



**קביעת סדר קדימויות  
בפיתוח תכנית  
בחינות לאומית**

וויליאם אנגורף



### סדר קדימויות בפיתוח תכנית בחינות לאומית (1), (2), (3)

בניה, פיתוח והפעלה של תכנית בחינות דורשת ראשית כל הכרה וזיהוי של מגוון בעיות וצרכים. לאחר שאלו זוהו - משימה לא קטנה כשלעצמה - יש להתמודד עימם באופן שלא רק ייצג את הפתרון הזמין הטוב ביותר מבחינה מקצועית ומעשית, אלא גם יציג זאת באופן שיטע אמון בציבור שאמנם הדבר הנכון מתבצע; שחוויה היא טובה והדרך הגיונית וסבירה. להלן נצביע על מספר נושאים חשובים בעריכת תכנית מבחנים בסדר גודל לאומי. לצורך הדגמה, נניח שהתכנית הנדונה מיועדת למטרות מיון אקדמי.

#### קביעת הרכב הסוללה

המשימה הראשונה הינה איתור המבחנים שישמשו כבסיס התכנית. מגוון מבחני יכולת (aptitude) מילוליים ומתמטיים עשויים להיות הולמים, וכך גם אולי מבחן ידע כללי המיועד לבחון את היקף הקריאה של המועמד ואת המידה בה הוא קשוב לעולם האינטלקטואלי השובב אותו. בכל מקרה, לאחר שאיתרנו סוללת מבחנים נסיונית, יש לערוך מחקרי תיקוף בכל אחד מן המוסדות, או במדגם מייצג של המוסדות. זאת על מנת לזהות מהי הסוללה התיפעולית האופטימלית. בחירת המבחנים שירכיבו סוללה זו תלויה במספר גורמים:

1. התקפויות של כל אחד מן המבחנים; 2. המידה בה הם מצליחים להימנע מחפיפה ביניהם; 3. המידה בה הם מרכיבים במשותף את תוקף הניבוי המקסימלי תחת המגבלות המעשיות הקיימות.

לצורך מחקרי תיקוף אלו, ידרשו תחילה ניתוחי רגרסיה. כמו כן, יהיה צורך לערוך ניתוחי גורמים (factor analyses) של המבחנים ושל הקריטריונים להצלחה (קריטריונים של המוסדות) - ניתוחים שמטרתם להגיע להבנה טובה יותר של התכונות הנמדדות.

- (1) מאמר זה הוזמן אצל ד"ר ו. אנגוף, חוקר בכיר ב-E.T.S. ויועץ למרכז הארצי. המאמר הוצג בכנס IAAP בירושלים, בקיץ 1986.
- (2) ב. נבו ערך מספר שינויים במאמר המקורי, בעיקר כדי להתאים למונחים ומצבים המקובלים בישראל.
- (3) תורגם על ידי נעמי שטרך.

בתהליך זה, יהיה זה שימושי לבחון את המהימנות ותקפות התוכן של הקריטריון עצמו. זאת משום שבמידה שהקריטריון - בד"כ ממוצע הציונים בשנה מסוימת - אינו הולם, הרי שיש צורך למצוא דרכים לפיתוח קריטריון הולם יותר, וכן, להציע יעוץ ועזרה למוסדות עצמם באשר לשיטת הערכה משופרת.

אף כי יש בידינו מספר הולך וגדל של עדויות לכך שתקפיות מבחן משתנות לעתים קרובות באופן שיטתי כפונקציה של ההטרוגניות של הקבוצה הנבחנת, הרי שיהיה זה נכון, לפחות בתחילה, לתקף את המבחנים בנפרד לכל מוסד, ואולי חשוב מכך - בנפרד לכל פקולטה וחוג. אם נתוני התיקוף מן המוסדות השונים יעמדו בהתאמה זה עם זה, אזי ניתן יהיה לצרפם. אולם, בהחלט יתכן שיהיו הבדלים משמעותיים במשקלות הרגרסיה בין המוסדות השונים, ואין להתעלם מאפשרות זו.

לבסוף, יש להדגיש שהתיקוף אינו פעולה חד-פעמית, כי אם מחקר מתמשך. לרציפות זו מספר סיבות: ראשית, על מנת לבדוק אם סדרת המבחנים ממשיכה להיות שימושית; שנית, כדי לבדוק את יציבותו ושימושיותו של הקריטריון ושלישית, במטרה לבחון את השימושיות האפשרית של סוגי פריטים ומדדים חדשים.

