



# מבחנים פסיכולוגיים בישראל ובארה"ב - תמונת מצב

ברוך נבו  
ועל בסיס מאמר של אנטוניו

## ה ק ד מ ה

קיימת הסכמה כללית, כי מבחנים פסיכולוגיים מהווים אחת התרומות היישומיות והמשמעותיות ביותר של מדעי ההתנהגות לחברה אנושית. די לנו למנות שימושים אחדים שנעשים במבחנים כדי להיווכח בנכונות הטענה.  
מבחנים פסיכולוגיים ממלאים תפקיד חשוב בפעולות האלה:

- ייעוץ והכוון מקצועי
- אבחון פתולוגיה נפשית למטרות קליניות
- איתור מחוננים
- זיהוי תת-משיגים
- אבחון פיגור שכלי
- ברירה ושיבוץ של עובדים
- ברירה ושיבוץ של צעירים למסלולי שירות צבאי
- ברירה של מועמדים למסלולי לימוד והשכלה
- הערכת קשיי תפקוד (וסיכויי שיקום) כתוצאה מליקוי אורגני
- מדידת מוראל, אקלים חברתי, אינטראקציות חברתיות
- יצירת קבוצות טרוגניות/הומוגניות לפי פרמטרים מוגדרים מראש
- מדידת אפקטיביות של הכנסת שינויים במערכות (מחקרי הערכה)
- מדידת משתנים תלויים, בלתי-תלויים, מפקחים ומתווכים לצורכי מחקר

מבחנים פסיכולוגיים הם חלק מ"ארגז הכלים" של כל הפסיכולוגים היישומיים - יהא תחום העיסוק הייחודי שלהם אשר יהא: תעסוקתי, חינוכי, קליני, שיקומי, רפואי, התפתחותי או אחר.

אוצר המבחנים העומד היום לרשות אנשי המקצוע כדי להשיג, לו גם חלקית, את המטרות הללו, הוא רחב ומגוון: מבחני משכל אינדידואליים, מבחני משכל קבוצתיים, שאלוני אישיות, מבחני השלכה, שאלוני נטיות, מבחנים למדידת כשרים מיוחדים, שאלונים ביוגרפיים, שאלונים סוציומטריים ועוד.

---

פרק זה התפרסם ב-1988 כמאמר בכתב העת: "פסיכולוגיה וייעוץ בחינוך".

מי שמשווה את ההוצאה השמינית (והאחרונה עד כה) של ה-  
Mental Measurement Yearbook (MMY), קטלוג המבחנים המפורסם, (BUROS 1978)  
להוצאות הראשונות של ה-MMY משנות ה-20 וה-30, מבחין בנקל בגידול עצום, הן  
בכמות המבחנים הן במידת גיוונם.

האם לפנינו סיפור של הצלחה?

במאמר זה ננסה לבדוק את השאלה לעומק. נעשה זאת על ידי שימוש בכמה "אמצעים  
דידקטיים":

ראשית, נביא כאן תרגום חופשי (מקוצר מעט) של מאמר פרספקטיבה של אן אנסטסי  
(Anne Anastasi) - המלכה הבלתי-מוכתרת של המבחנים הפסיכולוגיים, שהתפרסם  
לאחרונה בארה"ב. במאמר מצביעה אנסטסי על ההכרה הגוברת והולכת בהקשר של  
ההתנהגות, בכל מדידה פסיכולוגית. לתרגום זה נוסיף שני קטעים אשר עובדו מתוך  
מאמרי סקירה אחרים של אותה המחברת, על פי בקשתה.  
לאחר מכן, נביא פרק השלמות, בו יושם דגש על הבעיות ועל נקודות התורפה - החלקים  
ה"עצובים" בסיפור ההצלחה שלנו.

הפרק השלישי והאחרון יעסוק במצב בישראל. פרק זה יפתח בהצגת דו"ח ועדת מנכ"ל  
משרד החינוך לנושא השימוש במבחנים פסיכומטריים במסגרת בתי הספר ("ועדת באשי")  
ויסיים בשרטוט כמה קווים אופייניים לשימוש במבחנים בישראל והערכת התפתחויות  
עתידיות בתחום.

מגמות התפתחות אחדות במדידה פסיכולוגית:

בחמישים השנים האחרונות (1)

מאת א. אנסטסי

בבואי לבחון התפתחויות במדידה פסיכולוגית משנות ה-30 עד שנות ה-80, זיהיתי מגמות עיקריות אחדות שנראו בעיני משמעותיות. לצרכי דיון זה חילקתי אותן בהכללה לשלוש: (1) הרקע, או המסגרת ההקשרית של עריכת מבחנים; (2) נורמות של אוכלוסיות ושינויים בהישגיהן לאורך שנים; ו-(3) טיבען והיווצרותן של תכונות פסיכולוגיות.

כל אחד משלושת התחומים ניתן לבדיקה משתי נקודות מבט: אחת מתייחסת לדרכי היישום של המדידה הפסיכולוגית: השלכות על בניית מבחנים, השימוש בהם, פירוש הציונים במבחנים וכדומה. האחרת מתייחסת לתיאוריה פסיכולוגית ולמחקר בסיסי: קידום הבנתנו את התכונות הפנימיות - מהותיות ואת האטיולוגיה של ההתנהגות הנמדדת באמצעות מבחנים.

תקופת יובל השנים הנסקרת במאמר זה כוללת מחלוקות, מסקנות נוגדות והתחלות מוטעות לרוב. אף-על-פי-כן, כשמביטים בהן ממרחק מסויים, נוכחים שניסיונות מוטעים אלה עזרו במידה ניכרת לעצב את פני השטח ושתרומותיהם החיוביות הולכות ונקלטות בזרם העיקרי של מדידה פסיכולוגית מודרנית.

---

(1) תרגום חופשי של מאמרה של Anne Anastasi:

Some Emerging Trends in Psychological Measurement

אשר הופיעה ב: Applied Psychological Measurement 9, 1985, 121-138

1.1 כללי

כבר מתחילת השימוש במבחנים פסיכולוגיים, נהגו החוקרים להשתמש במבחנים לצורך השוואת אוכלוסיות נבחנים מתרבויות שונות. המסקנות מהשוואות כאלו היו בעצם מהותן דו-משמעיות, מאחר ואבחנה תרבותית ואבחנה ביולוגית בדרך-כלל שזורות זו בזו. שתיהן נובעות בעיקר מבידוד אוכלוסיות אנושיות בדרך של הפרדה גיאוגרפית, מנהגים, מסורות או תנאים אחרים, אשר כולם ביחד מקטינים שכיחות הכלאה בין קבוצות (interbreeding) ומגע תרבותי (Anastasi, 1958a, chap. 16). למרות שכבר מלכתחילה הכירו חוקרים אחדים בקשיים הקשורים בהפרדת גורמים ביולוגיים מגורמים תרבותיים, בבואם להסביר הבדלי הישגים של קבוצות שונות במבחנים, הרי ההדגשים שניתנו שיקפו לעתים קרובות את האוריינטציה התיאורטית של החוקר ואת האווירה החברתית של התקופה. כך, המתקרים ההשוואתיים הראשונים, מתחילת המאה ועד שנות ה-50 וה-40, זיהו את קבוצות המחקר כגזעים על בסיס תכונותיהם הפיזיות הנראות לעין, והניחו שיש בסיס ביולוגי להבדלים בהישגים. במחקרים מאוחרים יותר גברה הנטייה לזהות קבוצות תרבותיות ולחפש את ההסברים להבדלים בין קבוצות בהיסטוריה ההתנסותית האופיינית של חבריהן.

חלקית, בשל כלי המדידה אשר נהגו אז וחלקית, בשל תיאוריות שרווחו באשר להבדלים בין-גזעיים, עסקו המתקרים הראשונים בהבדלים הישתיים ותפיסתיים. בין המתקרים הראשונים היו אלה שנערכו ביריד העולמי בסיינט-לואיס ב-1904 (Woodworth, 1910; Bruner, 1908; ראה גם Anastasi, 1958a, chap. 17), אך עד מהרה נמצא כי הישגיהם המצוינים של אוכלוסיות ילידים מסוימות בשימוש במידע חזותי, שמיעתי או בנתוני ריח, ע"מ לזהות חפצים, לפרש עקבות, או למצוא את דרכם, נבעו לא מאיברי חישה עדיפים, אלא מתגובות נלמדות משוכללות.

עם הופעתם של מבחני המשכל הראשונים, דיווחו מתקרים רבים מאוד על הבדלים ממוצעים בציוני מבחן גלובליים או במנות המשכל בין קבוצות שמוינו לפי גזע, לאום, מיקום גיאוגרפי או קטגוריות ביולוגיות או תרבותיות אחרות (Anastasi, 1958a; Mern, 1940; Klineberg, 1935; Garth, 1931). בין המתקרים הראשונים מוצאים לעתים קרובות רשימות של קבוצות לאומיות וגזעיות, כשהן מדורגות לפי ציונים ממוצעים מן הסוג הנזכר לעיל. עם התפתחותם של מבחנים נפרדים לכשרים שונים, התברר עד מהרה שמקומן היחסי של קבוצות השתנה בהתאם לכושר הייחודי שנבחן. התפתחות מתודולוגית מאוחרת יותר ומשמעותית יותר היתה תקר ההבדלים בין קבוצות בהישגיהן במבחנים ובמקביל, גם תקר ההבדלים בהיסטוריה ההתנסותית או הלימודית שלהן. פסיכולוגים שעסקו במחקרים השוואתיים הלכו והתמקדו בחוויות המיוחדות לתרבות ספציפית, המתקשרות עם הבדלים טיפוסיים בהתנהגות קוגניטיבית

(Cole & Bruner, 1971; Rebelsky & David, 1976; Cole & Scribner, 1974; Anastasi, 1983a).  
קיימות ראיות לכך שחינוך פורמלי מערבי ממלא תפקיד חשוב בעידוד חשיבה מופשטת ומושגית. למשל, מחקר על תרבויות קדם-כתביות מצביע על כך שבני תרבויות כאלה, אשר נחשפו לחינוך מערבי, נוטים יותר להגיב במונחי מושגים רחבים והם פחות קשורים למסגרות קונטקסטואליות ייחודיות מאשר בני-גילם אשר קיבלו חינוך מסורתי קדם-כתבי. (Cole, Gay, Glick & Sharp, 1971; Cole & Scribner, 1974); Cole, Greenfield, 1966; Scribner, 1974; Sharp, Cole & Lave, 1979). במלים אחרות, אצל הראשונים התפתחה חשיבה מוכללת ומופשטת יותר.

