים מיוחדים באשר לשימוש במבחנים, בכל הנוגע לפרטים אשר לא זכו להזדמנות המתאימה ללמוד את שפת המבחן. יתכן ויהא זה רצוי לעשות שימוש במגוון רחב מהרגיל של מבחנים ותצפיות, במידה וציוני המבחן משמשים להכרעות חשובות.

עריכת מבחנים לפרטים בבתי ספר

בעת עריכת מבחן יש לזכור, כי התרבות ממנה בא הילד קובעת במידה רבה אם התנהגות מסוימת תיחשב נאותה או חריגה. כך, למשל, קבוצה מסוימת עשויה לראות בדיבור מהיר או מלל רב התנהגות מחוצפת, בעוד שקבוצה אחרת עשויה לראות בדפוס דיבור אלה אינדיקציה ליכולת שכלית גבוהה או לידידותיות. ילד רגיל שמוצאו בתרבות מסוימת, הנשפט במונחים של תרבות אחרת, עלול להיחשב בטעות לשקט, מופנם או בעל יכולת שכלית נמוכה. הפירושים ותוכניות הטיפול הנובעים מכך, עשויים להיות בלתי תקפים ואף מזיקים בכוח, לפרט הנבחן.

סטנדרט 12.1 מבחנים עבור פרטים שהעברית אינה שפת אמם, יש לעצב באורח שימזער את האפשרות לפגיעה במהימנות והתקפות של המבחנים, העלולה לנבוע מהבדלים בין השפות.

הערה: דוגמאות לאופנים בהם ניתן לעמוד בסטנדרט זה:

- (1) ניתן להפעיל את המבחן באמצעות סגל שהוכשר במיוחד למגע עם בני קבוצה זו.
- (2) במצבים בהם מתבצעת הערכה של פרטים, כגון מבחנים קליניים, רצוי להשתמש בשפת האם של הנבחן או בשתי השפות, על מנת להפיק את התגובות למבחן בצורה היעילה ביותר.
- (3) בעת הערכת פרטים, יש להביא בחשבון התנהגות מילולית הנחשבת לנאותה מבחינה חברתית בתרבותו של הנבחן.

סטנדרט 12.2 שינויים לשוניים עליהם ממליצים מו"לי מבחנים, יש לתאר באורח מפורט במדריך המבחן.

סטנדרט 12.3 מקום בו מתורגם מבחן משפה לשפה או מניב אחד למשנהו, יש לבסס באופן מחקרי את מהימנותו ותקפותו של המבחן, לצורך הפעלתו על הקבוצות הלשוניות הנבחנות.

סטנדרט 12.4 יש לוודא כי במבחני מיון למשרות, מתן רשיונות או תעודות לימוד, רמת הבקאות בשפה העברית הנדרשת במבחן, לא תהא גבוהה מהרמה הנדרשת למקצוע או משלח היד הרלוונטיים.

סטנדרט 12.5 אין להסיק על מידת הבקאות הכללית בשפה העברית אך ורק על פי מבחנים הדורשים מיומנות לשונית יחידה.
הערה: לדוגמה, מבחן לבדיקת אוצר מילים, שהוא מטיפוס בחירה, אינו מצביע על מידת הבנת השפה של הנבחן בצורתה המדוברת או כושר הביטוי של הנבחן בשפה זו.

13. עריכת מבחנים לאנשים מוגבלים

(Testing people who have handicapping conditions)

רקע

מבחנים נערכים לאנשים הסובלים ממוגבלות, במגוון של מצבים ולמטרות שונות. קיימים מספר אופנים לעריכת שינויים במבחנים ובהליכי הפעלת מבחנים, המאפשרים לאנשים הסובלים ממוגבלות מסוימת לגשת למבחנים שפותחו במקור עבור האוכלוסייה הכללית. כך, למשל, כאשר הנבחנים סובלים מליקויי ראייה, יתכן וחוברת המבחן תודפס באותיות גדולות, בדפוס רגיל באיכות גבוהה, או בכתב ברייל, או שהמבחן יוקלט או יוקרא בקול לנבחן.

יתכן ויהא צורך אף בשינוי השיטות לתייעוד התגובות. בדרך כלל, נעזרים נבחנים אשר אינם מסוגלים לרשום את תגובותיהם לשאלות המבחן, באדם הכותב או מסמן את התשובות, או ברשם-קול, מכונת הדפסה או מדפסת ברייל. יתכן ויהא צורך בשינוי המבחן, על מנת לאפשר לנבחן להצביע על התשובה בה ברר. במקרים רבים יהיה צורך בהארכת זמן המבחן.

לעיתים, נדרשים שינויים בתוכן המבחן, עבור נבחנים הסובלים מליקויי שמיעה או ראייה. כך, למשל, נכון הדבר לגבי פריטים המשתמשים בגירוי ויזואלי על מנת למדוד ידע הנרכש באמצעות חושים אחרים. בשל העובדה, שהחלפות פריטים עשויות לגרום שינויים במאפיינים אחרים של הפריטים, יש לנסות את הפריטים המתוקנים מראש, לפני שימשו בהפעלות מבחנים.