#### חיבור פריטים וניתוח פריטים

לאחר שעוצבו הכלים הבסיסיים של תכנית הבחינה, על הצוות והיועצים לחבר פריטים שיהוו את המרכיבים של המבחנים הטיפעוליים עצמם. ברור שהפריטים חייבים להיות משמעותיים והגיוניים לגבי הנבחנים. אין לעסוק בנושאים שוליים, או להתבסס על תוכן נדיר, מעורפל, או אזוטרי, כדרך להשגת הבחנות תקפות בקרב התלמידים בעלי היכולת הגבוהה יותר. במבחן הבחירה הקונבנציונלי, צריכה להיות אופציה אחת, ואחת בלבד, שהיא נכונה באופן חד משמעי. כל הפריטים יבדקו ע"י מעריכים בלתי תלויים, לא רק כדי לוודא את תקינותם הטכנית, אלא גם כדי להבטיח שלא נכנס למבחן כל אלמנט של חוסר כבוד או עלבון לתת-אוכלוסיה כלשהי של נבחנים, וכן, שאין פריטים הדורשים ידע מיוחד או הבנות הקשורות באופן רופף בלבד למטרת המבחן והידועים לנבחנים בודדים בלבד.

כחשלמה להערכות אלו, צריכה להתקיים תכנית פעילה (רצוי ממוחשבת) של ניתוח פריטים. תכנית זו תקבע את הקושי של כל פריט, מספר העונים לכל מסיח ורמת היכולת שלהם, מספר הנבחנים שהגיעו לפריט, וכח ההבחנה של הפריט. בנוסף, צריך ניתוח הפריטים גם לזהות/לאתר פריטים אשר דומה שהינם מוטים כנגד תת קבוצה זו או אחרת של האוכלוסייה. קיימות מספר שיטות להקירת ההטיה של הפריט וביניהן ניתן לאתר כאלה שהינן יעילות וזולות יחסית. את ניתוח הפריטים יש לערוך בשלב הנסיוני (pretest), לפני שהפריטים הופכים לטיפעוליים. במידה ואין הדבר אפשרי, יש לבצע זאת לאחר העברת המבחן, אך לפני שמציינים את התשובונים.

בתכנית בחינות מתמשכת, חיוני שיהיו מספר נוסחים זמינים להעברה. יש לחבר נוסחים חדשים באופן סדיר. על מנת לבנות נוסחים אלו, יש לפתח מערכת של הוראות והנחיות, הקובעות את טבעם ותכנם של הפריטים, מספר הפריטים מכל סוג בתוך כל קבוצת תוכן, רמת הקושי ופיזורו והרמה המינימלית של מדדי אבחנות של פריט (כגון המתאם פריט-מבחן; item-test biserial correlation).

יחד עם מאמצים אלו, יש לפתח תכנית pretest, שתנסה פריטים חדשים. דבר זה רצוי לבצע בהקשר של בחינה תפעולית. ניתן לפתח שיטות לבדיקה נסיונית סימולטנית של מספר מערכות של פריטים, וזאת באמצעות שינוי שיטתי של החומר הנסיוני מתלמיד לתלמיד בתוך כל כיתת מבחן. במידה והדבר מבוצע כיארת, יבטיח תהליך זה שעל כל מדגם של פריטים נסיוניים, ענה מדגם מקרי של הקבוצה השלמה (רצוי עם דגימה שכבתית).

### העברת הבחינה

כאשר מתכננים להעברה עצמה, יש לדאוג לכך שהוראות מפורטות תהיינה זמינות לבוחנים, אשר יהיו אחראים לשמירה על תנאים סובבים נוחים לבחינה, לביצוע העברה מסודרת ושיטתית, לשמירת סודיות המבחנים ולהיצמדות לכללים להעברת המבחן. הבוחן יהיה אחראי גם לקבלה ולהחזרה הבטוחה של חומרי (חוברות) המבחן. על מנת לשמור על האופי הסטנדרטי של הבחינה, ינתנו ההוראות לבוחנים בשפה המדויקת בה יאמרו לנבחנים במהלך הבחינה.