יש לציין כי החינוך בבית-הספר מתרחש בדרך-כלל בזמנים ובמקומות המיועדים לכך במיוחד, ומצב זה מבדד את הלומד מן המסגרות הספציפיות של חיי היום-יום (Bruner, 1966; Scribner & Cole, 1973). בתרבויות קדם-כתביות, ללא חינוך בית-ספרי, מתרחש חלק רב מן הלימוד המסורתי באמצעות ביצוע תפקידים אמיתיים על-ידי הילד והשתתפותו בפעילויות של מבוגרים. לעומת זאת, בהקשר של בית-הספר, אפילו בעיות מעשיות נשקלות במסגרת נייטרלית, המנתקת את הבעיות מהרקע הטבעי שלהן. כך מוצגת הדעת מנקודת ראות פחות מוגבלת, ומוערכת במונחים של קריטריונים הראויים לשימוש באופן אוניברסלי. הישגים טכנולוגיים מסוימים, ובמיוחד שיטות כתב פונטיות ושיטות מיספור, מקלים על התפתחותן של תפיסות הכרתיות מוכללות ושל ניתוח הגיוני-לוגי. תשומת-לב מיוחדת ניתנה להשפעות חיוביות של שפה כתובה, לעומת תקשורת בעל-פה, כאמצעי הסמלה והעברת הידע המצטבר של תרבות (Scribner & Cole, 1981; Olson, 1976, 1977).

אף-על-פי-כן, נראה שעצם קיומן של מערכות סימנים מילוליות ומספריות, משפיעה אך מעט על התפתחות כשרים קוגניטיביים מקיפים, אם מערכות סימנים אלה מוגבלות למסגרות של חיי המציאות, כגון התכתבות או ניהול תשכונות (Lancy, 1983; Reed & Lave, 1979; Scribner & Cole, 1978, 1981). כאשר המסגרת היא חופשית יחסית, משפיעים הישגים טכנולוגיים אלה באופן משמעותי על התפתחות החשיבה המופשטת.

## 1.2 מבחנים הוגנים

למרות שהתפתחויות חברתיות ומדיניות באמצע המאה המריצו מאוד התעניינות במבחנים בין-תרבותיים, הוכרה הבעיה כבר בשנת 1910. אחדים מהמבחנים הבין-תרבותיים הראשונים נועדו לבחון את גלי המהגרים אשר הגיעו לארצות-הברית בתחילת המאה הנוכחית. (Anastasi, 1954a, chap. 10; Knox, 1914.) ראה גם

בדרך כלל תוכננו מבחנים בין-תרבותיים על מנת לנטרל אחד או יותר מהפרמטרים שלאורם נבדלות התרבויות. הפרמטרים העיקריים הם השפה, ידיעת קרוא וכתוב (בשפה כלשהי), שימת דגש יחסי על ערך המהירות ותכנים תלויי-תרבות. הניסיונות הראשונים לסלק תכונות תלויות-תרבות אלה הביאו לתכנון המבחנים ה"נקיים מהשפעה תרבותית" (culture free) הקלאסיים. (Anastasi, 1954a, chap. 10; pp. 286-297.) הדוגמאות כוללות את ה-Leiter International Performance Scale, ה-Culture-Free Intelligence Test של Cattell, ה-Progressive Matrices של Raven וה-Draw-a-Man Test של Goodenough.

ניסיונות ראשונים אלה לפתח מבחנים שאין בהם הטיה תרבותית נכשלו הן מבחינה תיאורטית, הן מבחינה אמפירית. מבחינת תיאורטית - לא יכול להיות מבחן נקי לגמרי מהטיה תרבותית, מפני שההתנהגות האנושית היא תלויה תרבות. כל תרבות דורשת ומחזקת את דפוס הכשרים האופייניים לה. מרבית המבחנים שאין בהם כביכול הטיה תרבותית, איכזבו (Anastasi, 1982, chap. 10). הם היו בדרך-כלל טעוני כשרים מרחביים ותפיסתיים, יותר מכשרים מילוליים ומספריים, והוכיחו את עצמם ככחות תקפים ממבחי המשכל המסורתיים, בכל הקשור לחיזוי קריטריונים חינוכיים ומקצועיים. לעיתים קרובות, לא הגיעו קבוצות טעונות טיפוח מבחינה תרבותית להישגים טובים יותר במבחנים "נקיים", כביכול, מהטיה תרבותית, מאשר במבחני משכל אחרים. בסופו של דבר, המונח "נקי מהטיה תרבותית" הושמט משמות המבחנים, כמו גם מהלקסיקון הפסיכומטרי הכללי. המבחן של Cattell למשל, מכונה עתה "מבחן משכל הוגן מבחינה תרבותית" (Culture Fair Intelligence Test). אחדים ממחברי המבחנים המקוריים חזרו בהם מטענותיהם הקדומות (למשל - Goodenough & Harris, 1950). תיקונים ויישומים מאוחרים יותר, כווננו יותר ויותר כלפי מטרות מדידה אחרות. למשל, מבחן ציור-איש Goodenough-Harris (צורתו הנוכחית של ה-Draw-a-Man Test) משמש כיום בראש ובראשונה לשימוש קליני.

### 1.3 הטיית במבחן

בחברות פלורליסטיות כמו החברה האמריקנית קשורות הבעיות המעשיות של עריכת מבחנים בין-תרבותיים בעיקר לאוכלוסיות של מיעוטים תרבותיים. בהקשר זה הוטל ספק בהתאמתם של המבחנים הפסיכולוגיים לקבוצות טעונות טיפוח. אולם יש לציין, כי מצוקה תרבותית היא מושג יחסי; מבחינה אובייקטיבית, יש רק שוני תרבותי. כל תרבות או תת-תרבות מטפחת ומעודדת התפתחות של התנהגות המותאמת לצרכיה ולערכיה. כאשר יחידים נאלצים להסתגל ולהתחרות בתוך סביבה תרבותית שונה מהחברה שבקרבה חונכו, עשוי שוני תרבותי להפוך מצוקה תרבותית.

בהקשר זה התעוררה שאלת ההטייה במבחנים (Anastasi, 1982 pp. 183-191; Anastasi, 1985a, pp. 109-111). במהותה, הטייה במבחן מתייחסת להבדלים בהישגים במבחנים בין קבוצות, אשר אינם משתקפים בתחום ההתנהגות שהמבחן נועד למדוד אותו. ליתר דיוק, הגינות המבחן (או העדר הטייה תרבותית) משמעה שהמבחן תקף באופן זהה לגבי כל הקבוצות אשר ביחס אליהן ייעשה שימוש במבחן; משמעות נוספת היא שהמבחן אינו חוזר חיזוי חסר (underprediction) או חיזוי יתר (overprediction) את ביצוע הקריטריון של קבוצה כלשהי. במונחי מודל הרגרסיה הידוע, מתייחסות שתי מטרות אלה להימנעות מהטיית שיפוע והטיית אינטרספט. (ראה: Anastasi, 1982, pp. 182-191; Anastasi, 1985a pp. 109-111; Cleary, 1968; Petersen & Novick, 1976).

מתכנני מבחנים השתדלו מאוד להבטיח שמבחניהם יקיימו תנאים אלה. כיום יש מתקרים רבים המוכיחים שתנאים אלה מתקיימים במידה מספקת ברוב מבחני הכשרים הנוכחיים, (1) הן בהקשר לקריטריונים השכלתיים הן בהקשר לקריטריונים מקצועיים. גישה זו שונה מהניסיונות הקודמים לתכנן מבחנים נקיים מהטייה תרבותית או הוגנים מבחינה תרבותית. עקב המאמץ לכלול במבחנים כאלה רק אותן הפונקציות אשר אינן שכיחות בשום תרבות, או משותפות לתרבויות רבות, עשוי היה בונה המבחן לבחור תוכן, שמידת הרלוונטיות שלו לקריטריון כלשהו העשוי להימדד, היא נמוכה. פתרון אפקטיבי יותר הוא לבחור תוכן אשר הינו רלוונטי לקריטריון, ואז לחקור הבדלים אפשריים בין קבוצות ברגרסית ציוני מבחנים על מדדי הקריטריון.

---

(1) למעשה, באותם מתקרים אשר גילו הטיית אינטרספט משמעותית, נחזה בדרך

כלל ביצוע הקריטריון של קבוצות מיעוטים חיזוי יתר, ובכך מועדפים חברי אותן קבוצות בהחלטות ברירה המבוססות על אותה נקודת תתך לכל המועמדים (באשר לרציונאל התיאורטי וממצאים אמפיריים, ראה Anastasi, 1982, pp. 187-189; Anastasi, 1985a, pp. 109-111; Linn & Werts, 1971).



#### 1.4 עריכת מבחנים בהקשרים מוגדרים

השאלה שמעוררת עריכת מבחנים למיעוטים תרבותיים מהווה רק היבט אחד של שאלה רחבה יותר. במדידה פסיכולוגית, כמו כל תחום בפסיכולוגיה, יש להתחשב במסגרת התרבותית. כדי לנסח הכללות בנושא ההתנהגות האנושית, צריך לאסוף נתונים והשערות של השוואות תרבותיות שונות, מגוונות. (Perl, Trickett & Watts, 1984; Azuma, 1984; Russell, 1984).