מבחנים מילוליים עלולים להיות בעיה חריפה לגבי נבחנים חרשים. מכל מקום, מציאת המבחנים הבלתי מילוליים המתאימים, למדידת אותם תחומי היכולת או לחיזוי אותה ההתנהגות, עשויה להיות משימה קשה ביותר. למרות שזוהי בעיה מתחום המבחנים, הרי שהיא משקפת קשיים בסיסיים יותר, בהבנת טבעם של תחומי היכולת; אילו תחומי יכולת נדרשים למצבים מסוימים, ואילו תחומי יכולת קיימים יכולים לפצות על תחומי יכולת נגומים, בנסיבות מסוימות.

במרבית המקרים של שינוי מבחן עבור אנשים נכים או מוגבלים, קיימת בעיה, לאמור, שבדרך כלל נערך מספר מצומצם מאד של מתקרים אמפיריים בנושא השפעות של שינויים מיוחדים על תוצאות המבחנים, או על מהימנותם ותקפותם. הסיבה העיקרית לכך היא מספרם המצומצם יחסית של הנבחנים הנכים.

חרף העובדה, שככלל, עריכת שינויים במבחנים עבור אנשים מוגבלים, נתפסת כפעולה רצויה, מתעוררת השאלה, האם בודקים מדי המבחן המתוקן את אותם המבנים. האם משפיעים שינויים באמצעי ההתבטאות על התפקוד הקוגניטיבי ועל משמעות התגובות?

מכל ההיבטים הנוגעים לעריכת מבחנים לאנשים מוגבלים, ההיבט של דיווח על ציוני המבחן עורר את המחלוקת החריפות ביותר. מפתחי מבחנים רבים טענו, כי דיווח על תוצאותיו של מבחן שהופעל באורח לא סטנדרטי, ללא מתן זיהוי מיוחד, מפר את כללי המקצוע, מטעה את המשתמשים במבחן, ואולי אף גורם נזק לנבחנים הנכים, אשר תוצאות מבחניהם אינן משקפות נאמנה את יכולתם. מאידך, זיהוי נבחנים כנכים, שולל מהם את האפשרות להתחרות עם נבחנים אחרים על בסיס שוויוני. מחלוקת זו לא נפתרה.

סטנדרט 13.1 על אנשים העורכים שינויים במבחן עבור נכים, להיעזר בפסיכו-מטריקאי. בנוסף, עליהם לוודא כי מצוי בידיהם מידע לגבי השפעותיהם של תנאי מוגבלות שונים על ביצוע המבחן.

סטנדרט 13.2 עד לביצוע מחקרי תקפות למבחנים עבור אנשים הסובלים ממוגבלויות ספציפיות, על מו"לי מבחנים לפרסם הצהרות אזהרה במדריכים או במקורות אחרים, הנוגעות למידת הבטחון בפירושים המתבססים על תוצאות מבחנים כאלה.

סטנדרט 13.3 נוסחי מבחנים שעברו שינויים לצורך הפעלתם עבור אנשים הסובלים מנכות קשה, יש לנסות ראשית על אנשים הסובלים מנכות דומה, על מנת לבדוק את התאמתם של השינויים ואת המידה בה הם ניתנים לביצוע.

סטנדרט 13.4 בכל עת אפשרית, יש להשתמש בהליכים אמפיריים על מנת לקבוע מגבלות זמן רצויות לנוסחים מתוקנים של מבחנים לנבחנים מוגבלים, תחת לאפשר לנבחנים אלה באופן סתמי זמן כפול מהזמן הסטנדרטי למבחנים. כאשר מורחבים גבולות הזמן, יש לחקור את גורם העייפות כגורם בעל חשיבות פוטנציאלית.

סטנדרט 13.5 כאשר הדבר הינו בר ביצוע, על המוסד או המו"ל אשר עורכים שינויים במבחנים, לחקור ולדווח על תקפותם ומהימנותם של מבחנים המופעלים עבור נכים. על מחקרים אלה לבחון את השפעות השינויים שנערכו עבור אנשים הסובלים מתנאי נכות שונים, על התוצאות המתקבלות, כמו גם את השפעותיה של הפעלת מבחן בלתי מתוקן על קבוצה זו.

הערה: ביצוע מחקרי תקפות הוא הדרך היחידה לצבור מידע לגבי יעילותם של המבחנים לגבי נכים. יש לשקול את עלויות התיקוף למבחנים אלה, כנגד העלויות הכרוכות בהיעדר מידע מועיל ובאשר למשמעות התוצאות לגבי נכים.

סטנדרט 13.6 המשתמשים במבחנים, ואלה הבאים במגע מקצועי עם נבחנים פוטנציאליים הסובלים מנכות צריכים:

- (א) לדאוג לקבלת המידע הנדרש לבחירה נכונה של אמצעי הערכה חלופיים;
 - (ב) לדאוג לקבלת מידע שוטף בנוגע לאפשרות להשיג נוסחים מתוקנים של המבחנים הנדונים;
 - (ג) להודיע לנכים על קיומם של נוסחים מתוקנים, בעת הצורך;
 - (ד) לאפשר לנבחנים להיבחן בנוסחים כאלה, כאשר הדבר בר ביצוע ומתאים.
-

סטנדרט 13.7 בעת הערכת מאפייניו של פרט הסובל מנכות, יש להחליט על שימוש בנורמות רגילות או מיוחדות וזאת על פי מטרת המבחן; שימוש בנורמות רגילות לגבי המאפיינים הנידונים, הינו מתאים כאשר מטרת המבחן כוללת את בדיקת תפקודו של הנבחן ביחס לאוכלוסייה הכללית. מכל מקום, במידה וקיימות נורמות מיוחדות, יש לבחור בהן, בעת שנידונים תפקודיו של הנבחן ביחס לאנשים אחרים הסובלים מנכות.