מתפקידו של הבוחן גם למנוע, ובמקרה הצורך לאתר, רמאות. לעתים קרובות ניתן לאתר מתחזים באמצעות בדיקת תמונה וכתב יד בעת הכניסה לחדר המבחן. על מנת למנוע נסיונות העתקה, יש לדאוג לכך שהנבחנים ישבו במרחק מה זה מזה. חוברות הבחינה ימוספרו, ומפת ישיבה בחדר תוכן ע"י הבוחן מראש, או לאתר שהנבחנים מצאו את מקומם. באופן אידיאלי, יהיה כל עמוד במבחן בצבע שונה, או, לכל הפחות, יכלול סימן מזהה גדול, שייראה בנקל ע"י הבוחן. זאת על מנת שניתן יהיה לוודא שכל הנבחנים אמנם עובדים על אותו עמוד במבחן עליו התבקשו לענות. יחד עם זאת, בסופו של דבר תלויה מניעת ההעתקות וגילויין בהשגחה ובעירנות המתמדת של הבוחן.

לאחר העברת הבחינה, ניתן לחקור מקרים החשודים ברמאות, באמצעות השוואות כתבי יד (לאיתור מתחזים), וע"י השוואה של תשובות שגויות שסומנו באופן זהה (לאיתור העתקות). ניתן למצוא בספרות שיטות להשוואת תשובונים במבחנים סטנדרטיים.

תהליך הציינון עצמו הינו מרכזי ויש לוודא שיבוצע בסטנדרטים הגבוהים ביותר האפשריים. במידה ויתגלו טעויות בתהליך הציינון, יגרם בכך נזק חמור לשמו הטוב של המבחן ולאמון הניתן בו ובתכנית הבחינות כולה. לפיכך, יש לערוך בדיקות, התאמות וכיולים חוזרים, תכופים ונמשכים, על מנת שנוכל להיות בטוחים שלא נמנע מאיש מהתלמידים הציון לו הוא זכאי. מנהלי תכנית הבחינה צריכים להיות מוכנים לדווח בפומבי על שיעור הטעויות.

כאשר קיימים מספר נוסחים של הבחינה, חיוני שהנוסחים יכויילו כראוי או יהיו ברי השוואה זה לזה, והדיווח עליהם יתבצע על סולם רציף וקבוע. דבר זה הינו חשוב לא רק מסיבות של הוגנות (equity and fairness) - להבטיח שלא נותנים לתלמיד כלשהו יתרון על פני עמיתיו בכך שהוא נבחן במקרה בנוסח קל יותר - כי אם גם מסיבות חשובות אחרות: לדוגמה, כדי לערוך השוואות תקפות בין קבוצות תלמידים שנבחנו בנוסחים שונים, כדי לקבץ נתונים ולחקור לאורך זמן, כדי לשמש כנקודת ציון להקירת שינויים שיטתיים והבדלים שיטתיים בציוני קורסים, כדי לערוך מחקרי לפני-אחרי מבוקרים, וכן, כדי לספק מבחן חלופי לנבחנים שבחינתם הראשונה הפכה לבלתי תקפה.

בעיה הקשורה להשוואה (equating), היא הבעיה המועלית ע"י כל תכנית מבחנים (בדרך כלל מבחני הישג), המאפשרת לנבחנים לבחור באיזה מבחן (נושא) יבחנו. יש לדאוג לכך שתלמידים הבוחרים להיבחן בנושא קל יותר, או במבחן שבד"כ נבחנים בו תלמידים פחות טובים (אליהם יושוו בד"כ) לא יזכו ליתרון. קיימים פתרונות להתמודדות עם בעיות אלה, ואף שאינם אידיאליים, הם מסייעים במידה רבה לצמצום החטיות.

בעיה הקשורה לנושא היא זו של חיבורים אופציונליים. מערכת בחינות המציעה חיבורים אופציונליים, חושפת את עצמה להטיה בשתי דרכים: 1. בשל קשר הבחירה העצמית הקיים בין הנבחנים והחיבורים שתואר לעיל. 2. עקב העובדה שהחיבורים יקראו וידורגו ע"י מעריכים השונים זה מזה באשר לסטנדרטי הציינון האינדיבידואליים שלהם. יש לפתח שיטות שידאגו לפיזור החיבורים בין המעריכים השונים בדרך שתאפשר לבצע לאחר מכן התאמות ("הזזות ציונים") נוחות בין הקוראים. כאשר אין שיטות אלו זמינות, יש למצוא דרכים אחרות כדי לחקור הבדלים בין מעריכים ולבצע התאמות. מספר שיטות כאלה הינן זמינות. בכל מקרה יש להבטיח שכל חיבור יקרא באופן בלתי תלוי ע"י שני מעריכים שונים. באופן אידיאלי, יש להקצות את המעריכים לחיבורים באופן שיטתי ומאוזן.