אך למטרות מעשיות, המבחנים היעילים ביותר עשויים להיות מבחנים שפותחו לתכליות מוגדרות היטב ולשימוש בהקשר מוגדר. למרות שמבחנים אלה יהיו שונים באשר למידת הכלליות שלהם, אין אף אחד מהם מסוגל להקיף את כלל הגזע האנושי. מה שחשוב הוא לזהות את המקום (locus) ואת הטווח של ההקשר התרבותי (או כל הקשר התנסותי אחר), אשר לו מתאים מבחן נתון כלשהו. בבחירת מבחנים למטרה ספציפית ובפירוש הישגים במבחן, חייב הבוחן להכיר את הקשרים התרבותיים הרלוונטיים.

אנו חיים בחברה פלורליסטית לא רק בתוך מדינות גדולות והטרוגניות כמו ארצות-הברית, אלא גם בתוך החברה הרחבה יותר, הכלל-עולמית. מאמצע המאה הלך וגדל הצורך למדוד כשרים של אנשים בעלי רקע תרבותי שונה. השימוש במבחנים כדי להגיע לניצול מקסימלי של הפוטנציאל האנושי במדינות מתפתחות מוכיח זאת. מוסדות החינוך המתרחבים במהירות באותן ארצות, דורשים עריכת בחינות כניסה בנוסף על מבחנים לצורכי ייעוץ פרטני. עם הגידול בתיעוש, הולכת וגדלה הדרישה למבחנים שיסייעו בבחירת מקצוע והצבת כוח אדם, במיוחד בתחומי המכניקה והפקידות ובתחומים עתירי-ידע.

כל המבחנים מודדים את יכולת הביצוע של הפרט בעת המבחן. לעומת זאת, פירוש הישגיו במבחן עשוי להיות כרוך בהתייחסות אחורנית או קדימה. למשל, אם ברצוננו להבין מדוע רמת ביצועיהם של אנשים מסוימים היא כפי שהינה, עלינו לבדוק את התנסותם בעבר; לעומת זאת, אם מטרתנו היא להעריך "קדימה" את התאמתו של הפרט למסלול לימודי, לתפקיד ספציפי, לקריירה, לארץ שהוא מתכוון להגר אליה או לתרבות מתועשת במדינה מתפתחת, עלינו לברר אם הוא רכש את הכשרים הקוגניטיביים והידע הדרושים למעבר כזה. מבחנים המשמשים למטרות אלה צריכים להתבסס על דרישות ההקשר החדש אשר אליו רוצה הפרט להיכנס ואשר בתוכו הוא אמור לתפקד.

מנקודת הראות של המחקר הבסיסי, ניתן להשתמש במבחן בין-תרבותי כדי לחקור את תרומתם של נתונים ביוגרפיים או של תולדות למידה של הפרט, לעיצוב תכונות פסיכולוגיות, להתפתחות כשרים ספציפיים או לגיבוש האינטליגנציה של האדם. מחקר כזה דורש מבחנים פנימיים (indigenous) המיועדים להעריך מבנים התנהגותיים, משמעותיים מבחינה חברתית, בתוך כל תרבות. בתרבויות מערביות מתקדמות, המבנים הללו מתאפיינים בכושר לימוד, בכושר מילולי ובהיגיון כמותי. העברת מבנים אלה באופן אוטומטי (באמצעות מבחנים), למחקרים על תרבויות אחרות לא תיצלח לגילוי כשרים קוגניטיביים וידע המעוגנים בתרבויות אלה. על מנת לזהות דרישות סביבתיות בהקשרים תרבותיים שונים, יש צורך להכין מלאי של ניתוח תפקידים וטקסונומיות של ביצועים אנושיים. (Fleishman, 1975; Fleishman & Quaintance, 1984; McCormick, Jeanneret & Meacham, 1972; Pearlman, 1980).

מחקרים כאלה אינם דרושים רק להבחנה בין תרבויות בארצות שונות, אלא גם להבחנה בין תת-תרבויות באותה ארץ, כאשר החברה בה היא בעלת התפתחות גבוהה ותחומי ההתפתחות בה מגוונים. לדוגמה, עם ההתעניינות ההולכת וגדלה בפסיכולוגיה התפתחותית לאורך-החיים, (life-span developmental psychology) והתרחבות תאור המבוגרים, גדל והולך הצורך בטקסונומיות מיוחדות להתנהגות תעסוקתית ולפעילויות אחרות של מבוגרים. ניסיון לפתח מבחנים באמצעות ניתוח תפקידים (task analysis) של הפעילויות היומיומיות של אנשים מבוגרים, מדגים את ההבדלים בדרישות קוגניטיביות בשלבי חיים שונים (Demming & Pressey, 1957).

לסיכום, בניית המבחנים צריכה להתבסס על ניתוח שיטתי של דרישות סביבתיות-תרבותיות. גישה זו תעודד את פיתוחם ויישומם של מבחנים לצרכים מוגדרים בהקשרים מתאימים, הן בתוך תרבויות הן בין תרבויות.

## 2. נורמות של אוכלוסיות ושינויים בהישגיהן לאורך שנים

נורמות מבחן משקפות את אופן הביצוע של אוכלוסייה מוגדרת בזמן נתון ובמקום מוגדר. שינויי נורמות של אוכלוסייה לאורך זמן, נובעים הן משינויים בהרכבה, הן משינויים חברתיים. לא רק מתכנן המבחן, אלא גם המשתמש במבחן, חייב להיות ער לתנאים שמשפיעים על נורמות. בבואו לפרש את ההישגים במבחן, עליו להביא בחשבון את ההשפעות הייחודיות שהיו עשויות לפעול על האוכלוסייה שנבחנה.

נניח לדוגמה, שחוקר מגלה כי מועמדים מסוימים לבית-ספר להנדסה מגיעים להישגים העולים על הנורמה במבחן מסוים. החוקר ישקול בוודאי את ההשערה, שברירה עצמית מחמירה היא שגרמה לכך. אלא שבאותה המידה ייתכן שרמת ברירה עצמית נמוכה יותר איפיינה את המדגם הנורמטיבי, אולי בשל לחצים חברתיים מיוחדים שפעלו במשך תקופה קצרה לא-אופיינית. הלקח מדוגמה זו הוא פשוט: נורמות מבחן מייצגות נתונים אמפיריים שהושגו בזמן נתון ובמקום מוגדר. למרות ההבנה, שנורמות אינן קבועות ומוחלטות, יש עדיין נטייה לראות נורמות מסוג זה כחסינות יחסית.

### 2.1 שינויים חברתיים ופירוש ציוני מבחן

2.1.1 מהדורות מעודכנות של מבחנים ומנות משכל משתנות. שינויים באוכלוסייה באשר להישגי מבחן, עשויים להשפיע על פירוש ציונים בדרכים שונות. אחת הדוגמאות הידועות ביותר היא השימוש במהדורות מעודכנות של מבחני משכל. בתיקון מחדש של מבחן סטאנפורד - בינה (1972 לעומת 1937) וגם במהדורת ה- Wechsler Intelligence Scale for Children Reviser (1974 לעומת 1949), המדגם הנורמטיבי המאוחר יותר הגיע להישגים גבוהים יותר מאשר המדגם שלפניו.

פירושו של דבר, שאותם ילדים היו מראים מנת-משכל נמוכה יותר אילו נבחנו לפי המהדורה המעודכנת, רק מפני שבצועם היה מוערך לפי נורמות גבוהות יותר. אחת הסיבות לעלייה בנורמות היתה רמת ההשכלה הגבוהה יותר של הורי הילדים במדגם הנורמטיבי המאוחר.

השפעתה של רמת השכלה כללית ההולכת ועולה משתקפת ישירות בהישגיהם של מי שנכללים במדגמים הנורמטיביים של המהדורות השונות. החשכות על פירוש מבחנים הודגמו במחקר עם ה- Wechsler Adult Intelligence Scale (Wechsler, 1981). 72 נבחנים, בגילים 35 עד 44, אשר נבחנו לפי מהדורות 1955 ומהדורות 1981, השיגו מנות - משכל ממוצעות בגובה 111.3 במהדורה הקודמת ו-103.8 לפי המהדורה שלאחריה. הבדל זה של 7.5 נקודות מרמז שהנורמות הקודמות היו נמוכות מאלה שנקבעו לאחריה.

2.1.2 הזדקנות והידרדרות אינטלקטואלית. פירוש מוטעה אחר של נורמות, מודגם על ידי הזיהוי של משתנה הגיל עם משתנים אחרים כלשהם, אשר מבחינים בין קבוצות הגיל השונות. (Anastasi, 1956a; 1958a, pp. 240-243; 1982, pp 333-339). ב-WAIS, למשל, נבחני המדגם הנורמטיבי המבוגרים יותר סיימו בממוצע מספר קטן יותר של שנות לימוד מאשר הנבחנים הצעירים יותר. בהתאם לכך, ציוניהם היו נמוכים מציוני הצעירים. תוצאה זו יצרה את הרושם שהאינטליגנציה יורדת מגיל 30 ואילך (Doppett & Wallace, 1955), והיא נתפרשה כתופעה הקשורה לתהליך ההזדקנות. הממצאים תמכו בירידה שעליה דיווחו מחקרי חתך (cross-sectional) מוקדמים, בהם נערכו מבחנים סטנדרטיים למבוגרים בני גילים שונים באותה תקופת זמן (Jones & Conrad, 1933; Wechsler, 1944; Miles & Miles, 1932). מחקרי-אורך מאוחרים יותר, בהם נבחנו בוגרים שוב לאחר חמש שנים עד 40 שנים, הראו בדרך כלל מגמה הפוכה: הציונים הממוצעים נטו להשתפר עם עליית הגיל. עם זאת, סברו שמהלך החיים האינדיבידואלי של נבחנים משפיע על התוצאות (ראה Anastasi, 1982, pp. 333-35).