חלק רביעי

סטנדרטים להליכים מנהלתיים

Standards for Administrative Procedures

14. העברת מבחן, ציינון ודיווח
(Test administration, scoring, and reporting)

רקע

פירושים הניתנים לציוני מבחנים, הינם מהימנים ביותר כאשר המדידות מתקבלות בתנאים מתוקננים או מבוקרים. ללא התקנון, תרד איכותו של הפירוש, עד למידה בה הבדלי הליכים ישפיעו על הביצועים. חישוב ציונים ודיווח מדויקים, הינם חיוניים בכל תנאי.

סטנדרט 14.1 בהעברות טיפוסיות, על הבוחנים לבצע בשימת לב את ההליכים המתוקננים להעברה, המצוינים במדריך. יש למלא בדייקנות אחר מפרטים הנוגעים להוראות שיש לתת לנבחנים, למגבלות הזמן, לצורת הצגת פריטים או תגובות, לחומרי המבחן או הציוד הנדרש. יש לתרוג מהנהלים המפורטים אך ורק על בסיס שיפוט מקצועי שקול וזהיר, בעיקר ביישומים קליניים. יש לתאר בדו"ח עריכת המבחן כל שינוי שנערך בהליכים הסטנדרטיים להפעלת המבחן, בלוויית אזהרות מתאימות בכל הנוגע להשפעות האפשריות של שינויים כגון אלה, על התקפות.

סטנדרט 14.2 המקום בו נערך המבחן צריך להיות נוח במידה סבירה ובעל מספר מצומצם ככל האפשר של גורמים להסחת הדעת ולהפרעות. על חומר המבחן להיות קריא ומובן. **הערה:** יש לערוך בקרה בזמן המבחן, בעת הצורך, הן על מנת לסייע לנבחן בעת הנדרש, הן על מנת לשמור על נהלי העברה נאותים. בכל האמור במבחנים המועברים בצורה ממוחשבת, יתכן ולחידוש בצורת הצגת המבחן, תהיה השפעה בלתי ידועה על הפעלת המבחן.

סטנדרט 14.3 יש לעשות מאמצים סבירים על מנת להבטיח את תקפות ציוני המבחן באמצעות פעולות שתימנענה מהנבחנים לצבור נקודות באמצעים בלתי הוגנים. **הערה:** בתוכניות מבחנים גדולות, בהן עשויים הציונים להיחשב כבעלי השפעה כבדת משקל, במידה והדבר אפשרי ובר-ביצוע, יש לכלול במאמצים אלה: העברה בו-זמנית של המבחן לכל הנבחנים הנבחנים באותו נוסח של המבחן; עמידה על הדרישה להצגת מסמכים מזהים; עריכת רישום לסדר הישיבה; הושבת הנבחנים במקומותיהם; דרישה למרחק נאות בין המושבים; בקרה מתמשכת על תהליך המבחן. על הבוחנים לרשום ולדווח על כל מקרה חשוב של חריגה מנהלי מבחן תקינים.

סטנדרט 14.4 הגורם המחשב תוצאות מבחן, מתעד את ההליכים בהם השתמשו, על מנת להבטיח דיוק ואפשרות שיחזור בחישוב התוצאות. יש למדוד את תדירות השגיאות ולדווח עליה על פי בקשה.

סטנדרט 14.5 מקום בו הדיווח על הציונים עשוי להיות בסיס לקבלת הכרעות בעתיד הקרוב, ונתגלתה טעות מהותית בציינון המבחן או במידע חשוב אחר, יש להפיץ דו"ח מתוקן במהירות האפשרית. יהא הנזק למוניטין של הגורם המציין אשר יהיה, אין להעלים ולהסתיר טעויות ציינון שנתגלו.

סטנדרט 14.6 על המשתמשים במבחן לשמור על בטיחות וסודיות חומר המבחן.

סטנדרט 14.7 ביישומים כגון קבלת תלמידים למוסד חינוכי, מתן תעודות או אישורים, בהם הכרעות חשובות תלויות בביצועים במבחן נתון, יש לאפשר לנבחנים אמצעים לבדיקת הדיוק בחישוב הציונים. מקום בו לא ניתן לשחרר לידי הנבחנים את המבחן עצמו ואת המפתח לחישוב הציונים, יש לספק אמצעים אחרים לאימות ובדיקה.

סטנדרט 14.8 כאשר מוחזקים נתוני מבחן של הפרט, יש לשמור הן על פרטוקול המבחן, הן על כל דו"ח כתוב.