לאחר העברת הבחינות, יש להוציא ניתוח סטטיסטי פורמאלי, המתאר את התאמת המבחנים לאוכלוסיה שנבחנה, את המהימנות וטעות המדידה של כל אחד מהמבחנים, תת-מבחנים, והסוללה ככלל, בצרוף עם מדדים של הגבלת זמן (speededness), מתאמים בין תת-החלקים של המבחנים, כולל עד רמת היחידה המיוצגת ע"י אוסף הפריטים מאותו הסוג. דו"חות אלו של ניתוח המבחן, אמורים להיות זמינים בהקדם האפשרי לצוות הסטטיסטי ולצוות המפתח את המבחן, על מנת שיהיה בידם המידע הנחוץ להם להכנת הנוסח החדש הבא.

#### דיווח הציונים ומידוע

הציונים ידווחו לנבחנים ולמוסדות בהקדם האפשרי, ובצורה מובנת. הם ידווחו על סולם שנבנה והוסכם עליו כשהתכנית החלה, סולם שישמש באופן קבוע לכל אורך תהליכי החשוואה (equating) של הציונים. רק אם דבר זה יבוצע, יוכלו הנבחנים, הוריהם, המוסדות והציבור, להכיר את יחידות הסולם, כפי שהם מכירים יחידות בסולמות אחרים בחיי היום-יום. במידת האפשר, יש לדווח על הציונים בתוך טווח טעות ידוע. בספרות מצויים קירובים טובים מאוד של טעויות המדידה ברמות שונות של ציונים. יש להכין חוברות מידע אודות המבחן, וחוברות הסבר לפירוש הציונים, לשם הפצה בקרב הנבחנים, או לפחות שיהיו זמינים לבקשתם. מדריכים/חוברות אלו, יכללו תיאור של תכנית הבחינה והמבחנים, הוראות הרשמה, אזהרות בנוגע להתנהגות נבחנים בכיתת המבחן ומידע באשר למשמעות הציונים. התפלגויות הציונים, המראות את האחוזונים (פרצינטילות) לכל ציון, צריכות להיות זמינות, וכך גם דיון בטעות הטבעה בתוך הציונים והתקפות של הציונים. מידע דומה, אך ביתר פירוט יהיה זמין למשתמשים במבחן, יחד עם הנחיות מפורטות באשר לאבחנות בין שימושים הולמים לבין שימושים בלתי הולמים בציונים. לבסוף, יש להכין גם מדריך טכני, אשר אף הוא יעודכן במידת הצורך, המתאר את פיתוח המבחן והשיטות הסטטיסטיות בהן נעשה שימוש בתכנית, מסכם נתונים אודות קבוצות נבחנים, וסוקר נתוני תקפות וממצאים של מחקרים משמעותיים יותר, שבוצעו בקשר לתכנית הבחינות.

#### בחינות בשפות זרות

לאור הפלורליזם של האוכלוסיה הישראלית, הונח כמובן מאליו, שחייבות להיות תכניות בחינה מקבילות בשלוש שפות לפחות: עברית, ערבית ואנגלית. הכוונה כאן היא בבירור טובה: שתהליך הבחינה יהיה הוגן ככל האפשר לכל התלמידים, שיכולות אקדמיות של קבוצות שונות של נבחנים יוערכו באופן הוגן ובר-השוואה, ושההשוואות ביניהם יבוצעו ברמת הסטנדרטים הללו, ללא קשר למוצאם, תרבותם ושפתם. מאמצים שנעשו בארצות הברית הצליחו להשיג מטרה זו במידה סבירה. אולם, מאמצים אלו בוצעו תוך קשיים רבים, ונערכו רק כאשר השפות והתרבויות השונות היו דומות. כאשר השפות והתרבויות שונות זו מזו

באופן משמעותי, בדומה לאנגלית מול ערבית, אזי הסבירות להצליח בהשגת השוואתיות הינה קטנה למדי. מבחנים - תכנם, המשמעות שהם מציגים, והשפה בה הם מוצגים - משקפים בהכרח את התרבות. לניואנסים של השפה, ולדקויות הכלולות שלא במתכוון במילים ומשפטים - יש משמעות לקבוצה אחת ולא לחברתה. תרגומים הינם בד"כ מקורבים ולכן בלתי הולמים למטרות הבחינה. למעשה, קשה להגיע להקבלה קרובה ממש; ניתן רק לקוות שבעיות של הוגנות יוכלו להיפתר באמצעים אחרים.