מחקרי-חתך ומחקרי-אורך בלבד אינם יכולים לספק פירוש החלטי להבדלים בין הגילים. לאחרונה, השתמשו שי ועמיתיו בדגם ה-composite cross-sequential, הכולל מחקר-חתך ומחקר-אורך של יחידים ביחד עם השוואות בזמנים שונים (time-lay) (Schaie & Labouvie-Vief, 1974; Schaie, 1965). מתודת הזמנים השונים (time-lay) מחייבת מדידת קבוצות אנשים בני אותו גיל בתקופות זמן שונות. למשל, משווים בני 20 אשר נבחנו בשנת 1940 עם בני 20 שנבחנו בשנת 1970. באופן כללי, ההבדלים בין קבוצות גיל (cohorts) או בין הדורות, הקשורים לשינויים חברתיים רחבים, מסבירים במידה רבה את הפער בציוני מבחני כושר שכלי, אשר יוחסו בעבר לתהליך ההזדקנות.

### 2.1.3 ציונים עולים בהישגי אוכלוסיות במבחני משכל

מחקרים בקנה מידה גדול, אשר נערכו במשך כל המחצית הראשונה של המאה ה-20 בארצות-הברית ובבריטניה, גילו רמה שכלית עולה באוכלוסייה הכללית, כפי שנמדדה על ידי מבחני משכל (ראה Anastasi, 1958a, pp. 209-211). ממצאים אלה הפריכו את טענותיהם של כמה פסיכולוגים, חוקרי תורשה ומומחים לדמוגרפיה (Bradford, 1925; Burt, 1946; Catell, 1937; Dawson, 1932), בדבר ירידה אינטלקטואלית צפויה. הירידה שהם צפו התבססה על הקורלציה השלילית בין אינטלגנציה וגודל המשפחה, שדווח עליה בהרחבה.

את אי-ההתאמה בין הירידה החזויה לבין העלייה הממשית בהישגים, אפשר להסביר בעזרת שני סוגי טענות. הראשון - שמתקר על הקשר בין משכל וגודל המשפחה, כרוך במלכודות מתודולוגיות, במיוחד בקשר לבעיות בחירת המדגמים וארטיפקטים סטטיסטיים (Anastasi, 1954b, 1956b). בין המשתנים המזהמים, אשר אינם ניתנים לפיקוח, נמצאים שיעור התמותה, תדירות הרווקים וזוגות בלי ילדים, גיל ההורים ומספר המשפחות הלא-שלמות, שמספר הצאצאים הסופי בהן טרם ידוע בעת המחקר. כל אחד מהמשתנים הללו עצמם קשור לרמת משכל, לרמת השכלה ולרמה החברתית-כלכלית. הסוג השני של טענות קשור בשינויים החברתיים-תרבותיים שהתרחשו בעולם במקביל לעליית ציוני מבחני המשכל, כגון התפשטות ידיעת קרוא וכתוב, עלייה ברמת ההשכלה והתפתחות אמצעי התקשורת. נראה שהתפתחויות תרבותיות אלה מסבירות במידה רבה את השיפור שנמצא באינטליגנציה של הנבחנים<sup>(2)</sup>. מבחינה תיאורית, ייתכן, כמובן, שהכלאה סלקטיבית שלילית (adverse selective breeding) גרמה לירידה אמיתית במשכל, אך הוסתרה על ידי שינויים חברתיים חיוביים, שפעלו כמשקל-נגד.

#### 2.1.4 ציונים יורדים בהישגי אוכלוסיות במבחני משכל

שינויים החלים בממוצע הציונים במבחן משכל של אוכלוסייה, תלויים בגורמים רבים. תקופת הזמן הנחקרת והשינויים התרבותיים המתרחשים בה, היא כמובן הגורם הראשון במעלה. גיל הנבחנים משנה אף הוא. למשל, רמת השכלה עולה באוכלוסייה תשפיע ישירות על הישגיהם של מבוגרים, אך רק באופן לא-ישיר על הישגיהם של ילדים, מאחר והילדים בשכבות רצופות (קהורטים) מצויים באותה רמת השכלה פורמלית. שיקול חשוב נוסף, במיוחד במדידת תת-אוכלוסייה נבחרת, הוא כל שינוי בדרגת הסלקטיביות המתרחש בנקודות זמן שונות. למשל, בשנת 1960 למד חלק גדול יותר של האוכלוסייה האמריקנית בבית-ספר תיכון מאשר בשנת 1910 ולפיכך, תלמידי התיכון משנת 1910 מייצגים מדגם סלקטיבי יותר לגבי האוכלוסייה הכללית בת-זמנם, מאשר תלמידי התיכון משנת 1960. מספרם הרב ורמת הסיבוך הגבוהה של התנאים והגורמים אשר יכולים להסביר עלייה או ירידה במשכל הנבחן של אוכלוסייה, מודגמים היטב בניתוח ירידת הציונים המפורסמת במבחן הכושר הלימודי (SAT) של ה-College Board בשנים 1977-1963. הציון הממוצע במבחנים המילוליים ב-SAT ירד מ-478 ל-429, והציון הממוצע במתמטיקה ירד מ-502 ל-470. בניסיון להבין ירידה זו, שנמשכה באופן יציב במשך 14 שנה, בוצעו 38 מחקרים בידי מומחים בתחומים שונים ונבדקו כמה וכמה השערות סיבתיות (Wirtz, 1977).

---

<sup>(2)</sup> ניסיון קודם בהשתתפות במבחנים הוכיח את עצמו כבעל חשיבות מועטה לגבי השיפור שנמצא (ראה גם Scottis Council, 1953, pp. 121-24; Anastasi, 1956b, p. 199).

המסקנה העיקרית שכ תברי הועדה היתה, שהסיבות למגמת הירידה במחצית הראשונה היו שונות מאלה של המחצית השנייה של תקופת 14 השנים. בשבע השנים הראשונות, נבעה ירידת הציונים בעיקר משינוי שחל בהרכב האוכלוסייה שנבחנה ב-SAT. בעקבות הגידול המתמשך במספר בוגרי התיכון שהלכו למכללות במשך תקופה זו ועקב הנהגת מדיניות "הקבלה החופשית", נעשה המדגם פחות ופחות סלקטיבי. לעומת זאת, בשבע השנים של המחצית השנייה, כבר התייצבה אוכלוסיית המועמדים למכללות ורמת הסלקטיביות הסבירה חלק הרבה יותר קטן בירידת הציונים. לגבי תקופה זו, היה צורך לחפש את ההסבר בתנאים בבית, בבית-הספר ובחברה בכלל. הועדה ציינה כי הנתונים שנמצאו לא איפשרו לקבוע מה היו תרומותיהם היחסיות של שינויים תרבותיים שונים לירידת הציונים, אולם בין הגורמים הרבים שצוטטו כמשמעותיים היו: הפחתה בסטנדרטים האקדמיים, צמצום כמות שיעורי-הבית, גידול במספר ההיעדרויות מבית-הספר, הקדשת פחות תשומת-לב לרכישת מומחיות וידע, בילוי ממושך מול מרקע הטלוויזיה וגידול בכמות מסיחי דעת נוספים בחייהם של התלמידים. מן הראוי להוסיף שירידת הציונים, אף כי נחקרה במיוחד בהקשר ל-SAT, לא הוגבלה לכלי מדידה זה בלבד. לא זו בלבד שהתרחשה ירידה במבחני כניסה אחרים למכללות; אלא שגם בהישגים במבחנים בתיכון ובבית-הספר היסודי היתה ירידה במהלך אותה התקופה.

## 2.2 שינויי אוכלוסייה לגבי משתנים רגשיים (affective)

מרבית מחקרי-האורך עוסקים במשתנים הכרתיים (קוגניטיביים), ובמיוחד אלה שבאים לידי ביטוי במבחני משכל סטנדרטיים. עם זאת, היו מספר התחלות מבטיחות בחקר שינויים באוכלוסייה באשר לתכונות רגשיות. בחקשו זה משתמשים במלה affect במובנה הרחב, החל על כל התכונות שאינן קוגניטיביות: תחושות, מצבי-רוח, רגשות, עמדות ומניעים. לדוגמה:

### מחקר מוקדם

מחקר שכבות (קוהורט) חלוצי נערך עם השאלון X-0 של Pressey & Jones, (1955). משתתפי המחקר קיבלו הוראות לסמן את כל הפריטים שנחשבו בעיניהם ל"רעים", בשאלון של 125 פריטים. מרבית הפריטים התייחסו לפעולות גבוליות, כגון: עישון, שתייה, יריקה ולעג לזולת.

השאלון הועבר בשנת 1953 לתלמידי מכללה ולקבוצות מבוגרים, בגילים 20-60. אותו השאלון הועבר לקבוצות דומות של תלמידי מכללה בשנים 1923, 1933 ו-1943. הנתונים איפשרו השוואות בין-גיליות בתוך מדגם שנת 1953 והשוואות לאורך (בין-דוריות) של תלמידים שנבחנו במשך ארבעה עשורים.

ההשוואה הבין-גילית מראה, שעם עליית הגיל ניכרת נטייה ברורה לשפוט יותר דברים כ"רעים". לעומת זאת, חקר אוכלוסיית התלמידים לאורך התקופה גילה הבדלים חדים בין הדורות וירידה תלולה במספר הפריטים שנחשבו ל"רעים" במשך ארבעת העשורים. מעניינת במיוחד היא השוואת המבוגרים שנבחנו בשנת 1953 לסטודנטים שהיו במכללה, כאשר מבוגרים אלה היו בשנות העשרים לחייהם. למשל, בני ה-50 בשנת 1953 ציינו אותו מספר הפריטים שציינו תלמידים במכללה בשנת 1923. אין, אפוא, ראיות לכך שאנשים נוטים להיות שמרנים יותר בעמדותיהם ככל שהם מזדקנים. יש רמזים לכך שאנשים מבוגרים נוטים להחזיק באותן הדעות, שהינן אופייניות לתפיסות של התקופה בה הגיעו לבגרות.