סטנדרט 14.9 על האחראים על תוכניות עריכת מבחנים, לספק פירושים נאותים כאשר משוחרר מידע לגבי תוצאות המבחן לתלמידים, הורים, מיופי כח חוקיים, מורים או אמצעי התקשורת. הפירושים אמורים לתאר בשפה פשוטה את התחומים אותם מכסה המבחן, את משמעות התוצאות, לציין פירושים בלתי נכונים אך נפוצים של ציוני מבחן, ולתאר את אופן השימוש בציונים.

סטנדרט 14.10 על ארגונים המחזיקים בציוני מבחנים בתיק נתונים אישי או באמצעי תיעוד אישיים, לפתח מערכת ברורה של קוים מנחים באשר לפרק הזמן בו יחזיקו בנתונים אלה, באשר לנגישותם, והשימוש שיעשה בתוצאות אלה על פני זמן. הערה: במקרים מסוימים, יהפכו תוצאות המבחן למיושנות ובטלות במשך הזמן, ואין להשתמש בהן או לאפשר בהן שימוש חיצוני. במקרים אחרים, תוצאות מבחנים שהתקבלו בעבר עשויות להיות מועילות ביותר, לדוגמה, בהערכה ארוכת טווח.

15. הגנה על זכויות הנבחנים

(Protecting the rights of test takers)

רקע

כיום מקובלים מספר עקרונות כלליים באשר להיבטים אתיים של מבחנים. בפרק זה מצוינות מספר דרישות טכניות, שעוצבו על מנת להבטיח התאמה לעקרונות אלה. נדון גם נושא בטיחות המבחן וביטול תוצאות נבחנים בשל חריגות מנהלי מבחן תקינים.

סטנדרט 15.1 יש לקבל את הסכמתם בידיעה של הנבחנים, (או של באי-כוחם) טרם ביצוע המבחן, למעט המקרים הבאים:

(א) מקום בו עריכת מבחן ללא הסכמה, מותרת על פי חוק או נוהל ממשלתי (לדוגמה, מבחנים בצבא);

(ב) כאשר המבחן מתנהל כחלק שגרתי מפעילויות בית הספר (לדוגמה - תוכניות מבחן כלליות בבית הספר, ושיתוף פעולה בין בתי ספר לצורך יצירת נורמות ועריכת מתקרים);

(ג) מקום בו ההסכמה משמעת מאליה בעליל (לדוגמה - הצגת מועמדות למשרה או למוסד חינוכי).

מקום בו לא נדרשת הסכמה, יש ליידע את הנבחנים באשר להליך המבחן. **הערה:** הסכמה בידיעה משמעה, כי לנבחנים או לנציגיהם הובהרו, בשפה המובנת להם, הסיבות לעריכת המבחן, סוג המבחנים המופעלים, השימוש המיועד להם וההשלכות המעשיות של השימוש המיועד, ואיזה מידע לגבי המבחן יועבר לאילו גופים. לנבחנים צעירים יש לספק הסברים על הסיבות לעריכת המבחן. גם ילד בן שנתיים או שלוש, ורבים מהנבחנים הסובלים מפיגור שכלי, מסוגלים להבין הסבר פשוט המבאר להם לשם מה הם נבחנים.

סטנדרט 15.2 אין להציג ציוני מבחן, כאשר הנבחנים מזוהים בשמותיהם, לשום אדם או מוסד, ללא קבלת הסכמה שבידיעה מהנבחנים או באי-כוחם המשפטיים, אלא אם כן החוק מציג דרישות אחרות. יש להעמיד תוצאות מבחניהם של נבחנים המזוהים בשמותיהם אך ורק לרשות גופים אשר הינם בעלי ענין חוקי ומקצועי.

סטנדרט 15.3 במסגרות חינוכיות, קליניות, או מסגרות ייעוץ, מקום בו נעשה שימוש בתוצאות מבחנים על מנת לקבל הכרעות לגבי פרטים, יש להעמיד לרשות הפרט הנדון או בא-כוחו המשפטי, מידע לגבי תוצאות המבחן ופירושו, לצורך שימוש נאות.

סטנדרט 15.4 על נתוני מבחן מתויקים יש לשמור מפני חשיפה בלתי נאותה. יש להגביל את השימוש במחשבים המחוברים למחשב מרכזי על בסיס חלוקת זמן, מאגרי נתונים ומערכות עיבוד נתונים אלקטרוניות אחרות, למצבים בהם ניתן באורח סביר להבטיח סודיות.

סטנדרט 15.5 מקום בו הדיווח על תוצאות כולל סיווג פרטים לקטגוריות, אזי על הקטגוריות הנבחרות להתבסס על מערכת קריטריונים שנבחרה בקפידה. יש להקפיד לחשתמש בתוויות מדויקות הכרוכות בסטיגמה המזערית ביותר.

סטנדרט 15.6 בתנאים מסוימים, יתכן ויהא זה רצוי לבטל את תוצאות מבחניו של הפרט או לאסור את פרסומו, בשל אפשרות לחריגות מנהלי מבחן תקינים, לרבות חשד להתנהגות בלתי הולמת. לנבחנים, אשר תוצאותיהם בוטלו או לא שוחררו לפרסום, יש להסביר כראוי מהם סוגי העדויות והנהלים בהם נעשה שימוש על מנת לקבוע שיש לבטל או לא לפרסם תוצאה.