### מספר הערות כלליות

תכנית בחינות תהיה מוצלחת רק אם תעמוד במספר תנאים. תנאי אחד כזה הוא, שהתכנית תציג באופן משביע רצון מה היו הכוונות ומה הושג, לרבות כל המגרעות והטעויות. כל זאת יוצג לציבור אשר צפוי מלכתחילה להיות סקפטי, ואולי אף לא מאמין.

תנאי שני הוא שיערך מחקר מתמשך לגבי בעיות שעדין אינן מזוהות, לרבות הקירה עצמית, בכל הרמות של הרצף, החל מהתאורטי וכלה בישומי. נציג מספר דוגמאות של מחקרים מסוג זה:

יש לערוך בדיקות של יציבות הסולם, במרווחי זמן קבועים, זאת במגמה לוודא שהמשמעות של הציון לא עברה שינוי במהלך הזמן. באמצע שנות ה-70, כאשר התברר שציוני ה- College Board Scholastic Aptitude Test ירדו בקצב מעורר דאגה, עלה חשש מידי שירידה זו עלולה לבטא ליקוי כללי של מערכת הציונים. נערכו שלושה מחקרים בלתי תלויים בנוגע ליציבות הסולם. שלושתם הובילו לאותה המסקנה והיא, שמשמעות הציונים הינה בעיקרה זהה לזו שהייתה כ-10 שנים קודם לכן (ולכן יש לחפש את סיבת ירידת ציוני ה-SAT בסיבתיות אחרת). בצד מחקרי התקפות המתייחסים לקבוצה כולה, חשוב לבצע ניתוחי רגרסיה שיבדקו את האפשרות שקבוצה זו או אחרת סובלת מתת-ניבוי (underprediction) עקב המשוואה שפותחה עבור הקבוצה הכוללת.

יש לבצע גם סוגים אחרים של מחקרי תיקוף, על מנת לוודא, למשל, שכל הפריטים במבחני ההישג של התכנית, אמנם לקוחים מהמסגרת המוצהרת של הקורס, וכי התלמידים נחשפו לתוכן הרלוונטי.

לעתים קרובות יבחרו תלמידים לחזור על מבחן. חשוב לדעת באיזה קצב נוטים הציונים להשתנות, באיזו מידה נוטים הציונים בשני המועדים להיות מתואמים, ואילו גורמים - למשל, גיל, תכנית הלימודים, או משתני רקע אחרים - קשורים לשינויים בציונים.

חשוב גם לדעת אם המבחנים מועדים לאסטרטגיות מיוחדות להשגת ציונים גבוהים, אסטרטגיות שאינן קשורות ליכולת האמיתית של התלמידים. במידה ואסטרטגיות כאלה הינן אפשריות, יופיעו בתי ספר שונים המאמנים לקראת המבחן, המעוניינים ברווחים ועקב כך הם בהכרח מושכים במיוחד אותם סטודנטים היכולים להרשות



זאת לעצמם. התפתחות זו, בתורה, תוביל להאשמות וטענות אודות חוסר הוגנות. על מנת למנוע, לפחות חלקית, טענות מסוג זה, רצוי שתכנית הבחינות תפרסם חוברות היכרות עם המבחן, המכילות מספר ניכר של דוגמאות של פריטים, שימכרו בערכן הנומינלי. בנוסף, חשוב יהיה לבחון את יעילותן של אסטרטגיות האימון.

תהליך פיתוח המבחן, העברתו, סילומו, הפיכתו לבר השוואה או כיולו, וקביעת מהימנותו ותקפותו, דורש תקירה ושיפור תמידיים. מחקרים של שיפורים ועידונים של המתודולוגיה, מחקרים על שיפורים שיובילו להוזלת המבחן, מחקרים על הדגימה, על אוכלוסית התלמידים, סקרים שיקדמו את הקשר בין תכנית הבחינות לבין התלמידים, הוריהם והמוסדות - כל אלה הינם חיוניים. כללית, תכנית חזקה ופעילה של מחקר ולימוד עצמי תעזור להגיע לתכניות בחינה תקינות מבחינה טכנית. פעולות שיונחו ע"י ממצאי מחקרים כגון אלה תעזורנה לרכוש את אימון הציבור, שתמיכתו תקבע בסופו של דבר את הצלחתה או כשלונה של התכנית.