### 2.3 הבדלים בין המינים ושינוי תרבות

חקר הבדלים בתכונות פסיכולוגיות בין המינים, הוא תחום מבטיח במיוחד ללימוד השפעותיהם של שינויים חברתיים על ההתנהגות לטווח ארוך (Anastasi, 1981). התנועה הפמיניסטית העניקה תנופה לחקר ההבדלים בין המינים במיוחד בשנות ה-60, אולם לאחרונה מורגש צורך להפעיל דרכי מחקר חדשות. מחקרי אורך של שכבות לגבי הבדלים בין-מיניים לאורך תקופות קריטיות של שינויים חברתיים, יכולים לתרום, כנראה, להבנת האטיולוגיה של הבדלים בהתנהגות (במידה שהם מתגלים). כאשר הנתונים מוגבלים לקונטקסט אחיד ויציב יחסית, עשויות התוצאות להישאר ברמה תיאורית. מסיבות אלו ומעשיות ברורות, אסור שחקר בני-אדם יחשוף את המשתתפים לשינויים דרסטיים ומתמשכים בתנאי החיים. לעומת זאת, שינויים חברתיים עשויים לספק הבדלים במשתנים התנסותיים קיצוניים יותר ומתמשכים יותר, שהשפעתם על ההתפתחות הפסיכולוגית רבה יותר מכפי שניתן להשיג באמצעות הפעלה ניסויית. לכן, מחקרי אורך של שכבות, המתפרשים על פני עשור אחד או שניים - תקופה בה מתרחשים שינויים חברתיים משמעותיים - מייצגים "ניסוי" טבעי בעל ערך. אפילו השוואה פשוטה של ממצאים מ-25 השנים האחרונות, עם ממצאי מחקרים קודמים שפורסמו, עשויה לתרום להבנת מקורותיהם של הבדלים פסיכולוגיים בין המינים, אם כי מחקרי אורך שיטתיים של שכבות, עם מדגמים דומים ובכלים מתוכננים היטב, יספקו, כמובן, תשובות מדויקות יותר. כבר הוחל במחקרי אורך של שכבות בהבדלים בין-מיניים במשתנים אפקטיביים. התוצאות אמנם מוגבלות, אך מעוררות מחשבה. בין הדוגמאות: העדפותיהם של ילדים לגבי פעילויות משחק שונות (Rosenberg & Sutton Smith, 1960); דחף להישגים (achievement drive); מה שמכונה "פחד מהצלחה" (Spence, 1974); תפיסת העצמי ותפיסת תפקידי המינים (Urberg & Labouvie-Vief, 1976). (Urberg, 1979).

### 2.3.1 השלכות על ההתפתחות ההכרתית (קוגניטיבית)

שינויים רגשיים עשויים להביא לשינויים הכרתיים. מאגר נתונים הולך וגדל, מרמז על כך שלא רק מצבים אפקטיביים מתחלפים משפיעים על ביצוע במבחני משכל, אלא שתכונות אופי יציבות משפיעות על התפתחותם של כשרים (ראה, Anastasi, 1982; pp. 352-354; 1985b). מדענים החוקרים את הבדלי המינים הקדישו תשומת-לב מיוחדת לאיסוף ראיות על הקשר בין משתנים אפקטיביים וקוגניטיביים. למשל, נמצאו מתאמים משמעותיים בין משתנים אפקטיביים כגון מדדי גבריות-נשיות, הזדהות עם תפקידי המינים ותפיסת תפקידי המינים, לבין הישגיהם של גברים ונשים במבחנים: בפתרון בעיות (Milton, 1957), כשרים מרחביים (Nash, 1979) וכשרונות מילוליים וכמותיים (Dwyer, 1974; Entwisle & Baker, 1983; Fitzpatrick, 1978; Nash, 1979). באחדים מהמקרים הללו, היחס בין המשתנים האפקטיביים והקוגניטיביים נמצא לא רק כאשר הושוו קבוצות של גברים ונשים, אלא גם בתוך כל מין ומין. פועל יוצא מממצאים אלה הוא, שככל שמשנתות תפיסות וציפיות לגבי תפקידי המינים בעקבות תהליכים חברתיים, כך עשויים להשתנות גם הבדלים בין-מיניים באשר להתפתחות כשרים מסוימים (Anastasi, 1981).

### 2.3.2 השלכות מתודולוגיות לגבי הבעיה של תורשה מול סביבה

היחסים בין התכונות האפקטיביות והקוגניטיביות, עשויים גם הם להאיר את הדרך לחקר תרומתם של תנאים תורשתיים וסביבתיים להבדלים אינדיבידואליים. במידה ויוכח כי עמדות והנעות משפיעות על התפתחות כשרים שכליים, ייתכן שניתן יהיה לטעון, כי הן מייצגות חוליה מקשרת בשרשרת האירועים, מגורמים גנטיים ופיסיולוגיים בסיסיים עד הבדלים אינדיבידואליים סופיים בתיפקוד קוגניטיבי. גם לפני שיכולנו להעריך באופן אמפירי הבדלים אינדיבידואליים בכשרים, היה מקור ההבדלים הללו בתורשת הפרט לעומת סביבתו נושא לוויכוח חריף, הנמשך בעוצמה עד היום. לפני כ-25 שנים טענתי, שיתכן שחוקרים אינם שואלים את השאלות הנכונות בקשר לתורשה והסביבה ושבהצבת השאלה "איך?", תבוא לידי ביטוי, אולי, גישה פורייה יותר (Anastasi, 1958b). במקום לשאול אֵלן הבדלים הם תורשתיים ואלו נרכשים, או באיזו מידה ניתן לייחס את ההבדל לתורשה ובאיזו מידה לסביבה, טוב יותר יהיה לחקור את דרך פעולתן של השפעות תורשתיים וסביבתיות על התפתחות הבדלים אינדיבידואליים. לאור הידוע עד כה בנושא התפתחות ההתנהגות האנושית, אין זה מתקבל על הדעת שגנים יכולים לקבוע ישירות הבדלים אינדיבידואליים בין כשרים מילוליים או מתמטיים, למשל. מה שנחוץ במידה רבה יותר הינו מידע על השלבים המתווכים הרבים בשרשרת האירועים האטיולוגית מן הגנים ועד להתנהגות. ייתכן שתפקיד המוטיבציה מייצג שלב אחד כזה.



חוקרים אחדים הכירו בתרומת המוטיבציה להתפתחות השכלית, אך תפקידה כמשתנה מתווך הוכר באופן המפורש ביותר על ידי הייאס (Hayes, 1962 p. 302). הוא טען שהאינטליגנציה נרכשת באמצעות ההשכלה ושהרכב ההנעתי, שהוא פרי תורשה, משפיע על טיב ההשכלה הנרכשת ועל כמותה. הבסיס התורשתי של האינטליגנציה מורכב מדחפים, ולא מכשרים כשלעצמם.

השתמעות מתודולוגית תשובה של שאלת התורשה-סביבה, שנוסחה כאן מחדש, היא שתשומת-הלב של החוקרים צריכה להיות מכוונת עתה אל השפעות סיבתיות ספציפיות ואל משתנים ספציפיים במקום אל התחומים הרחבים, ההטרוגניים והגלובליים, היינו, "תורשה וסביבה" בכלל. (ראה גם, Lewontin, 1974; Mackenzie, 1984, במיוחד עמ' 1230-1232).

### 3. טיבן ותבניתן של תכונות פסיכולוגיות

#### 3.1 זיהוי גורמים וארגונם

השיטות הסטטיסטיות המקובלות של ניתוח גורמים התפתחו מתוך מחקרים על טבע האינטליגנציה האנושית והרכבה. תיאוריית התכונות הראשונה שהתבססה על ניתוח סטטי של ציוני מבחן, היתה תיאוריית שני הגורמים שהציע הפסיכולוג הבריטי צ'רלס ספירמן (Spearman, 1904, 1914, 1927). בתיאוריה הועלתה הטענה, שכל הפעולות השכליות משתקפות בגורם משותף יחיד, המכונה הגורם הכללי,  $g$ . כמו כן, הניחה התיאוריה גורמים ייחודיים ( $s$ ) רבים - כל אחד ייחודי למבחן. למרות שהתיאוריה מניחה שני סוגי גורמים, רק גורם ה- $g$  מסביר את המתאם. לפי המינוח המקובל של תיאוריות תכונות, אפשר לראות בתיאוריה זו תיאוריית גורם יחיד (single factor theory), אך הכינוי המקורי (two factors theory) עדיין דבק בה. בשלב מוקדם בפיתוח המתודולוגיה שלו (Spearman, 1927), הכיר ספירמן בכך שהיה צורך בסיוג ה- $two\ factors\ theory$  שלו. כאשר המבחנים דומים מאוד, אפשר למצוא מידה מסוימת של קורלציה, מעל ומעבר לזו שניתן ליחס ל- $g$ . מכאן, ייתכן שקיים גורם מסוג אחר, פחות רחב מ- $g$  ופחות ייחודי מגורמי  $s$ . גורם מתערב כזה, המשותף לקבוצת פעולות שכליות אך לא לכולן, הוגדר כגורם קבוצתי (group factor). בכתביו המוקדמים, הכיר ספירמן בכך שישכולים להתקיים גורמים קבוצתיים צרים מאוד. בעקבות מחקרים שעשו עמיתיו ותלמידיו, הוא התחיל לכלול גורמים קבוצתיים רחבים יותר אשר הלמו כשרים תשובניים, מכניים ולשוניים. אף-על-פי-כן, חלק הארי של common variance יוחס עדיין לגורם  $g$ .