סטנדרט 15.7 במסגרות כגון משרד קבלת תלמידים למוסד חינוכי, מתן תעודות/רשיונות או מתן תעודות השכלה, טרם ביטול תוצאה או עיכוב פרסומה מעבר לתקופת תקירה קצרה ומקובלת, יש להזהיר מראש את הנבחנים, ולספק להם הזדמנויות להציג עדויות התומכות באי-ביטול או אי-עיכוב תוצאותיהם.

הערה: מספר ארגונים העוסקים בעריכת מבחנים, מטפלים הזדמנות לביצוע חוזר של המבחן תוך זמן קצר, או קיום בוררות.

מילון עברי-אנגלי

במילון זה מצויינים המונחים באנגלית שהם המקור למילים העבריות המופיעות במסמך.

עבור מרבית המונחים המוגדרים במילון זה, קיימות מספר הגדרות בספרות. ההגדרות הבאות מציינות כיצד משמש כל מונח בטקסט זה.

לגבי אותם מונחים אשר שכיחותם בספרות המקצועית הפסיכולוגית והחינוכית הכללית, אינה גבוהה, נמסר כאן גם פירוט המשמעות (ולא רק תרגום). גם בקטע זה של המסמך נעזרנו ב- Glossary של ה"סטנדרטים" האמריקאי.

Percentile	* אחוזון
Coaching	* אימון
Simultaneous estimation	* אמידה בו זמנית
<p>תהליך (קלאסי, ביזיאני, או ביזיאני-אמפירי) לאמידה בו זמנית של שני ערכים או יותר; המטרה היא לשפר את האומדן לכל פרמטר באמצעות שימוש במידע שנאסף במשותף עבור משפחה שלמה של פרמטרים.</p>	
Feasible	* בר ביצוע
Factor	* גורם
Domain sampling	* דגימת תחום
<p>תהליך בחירת פריטי מבחן, אשר ייצגו את מכלול הביצועים הספציפיים אשר בו מעוניין מפתח המבחן.</p>	
Interpretive report	* דו"ח פרשני
<p>סיכום של ציון מבחן ומידע הקשור בו, המספק פירוט משמעותי של הציון.</p>	
Percentile rank	* דירוג אחוזי
Item bias	* הטייה של פריט
<p>פריט נחשב כמוטה שלילית או חיובית לגבי קבוצה מסוימת באוכלוסיה אם:</p> <p>(א) התוצאה הממוצעת הצפויה לגבי פריט זה גבוהה או נמוכה באורח ניכר מזו של האוכלוסיה הכללית, ו-</p> <p>(ב) אי ההתאמה נובעת מגורמים שהפריט אינו אמור למדוד.</p>	

Differential prediction	חיזוי דיפרנציאלי *
<p>המידה בה מבחן משמש לניבוי מידת הישגיהם היחסיים של אנשים, מוביל לחיזוי שונה - על פי אותה מערכת של קריטריונים - עבור קבוצות בעלות מאפיינים דמוגרפיים שונים, התנהגויות קודמות שונות, או לקבוצות אשר עברו טיפול שונה.</p>	
Expectancy table	טבלת ציפיות *
<p>טבלה של ערכים נבדלים המשמשת לחיזוי רמות של ביצועי קריטריון עבור ערכים מוגדרים, או במרווחים מוגדרים של תוצאות חזאי.</p>	
Projective technique	טכניקת השלכה *
Error of measurement	טעות מדידה *
<p>ההפרש בין התוצאה הנצפית והתוצאה האמיתית המתאימה לה. ראה גם: <u>טעות תקן במדידה, ותוצאה אמיתית.</u></p>	
Standard error of measurement	טעות תקן במדידה *
<p>סטטיית התקן של טעויות מדידה הקשורה לתוצאות מבחן.</p>	
Standard error of estimate	טעות תקן של אומד *
<p>סטטיסטי המתאר את ממוצע עצמת הטעויות בחיזוי, הכרוכות בשימוש במשוואת רגרסיה לחיזוי y בהינתן x.</p>	
Discrimination	יכולת אבחנה *
<p>מידת יכולתו של מבחן או פריט ממבחן, להבדיל בין פרטים, על פי מדידת הדרגה בה מבטאים הפרטים את התכונה הנמדדת על ידי אותו המבחן או פריט המבחן.</p>	
Decision rule	כלל הכרעה *
<p>כלל לבחירת נתיב פעולה על בסיס עדויות. לדוגמה, כלל שעל פיו מורים מעבירים או מכשילים תלמיד בקורס לימודים, בהתבסס על ציוני המבחנים של התלמידים וביצועים אחרים בקורס; כלל על פיו סוכנות ממשלתית מדרגת הצעות לפרויקטים, לשם הכרעה על הקצאת כספים, בהתבסס על תכולת ההצעות והדירוג המיוחס להן על פי הגופים השופטים; או כלל על פיו מעריך מחליט אם להבדלים בין ציוני המבחנים של סטודנטים הנחשפים לתוכניות לימודים שונות, יש משמעות מעשית.</p>	
Configural scoring rule	כלל תבניתי לקבלת תוצאות *
<p>כלל המשמש למתן משקלים לקבוצת תנאים כך שהפירוש של ציון חזאי תלוי ברמה של ציוו החזאי השני.</p>	