באותה התקופה התמקד מחקר התכונות בארצות-הברית בתיאוריית מרובות גורמים (multiple trait theories). הספר Crossroads in the Mind of Man (Kelley, 1928) סלל את הדרך בפני מחקרים שחיפשו גורמים קבוצתיים ייחודיים. הגורמים העיקרים שזיהה קלי במדגמים של תלמידי בית-ספר תוארו כגורמים מילוליים,

מספריים, מרחביים, גורמי זיכרון וגורמי מהירות. אחד החוקרים המרכזיים בתחום זה היה תרסטון. על בסיס מתקר נרחב, הוא הציע מספר קטן של גורמים קבוצתיים וכינה אותם "כשרים שכליים ראשוניים". בין אלה שאומתו לא פעם, גם בעבודתו וגם בעבודתם של חוקרים אחרים, נכללים: הבנה מילולית, שטף דיבור (word fluency), כשרים מספריים, כשרים מרחביים, זיכרון אסוציאטיבי, מהירות תפיסתית והיגיון כללי (Ekstrom, French, Harman & Dermen, 1976; Thurstone, 1938; Thurstone & Thurstone, 1941).

העשורים הבאים היו עדים ל"שגשוג" מהיר של מספר גורמים, ובמהלכו של תהליך זה חולקו אחדים מגורמיו של תרסטון לגורמים קבוצתיים צרים יותר; למשל, כאשר נערכו מחקרים אינטנסיביים לגבי תחומים שבמקור זוהו כמילולי, תפיסתי, זיכרון והיגיון.

אחת הדרכים להתמודד עם מספר גדול והולך של גורמים מודגמת על ידי מודל מבנה האינטלקט (structure of intellect) של גילפורד (Guilford & Hoepner, 1971; Guilford, 1967). שלושת מימדיו של מודל זה מייצגים פעולות, תוכן ותוצאות; ניתן למיין כל מבחן לאורך כל שלושת הממדים (או שטחות). המודל מספק בסך הכל 120 תאים, שבכל אחד מהם צפוי לפחות גורם אחד, וחלק מן התאים עשויים לכלול יותר מגורם אחד. יש להודות שמספר הגורמים הצפויים במודל הזה הוא גדול, אך גילפורד טען שהטבע האנושי מורכב מאוד ושאי-אפשר לתארו כחלכה בעזרת גורמים ספורים בלבד.

סכמה אחרת למיון גורמים משתמשת במודל היררכי, הדומה לעץ הפוך (Vernon, 1961; Hompfreys, 1962; Burt, 1949). בצמרת נמצא גורם כללי; במדרגה שמתחתיו נמצאים גורמים קבוצתיים רחבים, הדומים לכמה מן הקשרים השכליים הראשוניים של תרסטון; גורמים קבוצתיים עיקריים אלה נחלקים לגורמים קבוצתיים צרים יותר ברמה אחת או מספר רמות; הגורמים הספציפיים הם ברמה התחתונה. המודל ההיררכי מאפשר לשלב את ה-g של ספירמן עם דפוסים רב-גורמיים (multiple factor patterns). באשר למתודולוגיות ניתוח גורמים, התאמה זו התאפשרה באמצעות פיתוח טכניקות רוטציה המניבות צירים אלכסוניים (oblique axes), המייצגים גורמים מתואמים. אל הקשרים הפנימיים בין גורמים כאלה ניתן להגיע בנייתו גורמים, וכך לזהות גורמים מהדרגה השנייה. כאשר ישנם די מבחנים בסוללה הראשונית, ניתן לחשב גורמים מדרגה גבוהה עוד יותר. תרסטון (Thurstone, 1948) היה מן הראשונים שהצביעו על כך שה-g של ספירמן יכול היה להופיע כגורם מדרגה גבוהה יותר בין גורמים מתואמים מהדרגה הראשונה. לאחרונה, מאמצים מנתי גורמים אחרים מודלים היררכיים, בנצלם גורמים מתואמים ומתקדמים דרך רמות אחדות של גורמים מהדרגה הראשונה (Cattell, 1963; Guilford, 1981; Hakstian & Cattell, 1978; Horn, 1968).

אפשר לראות שההבדל בין גורם מהדרגה הראשונית וגורם מדרגה גבוהה יותר מצוי במספר המשתנים אשר תוחמים אותו (Leiman & Schmidt, 1957); גורמי הדרגה הגבוהה יותר הם במהותם רחבים יותר. החוקר רשאי לבחור אותה רמה בהיררכיה, ההולמת ביותר את מטרתו. תפיסה זו של גורמים זוכה להכרה הולכת וגדלה (Carroll & Horn, 1981; Corn, 1964; Humphreys, 1962).

יש להוסיף מילת אזהרה באשר לגורם הכללי הנמצא בסוללה ספציפית כלשהי. גורם זה תואר לעתים קרובות (מדי), באופן כללי, כ-g של ספירמן. למעשה, הוא מייצג גורם כללי, המשותף רק למבחנים של אותה הסוללה. ייתכן אפוא, שאין להסיק מניתוח כזה שמבחן נתון כלשהו טעון g של ספירמן באופן מאסיבי. למעשה, יש לנסח את התיאור הטוב ביותר של הגורם הכללי המזוהה בסוללה ספציפית כלשהי, באמצעות בחינה מקרוב של כל המבחנים, הטעונים בצורה ניכרת באותו גורם.

### 3.2 השלכות על פיתוח מבחנים והשימוש בהם

מבחינת בניית מבחנים, המודל ההיררכי משלב מסגרת תיאורטית מקיפה עם גמישות מעשית. בונה המבחן או המשתמש בו יכול לבחור אותה רמה בהיררכיה המתאימה ביותר לבחינה ספציפית כלשהי. פתרון זה מייצג את מה שנעשה בשטח הלכה למעשה. אפשר להדגים שלוש רמות של היררכיה זו באמצעות כמה סוגי מבחנים המצויים בשימוש שוטף. ברמה המוגדרת באופן הברור ביותר, נמצאות אחדות מן הטכניקות להערכת התנהגות אשר תוכננו עבור תכניות לטיפול התנהגותי (ראה Anastasi, 1982, עמ' 483-488); באותה רמת פירוט, לערך, נמצאים מה שמכונים מבחנים תלויי-קריטריון (criterion-referenced tests)<sup>3</sup>, במיוחד אלה שפותחו לשימוש בשיטות לימוד עצמי, המותאמות לקצב הפרט המשתמש בהן. בתכניות לימוד כאלה נערכים מבחנים במשולב עם הדרכה. עורכים מבחנים לפני השלמת כל יחידה לימודית, במשך היחידה ואחריה. (ראה Anastasi, 1982, עמ' 94-101).

סוללות רב-תחומיות (multiple aptitude batteries) מטפלות ברמה של היררכיית התכונות המוגדרות באופן רחב יותר. הדוגמאות כוללות את סוללת ה-Primary Mental Abilities (PMA), אשר צמחה במישרין מהמחקר המוקדם של תרסטון בנושא ניתוח גורמים (Thurstone & Thurstone, 1946-1965); סוללת Differential Aptitude Tests (DAT) (Bennett, Seashore & Wesman 1984-1947); סוללת ה-General Aptitude Battery (GATB), שהוכנה בידי שירות התעסוקה של ארה"ב (USES, 1977-1946), וסוללות מיון אחדות שפותחו על ידי הצבא.

פיתוח הסוללות הרב-תחומיות והשימוש בהן הגיע לשיאו בשנות ה-50, כאשר פסיכולוגים סברו שגורמים קבוצתיים רחבים אחדים יכולים להסביר את מרבית ההבדלים האינדיבידואליים, ושפרופיל ציוני מבחנים, המודדים את הגורמים הללו, יועיל ביותר בבעיות כגון ייעוץ חינוכי ומקצועי ושיבוץ כוח-אדם לתחומי התמחות מקצועית, אשר ינצלו ככל האפשר את דפוס הכשרים הקוגניטיביים והידע של הפרט. דרכי יישום כאלה מניחות מראש תקפות דיפרנציאלית גבוהה למבחנים הנפרדים בסוללה. אולם בהקשר זה התוצאות הכזיבו. ב-DAT, למשל, נמצא עד מהרה שמבחן ההיגיון המילולי (VR) הניב מתאמים גבוהים עם הישגים ברוב הקורסים האקדמיים, ללא קשר לתוכן. בעיקר בשל תוצאות כאלה, הוכנס כמדד כללי לכשר אקדמי ציון מורכב, המייצג את סכום הציונים במבחני ההיגיון המילולי (Verbal Reasoning) וכשרון מספרי (Numerical Aptitude).

<sup>(3)</sup> בשל השימוש הרווח במונח קריטריון בפסיכומטריקה, criterion-referenced testing איננו מונח מוצלח כל כך. במבחנים אלה, ה-frame of reference המפרש הוא תחום התוכן ולא אוכלוסיית בני-אדם. לפיכך, נכון יותר לראות אותם כמתייחסים לתוכן (context referenced) או כמתייחסים לתחום (domain-referenced).

התפתחות דומה מודגמת במחקר המתמשך עם ה-GATB. ניתוחים מחדש של הנתונים שהושגו לגבי כ- 12,000 משרות שהופיעו ב- (1977, Dictionary of Occupational Titles, U.S. Department of Labor), הראו שגורם קוגניטיבי רחב הסביר חלק ניכר משונות הסוללה והניב מתאמים בעלי תקפות גבוהה לגבי רוב המשרות. אמנם הוספת מדידות של כשרים תפיסתיים ופסיכומוטוריים שיפרה את תקפות החיזוי לגבי כמה משרות, אך הדבר אירע רק לגבי המשרות המסובכות פחות (U.S. Dept. of Labor, 1983).

בימת היירכית התכונות הרחבות ביותר, נמצאים מבחני המשכל המסורתיים. מבחינת תוכנם, מבחנים אלה דומים מאד לצירופי המבחנים אשר הוכיחו את עצמם (בדוגמאות שהוזכרו לעיל) ככלי חיזוי טובים של קריטריונים אקדמיים ומקצועיים כאחד. יש ראיות מצטברות לכך שאותו מקבץ של כשרים קוגניטיביים וידע, שנבדקים בעזרת מבחני משכל מסורתיים, מושפע מהמתרחש בחברות מודרניות מפותחות מבחינה טכנולוגית ומשפיע עליהן; חרף זאת, המונח הבלתי-מסויג "משכל" (אינטליגנציה), רחב מכדי לאפיין מבחנים מסוג זה. קיימים סוגים רבים של אינטליגנציה אנושית. מסגרות תרבותיות או התנסותיות שונות מטפחות ומחזקות מערכות כשרים שונים, אשר מהוות את ה"אינטליגנציה" בתוך אותן מסגרות. ניתן לתאר את מבחני המשכל הידועים בעולם המערבי כמדדים לאינטליגנציה אקדמית או לכושר לימוד. הם מודדים סוג של אינטליגנציה, שהינה מפותחת באמצעות החינוך הפורמלי וגם מהווה תנאי מוקדם להתקדמות בתוך מערכת החינוך הפורמלי. וכך, בפירוש של ציוני מבחני משכל המושג "קטע מתחום המשכל" הוא אמנם קטע הניתן ליישום באופן רחב, ולו ביקוש רב, והוא עומד במקום המושג אינטליגנציה אנושית כללית ואוניברסלית.

### 3.3 שאלת הייחודיות של השפעות סביבתיות (situational specificity)

מושג הייחודיות של תנאי המדידה ונסיבותיה היה למוקד מחלוקת ביותר מתחום אחד של הפסיכולוגיה. הבלבול התעורר משום שנעשה שימוש במושג במובנים שונים, בהקשרים שונים. בחקר מבחנים קוגניטיביים הוא מציין את הקושי להכליל מסקנות של מחקר מעקב; במחקר אישיות על משתנים אפקטיביים, הוא מתייחס בראש ובראשונה לייחודיות של התנהגות אינדיבידואלית במצבים שונים.

באשר לסוג הראשון, יש לציין כי רק לעתים רחוקות, אם בכלל, משתמשים במבחנים בתנאים הזיהים לתנאים שבהם נאספו נתוני התקפות. לפיכך, דרושה מידה מסוימת של יכולת הכללה (generalizability) כשמדברים על שימוש מעשי במבחנים. כאשר מבחני כושר סטנדרטיים מתואמים (correlated) עם ביצוע בתפקידים דומים לכאורה - למשל, במחקרי תקפות בתעשייה - מתגלים מקדמי תקפות שונים למדי (Ghiselli 1966, 1959). ממצאים כאלה הפחיתו את הסיכויים לאפשרות ההכללה של תקפות מבחנים מעבר למצבים. עד אמצע שנות ה-70 נחשבה "ייחודיות סביבתית" למגבלה ניכרת בשימוש של מבחנים סטנדרטיים בברירת כוח-אדם.

לאחרונה הוכח באמצעות טכניקות סטטיסטיות חדשות, כי חלק ניכר מן ההבדלים בין מקדמי התקפות הוא בעצם ארטיפקט סטטיסטי, הנובע מגודל (קטן) של המדגם, או

ממהימנות נמוכה של הקריטריון (Schmidt & Hunter, 1977). הצטברות הראיות האמפיריות לפי שמיט והאנטר, מרמזת על כך שתקפותם של מבחנים ניתנת להכללה מעבר למקצועות באופן הרבה יותר רחב מכפי שנחשב קודם (Schmidt, Hunter, Pealman & Shane, 1979; Pealman, Schmidt & Hunter, 1980; Ghiselli, 1973) בהערכת תכונות אישיות; לעומת זאת, ייחודיות סביבתית ממלאת תפקיד משמעותי יותר (Mischel & Peake, 1982; Mischel, 1977, 1968; Anastasi, 1983). למשל, אדם עשוי להיות חברתי מאוד בעבודה, ואילו במפגשים חברתיים מחוץ לעבודה, יהיה ביישן ומאופק. גוש ניכר של ראיות אמפיריות מראה שקיימת ייחודיות סביבתית ניכרת בכמה מימדים לא-קוגניטיביים, כגון תוקפנות, הסתגלות לצורות חברתיות קבועות ותלות (Peterson, 1968; Mischel, 1968; Anastasi, 1983b).

#### 3.4 עיצוב תכונות (trait formation)

ניתוח גורמים כבר אינו נחשב לשיטת חיפוש אחר יחידות ההתנהגות הבסיסיות, הקבועות והאוניברסליות. אדרבה, הוא מוכר כשיטת ארגון של נתונים אמפיריים בקטגוריות מועילות. הגורמים, כמו ציוני המבחנים ונתונים תצפיתיים אחרים שמהם הם מופקים, הם בטבעם תיאוריים, לא מבארים. אין הם מייצגים ישויות סיבתיות יסודיות. כאשר מסתכלים לאור האמור לעיל בתכונות המזוהות בניתוח גורמים, אפשר לחקור את הסיבות לעיצוב תכונות ולשאל: אלו תנאים מביאים להתפתחות של עקביות התנהגותית או למבנה תכונות מיוחד. אין לערב את המונח "מבנה תכונות", כפי שהוא משמש כאן, עם פרופיל ביצוע, המראה את מקומו היחסי של הפרט בתכונות שונות; למשל, אם אדם מפותח יותר בצד הכושר המילולי או בצד הכושר הכמותי. ההיפך הוא הנכון - "מבנה תכונות" מתייחס לעיגון התנהגות בתכונות שלפיהן מתואר ביצועו של הפרט. מה גורם לדפוס הקורלציוני, אשר מלכתחילה מביא לזיהויין של תכונות מילוליות וכמותיות? הוצעו מספר מנגנונים העשויים להסביר את תהליך עיצוב התכונות (הפניות - ראה Anastasi, 1983). סוג אחד של הסבר מתמקד בקירבה או בהתרחשות בו-זמנית של חוויות לימודיות; סוג אחר הוא יישום חינוך למטרה אחרת (transfer of training), או לכל האפשרויות ללמוד על הכלל מהלימוד הספציפי. מחקרים השוואתיים על מבני תכונות באוכלוסיות בעלות היסטוריות התנסותיות שונות, מספקים כמה רמזים מבטיחים (Anastasi 1970, 1983a). אביא דוגמה אחת מתוך מחקר בנושא ההבדלים בין גילים. מחקרים מסוג ניתוח גורמים על אוכלוסיות בתי-ספר יסודיים ותיכונים, המתבצעים הן באמצעות שיטות השוואתיות, הן באמצעות מחקרים לאורך, הביאו קודם לניסוח השערת האבחנה המתפתחת (Garrett, 1946; Burt 1954) (differentiation hypothesis). לפי השערה זו, האינטליגנציה היא בלתי-מאובחנת יחסית בשנות הילדות הראשונות והולכת והופכת לנבדלת (specialized) במשך ילדות והתבגרות, עם התפתחותם של גורמים קבוצתיים ברורים בתחומים מילוליים, מספריים ומרחביים. התומכים בהנחת האבחנה המתפתחת, ייחסו ההבדלים בין הגילים בעיקר לבשילה (maturation). אולם בדיקה מעמיקה יותר של מחקרים, הכוללת כמה

ממצאים מן המחקרים המוקדמים, מרמזת על כך שהשינויים בדפוסי הגורמים שעליהם דווח קשורים יותר לחינוך מאשר לגיל (Khan, 1948, 1970, 1983b); Anastasi, 1970, 1972).

ממצאים כאלה מעוררים את שאלת תפקידו האפשרי של החינוך הפורמלי בעיצוב תכונות. ככל שהפרט מתקדם בבית-הספר, כך הוא נתקל בתכנית לימודים המחולקת יותר ויותר לתחומים מסורתיים (traditional subject-matter areas). למשל, הדרכה בתחומים מילוליים ומספריים נפרדת בהדרגה לשעות לימוד שונות, ולבסוף מועברת על ידי מורים שונים. הפרדה דומה נעשית בהקשר לתחומי הדרכה שבהם כרוכים כשרים שהם בעיקר תפיסתיים או מרחביים, כגון אמנות, מכניקה וקורסים "מעשיים" אחרים. ההבחנה בין חוויות לימודיות מלווה בהענקת משקל הולך וגדל לגורמים קבוצתיים צרים יותר. על כן, קיימות ראיות לכך שסוגים שונים של תכניות לימודים, כגון תכניות אקדמיות וטכניות, מיוחסים להתפתחות גורמים קבוצתיים השונים זה מזה מבחינה איכותית (Filella, 1965; Dockrell, 1980).

במחקרי קבוצות המייצגות קונטקסטים תעסוקתיים, חברתיים-כלכליים ותרבותיים שונים, נמצאו ראיות לכך שהיסטוריה התנסותית תורמת לעיצוב תכניות תכונות שונות.

#### 4. סוף דבר

בהכנת מאמר זה בחרתי לדון בהתפתחויות מרכזיות אחדות שחלו במדידה הפסיכולוגית במישים השנים האחרונות. לא הבטחתי להציג נושא אחיד, אולם כאשר אני קוראת את הדברים, אני מגלה בכל זאת מכנה משותף. בקצרה אפשר להגדירו כהכרה הולכת וגוברת בצורך להתחשב בהקשר הסביבתי של התנהגות בכל מדידה פסיכולוגית. אוריינטציה כללית זו מרמזת על כיוונים מסוימים למחקר עתידי, מצביעה על בניית מבחנים ועל שימוש במבחנים תוך התייחסות הולכת וגדלה להקשרים שבהם ינוצלו ודורשת ניתוחי משימות של התנהגות במגוון מצבים ובשלבם שונים בחיים. יש להקדיש מאמצים לחיפוש אחר מידע על רקעם ההתנסותי וההיסטוריה הלימודית של נבחנים אינדיבידואליים, בנוסף על הרקע ההתנסותי וההיסטוריה הלימודית של קבוצות המדגמים הנורמטיביים, המשמשים בסיס לתקנון המבחנים. מידע כזה עשוי לתרום לפירוש היעיל של ציוני מבחנים למטרות אין-ספור, החל בברירת מועמדים לתעסוקה ושיבוצם, דרך קבלת מועמדים לתכניות השכלה, ייעוץ פרטני והערכת קשיי לימוד וכלה בזיהוי ילדים מפגרים ומחוננים. כל זה עשוי לסייע לנו להבין את הגורמים המביאים להתנהגויות שונות של אנשים ואת התארגנות ההתנהגויות לקטגוריות תכונות, הניתנות לזיהוי אמפירי.