Adaptive testing	מבחן אדפטיבי *
מבחן בעל צורה סדרתית, בו נבחרים פריטים עוקבים במבחן, על בסיס התגובה לפריטים קודמים.	
Achievement test	מבחן הישג *
Ability test	מבחן כושר *
Speed test	מבחן מהירות *
מבחן בו נמדדים ביצועים על פי מספר המטלות המבוצעות בתחומי פרק זמן קצוב נתון. דוגמאות לכך הינן מבחני מהירות הדפסה וקריאה. כמו כן, מבחן המודד דיוק, בו עובד הנבחן תחת לחץ של זמן.	
Low-volume test	מבחן מצומצם *
בחקשר הנוכחי - מבחן אשר בו מספר הנבחנים קטן מכדי שיאפשר את ביצוע ניתוחי הנתונים המצויינים בסטנדרטים המותנים.	
Anchor test	מבחן עוגן *
מערכת משותפת של פריטים, המופיעים בשני נוסחי מבחן שונים או יותר, למטרת השוואת תוצאות הנוסחים הללו.	
Objective-referenced test	מבחן תלוי-מטרות *
ראה: <u>מבחן-תחום</u> .	
Norm-referenced test	מבחן תלוי-נורמה *
כלי שפירושו מתבסס על השוואת ביצועיו של נבחן לביצועיהם של אנשים אחרים, המשתייכים לקבוצה מוגדרת.	
Screening test	מבחן סינון ראשוני *
מבחן המשמש ליצירת מערכת קטגוריות נרחבת כשלב ראשוני בבחירת תהליכי קבלת הכרעות או החלטות אבחון.	
Criterion-referenced test	מבחן תלוי-קריטריון *
מבחן המאפשר למשתמשים בו לפרש תוצאות ביחס לרמת ביצוע תפקודי, להבדיל מפירושים אשר מתבצעים ביחס לביצועיהם של אחרים. ראה גם <u>מבחן תלוי-תחום</u> .	

Domain-referenced test	* מבחן תלוי-תחום
מבחן המאפשר למשתמשים לאמוד את היקף תחום תוכן ספציפי, שלמד פרט. תחומים עשויים להתבסס, לדוגמה, על מטרות לימודיות. ראה גם <u>מבחני קריטריון</u> , עדות לתקפות תוכן.	
Construct	* מבנה
מאפיין פסיכולוגי (לדוגמה - יכולת מספרית, יכולת מרחבית, מופנמות, חרדה), הנחשב כמאפיין אשר משתנה בין פרטים. מבנה (המכונה לעיתים משתנה סמוי) אינו ניתן לתצפית ישירה; תחת זאת, הוא מהווה תפיסה תיאורטית הנגזרת ממחקר והתנסויות אחרות, אשר נבנו על מנת להסביר דפוסי התנהגות נצפים. כאשר מפרשים תוצאות מבחן תוך שימוש במבנה, התוצאות מוכנסות למסגרת מושגית.	
Item discrimination index	* מדד לכושר הבחנה של פריטים
מדד למידה בה נבחנים - אשר נחשבים כמצויים ברמה גבוהה ביחס לקריטריון מסוים - מפגינים ביצועים טובים יותר באשר לפריט מסוים, מאשר נבחנים הנחשבים כמצויים ברמה נמוכה יותר ביחס לאותו הקריטריון.	
Item difficulty index	* מדד לקושי של פריטים
User's guide	* מדריך למשתמש
הצהרה הדנה במטרות המבחן, תוכנו, השימושים המתאימים לו, בנוסף למידע על הפעלת המבחן, חישוב התוצאות, נתונים נורמטיביים, ופירוש לתוצאות. לעיתים, המדריך למשתמש הינו מקטע מהחומר המצוי במדריך המבחן.	
Reliability	* מהימנות
Inter-rater reliability	* מהימנות בין מעריכים
Longitudinal study	* מתקר אורך
מתקר העוסק במדידת דגימות חוזרות במספר נקודות בזמן.	
Test specifications	* מפרטי מבחן
תיאור כללי של תוכן המבחן, המפרט איזה חלק יחסי מהפריטים יעסוק בכל תחום תוכן, ובכל סוג של יכולת. תיאור כללי זה עשוי אף לכלול פרטים כגון מספר הפריטים במבחן, הזמן המותר להפעלה, התבנית הטיפוסית של הפריטים, ומפרטים סטטיסטיים כגון התפלגות קושי הפריטים, ומדדי יכולת הבחנה.	
Reliability coefficient	* מקדם מהימנות

Split-half reliability coefficient	מקדם מהימנות מבחן חצוי *
Test-retest coefficient of reliability	מקדם מהימנות של מבחן חוזר *
Validity coefficient	מקדם תקפות *
Variance components	מרכיבי שונות *
<p>שונויות של חלקי מבנה נפרדים, אשר נחשבים כמשולבים בתוצאה הנצפית. שונויות אלה אומדים באמצעות שיטות ניתוח שונות. לעיתים קרובות, משקפות שונויות אלו מיקום, זמן, צורה, מעריכים והשפעות קשורות אחרות.</p>	
Variable	משתנה *
Scoring formula	נוסחה לקבלת תוצאות *
<p>הנוסחה אשר באמצעותה מתקבלת תוצאה גולמית של המבחן. הנוסחה הפשוטה ביותר לקבלת תוצאה היא "התוצאה הגולמית שווה למספר התשובות הנכונות". נוסחאות אחרות מייחסות משקלים שונים לתגובות לפריטים, לעיתים במטרה לתקן תוצאה כוללת הנגרמת מניחוש או אי מתן תשובה, ע"י מתן משקל אפס לאי מתן תגובה, ומשקל שלילי לתשובות שגויות.</p>	
Alternate forms	נוסחים חלופיים (מקבילים) *
<p>שני נוסחים או יותר, אשר דומים זה לזה בטבעם, ומיועדים לאותה המטרה. קיימות שלוש קטגוריות לנוסחים חלופיים. נוסחים בני השוואה (comparable) הינם דומים בטבעם, אולם לא הוכח כי יש להם מאפיינים סטטיסטיים דומים. על מנת שיכוננו מקבילים, (parallel), על הממוצעים, סטיות התקן והמתאמים עם כלים אחרים, להיחשב כשווים בקירוב. על מנת שיכוננו שווי ערך (equivalent), עליהם להיות זהים בכל האיפיונים הרלוונטיים.</p>	
Equated forms	נוסחים מכילים *
<p>שני נוסחי מבחן ומעלה, אשר תוצאותיהם שוות ערך או מקבילות, לגבי קבוצות מסוימות של נבחנים.</p>	
Parallel forms	נוסחים מקבילים *
<p>ראה: <u>נוסחים חלופיים</u></p>	
Equivalent forms	נוסחים שווי ערך *
<p>ראה: <u>נוסחים חלופיים</u>.</p>	

Norms	* נורמות
Normative	* נורמטיבי
Factor analysis	* ניתוח גורמים
Job analysis	* ניתוח עיסוקים
	כל אחת ממספר השיטות הקיימות לזיהוי המטלות המבוצעות במסגרת עיסוק, או הידע, המיומנויות והיכולת, הנדרשים לביצוע דרישות העיסוק.
Internal analysis (of reliability)	* ניתוח פנימי (של מהימנות)
	משפחה של שיטות לאמידת מהימנות, המבוססות על העברת נוסח יחיד של מבחן. למשפחה זו משתייכים: מקדם α , נוסחאות קודר-ריצ'רדסון, ושיטות הכוללות חציית המבחן.
Item analysis	* ניתוח פריטים
Battery	* קוללה
	מערכת של מבחנים שתוקננו על פי אותה האוכלוסיה, כך שאת הציונים שלהם ניתן להשוות, או להשתמש בהם בשילוב, לצורך קבלת החלטות.
Classification	* סיווג (גם - "שיבוץ")
	פעולה הקובעת איזו משימה או טיפול מתוך מספר אפשרויות, יש לתת לאדם, או ליעץ לו לבחור.
Content-related evidence of validity	* עדויות לתקפות-תוכן
	עדות המורה על המידה בה תחום תוכן של מבחן הינו מתאים בהקשר למטרתו המיועדת. עדות כזו משמשת לקבוע אם המבחן כולל דגימה מייצגת או קריטית של תחום התוכן הרלוונטי, ואם הוא אכן אינו כולל תוכן שאינו שייך לתחום התוכן.
Construct-related evidence of validity	* עדות לתקפות-מבנה
	עדות התומכת בפירוש מבני המוצע לציונים, במבחן אשר מבוסס על השלכות תיאורטיות הקשורות לתווית המבנה.
Criterion-related evidence of validity	* עדות לתקפות קריטריון
	עדות המורה על היחס בין תוצאות של מבחן לבין מדד הקריטריון.
Concurrent criterion-related evidence of validity	* עדות מקבילה לתקפות קריטריון

עדות לתקפות קריטריון, בה המנבא והקריטריונים מתקבלים בעת ובעונה אחת בקירוב.

* **Predictive criterion-related evidence of validity** עדות נבויית לתקפות-קריטריון

עדות לתקפות-קריטריון, בה תוצאות קריטריון נבדקות בתאיך מאוחר יותר; לדוגמא, בהקשר לביצועים בבית הספר או בעבודה.

* **Decision consistency** עקביות הכרעות

אחוז הפעמים בו מתקבלת אותה ההכרעה כאשר נעשה שימוש בכלל הכרעה מוגדר מעבר למבחנים חלופיים, בהנחה שלא אירעו שינויים בכושרם האמיתי של הנבחנים.

* **Item response curve** עקומת תגובה לפריט

פונקציה הנותנת את היחס בין הסתברות להצלחה בפריט מסוים, לבין רמת התכונה הנמדדת ע"י הפריט (מכונה גם עקומת אפיון).

* **Psychometric** פסיכומטרי

המונח קשור למדידות של מאפיינים פסיכולוגיים, כגון יכולת, כשרים, הישגיות, תכונות אישיות, מיומנות וידע.

* **Profile** פרופיל

* **Computer-based interpretation** פירוש המבוסס על מחשב

שיטה של מתן פירושים לציוני מבחן, אשר מסתמכת על אלגוריתמים של מחשב, ודו"חות מודפסים של ציונים שבוצעו ע"י מחשב.

* **Score** ציון

* **T score** ציון T

ציון תקן בסולם בו ממוצע של 50 יחידות, וסטיית תקן של 10 יחידות.

* **Z score** ציון Z

סוג של סולם ציונים סטנדרטי, אשר בו הממוצע שווה לאפס, וסטיית התקן שווה ליחידה אחת, לגבי הקבוצה המשמשת להגדרת הסולם.

* **True score** ציון אמיתי

בתיאורית מבחנים קלאסית - ממוצע הציונים של פרט, אשר יתקבל במספר אין-סופי של נוסחים של אותו המבחן, המקבילים זה לזה בצורה מושלמת.

Cut score	ציון חתך *
Composite score	ציון מורכב *
ציון המשלב מספר ציונים על פי נוסחה מוגדרת.	
Derived score	ציון נגזר *
ציון המתקבל על ידי חמרת תוצאות גולמיות באמצעות טרנספורמציה מספרית (לדוגמה - דירוג באחוזים או תוצאות מתוקננות).	
Scaled score	ציון סולמי *
ראה: <u>ציון נגזר</u>	
Standard score	ציון תקן *
Comparable scores	ציונים בני השוואה *
ציונים של מבחנים שונים, הנמדדים על פי אותו סולם ציונים, ואשר להם אותה משמעות יחסית, במסגרת קבוצת התיחסות משותפת.	
Restriction of range	קיצוץ תחום *
מצב בו - בשל הגבלות על נטילת דגימות - מידת השוני של נתונים בדגימה הינה נמוכה ממידת השוני הקיימת באוכלוסייה הנידונה.	
Criterion	קריטריון *
אמצעי לציון הערך המקובל של תוצאת פעולה, כגון ממוצע ציונים, שיעור תפוקה, שיעור תאונות, רמת ביצוע, רמת היעדרות, דרגת דחיה וכד'. זוהי בדרך כלל אמת מידה שכנגדה מתבצעת הערכה של אמצעי חיזוי.	
Confidence interval	רווח בר-סמך *
טענה הסתברותית, לפיה רווח המוגדר על ידי ערכים מסולם הציונים, מכסה את הפרמטר המבוקש בהסתברות מסוימת. המונח משמש בסטנדרטים אלה אף בהקשר לרווחי סמך ביזיאניים, המגדירים רווח אשר בו יפול הפרמטר הבלתי ידוע בהסתברות מסוימת.	
Critical score level	רמת ציונים קריטית *
נקודה מכרעת בסולם ציונים למבחן, אשר מעליה ומתחתיה קיים שוני בין הכרעות והערכות. ראה גם <u>ציון חתך</u> .	

Interest inventory	רשימת (שאלון) נטיות	*
סדרת שאלות או הצהרות המשמשת להסיק על תחומי הענין, ההעדפות, ודברים האהובים או השנואים על המשיב.		
Inventory	שאלון	*
Personality inventory	שאלון אישיות	*
Variance	שונות	*
Equating method	שיטת כיוול	*
תהליך המשמש להמרת סולם ציונים של נוסח אחד של מבחן, לסולם ציונים של נוסח אחר, כך שהתוצאות תהיינה שוות או מקבילות.		
False negative	שלילה מוטעית	*
בתחום המיון - טעות בהערכה, על פיה, מצופה שאדם יכשל, אולם למעשה היה מצליח אילו נבחר. לחלופין - סווג אדם באורח מוטעה, כשייך לקבוצה נמוכה יותר ברמתה.		
Generalizability theory	תיאורית אפשרות הכללה	*
שלוחה של תיאורית המהימנות הקלאסית, בה ניתוח שונות משמש לאמידת מרכיבי השונות, ומנקודה זו, משמש לתאר את עוצמת הטעויות הנובעות ממקורות מוגדרים. הניתוח משמש לאמוד את יכולת ההכללה של תוצאות, מעבר לדגימה ספציפית של פריטים, אנשים, ותנאים נצפים אשר נבדקו.		
Free response	תגובה חופשית	*
תגובה לפריט של מבחן, אשר הנבחן חייב לספק, בניגוד לתגובה אשר על הנבחן לבחור מרשימת אפשרויות תלופיות.		
Utility	תועלת	*
ערכה היחסי של התוצאה ביחס למערכת של תוצאות אפשריות אחרות.		
Raw score	תוצאה גולמית	*
Weighted scoring	תוצאה משוקללת	*
Gain scores	תוצאות שיפור	*
ההפרש בין הציון במבחן והציון בהפעלה קודמת של אותו המבחן או מבחן שווה ערך לו.		

Residual score תוצאת שארית *
ההפרש בין הציון הניצפה לבין הציון האמיתי.

Content domain תחום תוכן *
גוף של מידע, מיומנויות או כשירויות, המוגדר כך שפריטי מידע או מטלות מסוימות יהיו ניתנים לזיהוי ברור כשייכים או בלתי שייכים לתחום (גם "עולם תוכן").

Validation תיקוף *

Cross-validation תיקוף צולב *
יישום של מערכת מתן ציונים שמקורה אמפירי, או מערכת משקלות, של דגימה אחת על דגימה שניה, במטרה לבדוק את יציבות החיזוי לגבי מערכת מתן הציונים או המשקלות.

Validity תקפות *