## השלמות

בשני מאמרי סקירה נוספים מעלה אנסטסי עוד שני נושאים מרכזיים הקשורים במגמות התפתחות בתחום המבחנים והמדידה.

הקטעים להלן הם עיבוד חופשי שעשה המחבר מפיסקאות עיקריות במאמרים הללו. האחריות חלה על המחבר בלבד.

### 5. גישות חדשות לתיקוף מבחנים<sup>(4)</sup>

בתקופת ה"ילדות" של המבחנים המתוקננים, נמדדה התקפות בטכניקות שונות ומגוונות ונקראה בשמות שונים ומגוונים לא-פחות. בתחילת המאה אנו פוגשים במחקרי תקפות שהפעילו קריטריונים כגון דיפרנציאציה בין-גילית, הערכות מורים, חומר אנמנטי, איבחונים פסיכיאטריים והערכות עצמיות ושדיווחו על הממצאים בצורה שהיתה לא פעם עמומה ומבלבלת. המונחים שהופיעו אז בספרות היו מבלבלים לא פחות:  
face validity, validity by definition, intrinsic validity, logical validity, factorial validity.

ב-1954 נקצר פרי המאמץ של שנים אחדות של דיונים בוועדות מכינות. התפרסם Technical Recommendations for Psychological Test and Diagnostic Techniques. מסמך זה, שמהדורות מעודכנות שלו יצאו ב-1974 וב-1985, הכניס סדר בעולם המונחים וקבע סטנדרטים לבניית מבחנים ושימוש בהם.

ב- Technical Standards (שמו המקוצר של המסמך הנ"ל) של 1954, נקבעו פורמלית הקטגוריות של תקפות תוכן, תקפות מקבילה ותקפות חיזוי (שאוחדו ב-1974 בשם תקפות אמפירית) ותקפות מבנה.

בעוד שמשנות ה-50 עד אמצע שנות ה-70 היתה האוריינטציה אמפירית בעיקרה, הרי בעשור האחרון אנו רואים בפסיכולוגיה האמריקנית בכלל, ובתחום המבחנים בפרט, עניין גובר והולך בתיאוריה. תקפות המיבנה קבלה אפוא משנה חשיבות ביחס לתקפות האמפירית. "מבנה", במינוח הפסיכומטרי, הינו מושג תיאורטי הקרוב במשמעותו ל"תכונה". מבנים יכולים להיות פשוטים ומוגדרים באופן צר, כמו, למשל, מהירות הילוך או רמת הדיוק בכתיבה; הם גם יכולים להיות מורכבים וכלליים, כמו, למשל, היגיון מתמטי, כושר סכולסטי, נזירותיות וחרדה.

<sup>(4)</sup> מתוך - Annual Review of Psychology, 1986, 37, 1-15. Evolving Concepts of Test Validation.



לאחרונה הולכת והודרת ההכרה, כי פיתוח של מבחן תקף מחייב נקיטה של כמה הליכים המופעלים בזה אחר זה בשלבים שונים של בניית המבחן. כך, נבנית התקפות אל תוך המבחן מתחילתו, וזאת בניגוד לגישה המסורתית, שנקטה בדיקת תקפות אמפירית מול קריטריון בשלבים האחרונים של הפיתוח. הדגשת-היתר של הליכים אמפיריים בעשורים המוקדמים, נבעה בחלקה מהתנגדות בסיסית ל"פסיכולוגיה של הכורסה", אשר שימשה בסיס לכתיבה הפסיכולוגית בסוף המאה הקודמת ובתחילת המאה שלנו. כעת אנו מבינים, שאמפיריציזם אסור שיהיה עיוור; לעומת זאת, תיאוריה אסור שתהיה ספקולטיבית וסובייקטיבית. במלים אחרות, תיאוריה יכולה לנבוע מניתוח של ממצאי מחקר, ויכולה גם להוביל לניסוח של השערות שניתן לבדקן הלכה למעשה. דוגמה לאינטגרציה נכונה של הגישה האינדוקטיבית והדדוקטיבית אפשר למצוא בשיטות השמת פריטים לתת-מבחנים, המבוססות הן על נתונים סטטיסטיים, הן על אפיוני תוכן. פריט נכלל במבחן/בשאלון אם הוא (1) נכתב מראש על פי ספציפיקציות הלקוחות מההגדרה המבנית של המבחן (תיאוריה) וכן - (2) על פי נתונים סטטיסטיים, הוא שייך אמנם לסולם (או התת-מבחן) הספציפי.

#### 6. תיאורית ה-IRT והמבחנים הממוחשבים<sup>(5)</sup>

טכניקות סטטיסטיות של ניתוחי פריטים מילאו תפקיד חשוב בבניית מבחנים מתוקננים למן ההתחלה. הן עסקו במדידה של דרגת הקושי של הפריט ובמידת הערך המאבחן שלו (discriminative value). שני מאפיינים אלה נמדדו תמיד כך שהתוצאה היתה תלויה במדגם המסוים ששימש להעברה הניסויית, והחלה על אוכלוסייה רחבה יותר התאפשרה רק אם המדגם ייצג באופן מלא את האוכלוסייה. כאשר נדרש מעבר מנוסח ממבחן למבחן ומאוכלוסייה לאוכלוסייה, נעשה בדרך כלל שימוש בפריטי עוגן ובנוסחאות כיוול, ובכל מקרה נדרשו העברות נוספות של מבחנים. הופעתם של המחשבים בזירה הפסיכומטרית הביאה לשימוש הולך וגובר של הליכים סטטיסטיים-מתמטיים מדויקים, שנועדו לתשב פרמטרים של פריטים באופן בלתי-תלוי במדגם הספציפי. הליכים אלה נקראו במקור - Latent Trait Models. משמעותה של trait אינה אלא מבנה סטטיסטי, הנגזר באופן אמפירי מתוך מערכת קשרים בין פריטים. כדי למנוע את הרושם המטעה הנוצר על ידי הביטוי "תכונה חבויה" (latent trait), החליפו כמה מחמובילים בתחום את המונח הביטוי הפחות יומרני (IRT) Item Response Theory, המקובל כעת בספרות הפסיכולוגית.

(5) מתוך -

Mental Measurement: Some Emerging Trends, Introduction to the 9th Mental Measurement Yearbook, Gryphon Press (in press).

ההליכים הללו משתמשים בשלושה פרמטרים של הפריט: הערך המאבחן של הפריט, דרגת הקושי שלו והאסימפטוטה התחתונה, שהיא פרמטר המבוסס על הסתברות הניחוש. כמה הליכים פשוטים יותר של IRT, כמו המודל של Rasch משתמשים רק בפרמטר אחד - דרגת הקושי - תוך הנחה שהשפעת השניים האחרים היא זניחה. אגב, הנחה זו ניתנת לבדיקה אמפירית ויש לבדוק אותה הלכה למעשה.

השימוש ב-IRT מאפשר לתאר התנהגותו של פריט ולשער את הפרמטרים שלו גם במדגמים אחרים מאלה בהם נוסה, וזאת ללא העברה נוספת (מכאן המונח sample-free).

ה-IRT כובש לו בהדרגה מקום חשוב בפרויקטים של מבחנים המועברים בקנה-מידה גדול (ה-Scholastic Aptitude Test למשל, המועבר לכמיליון נבחנים בשנה).

אחד היישומים החשובים של IRT מתרחש בשטח המבחנים הממוחשבים המסוגלים (Computer Administered Adaptive Testing).

הרעיון המרכזי של השיטה הוא, שהנבחן עצמו מותאם את מהלך המבחן באמצעות תגובותיו, והפריט הבא בתור נבחר על ידי המחשב (ומוצג לנבחן) על פי תגובות הנבחן לפריטים הקודמים. בעיקרו של דבר, הליך זה אינו שונה ממה שעושה הבוחן המעביר מבחן אינטליגנציה אינדיבידואלי כמו סטנפורד-בינה.

מבחנים מסוגלים הם חסכוניים יותר ומניבים ציונים מדויקים יותר ממבחנים אינדיבידואליים וממבחנים קבוצתיים, שבהם מוצגות לכל נבחן כל השאלות.

קידום מבחנים ממוחשבים מסוגלים מחייב הכנות בכמה תחומים: פיתוח סטטיסטיקות מתקדמות (באמצעות מודלים של IRT), תוכנה משוכללת, חומרה נוחה למשתמש, סיגול פסיכולוגי של הנבחנים ומחקר משווה, שיאפשר שימוש בידע אשר הצטבר על מבחני נייר ועיפרון, כאשר יתבצע המעבר.

בכמה ארגונים המשתמשים במבחנים בקנה-מידה רחב, הן צבאיים, הן אזרחיים, נעשות כעת הכנות למעבר הצפוי. אנשי מקצוע רבים רואים במיחשוב המבחנים את השינוי החשוב ביותר בפסיכומטריקה מאז תחילת השימוש במבחנים מתוקננים.

המבחנים הפסיכולוגיים תחת שבט הביקורת

התרחבות השימוש במבחנים פסיכולוגיים בארצות-הברית נמשכה כמעט ללא הפרעה במשך יובל שנים, למן שנות העשרה. בשלב מסוים באמצע שנות ה-60, החלו מטחי הביקורת לפחות. בתחילה טיפין טיפין ואחר-כך בקילוחים. למעשה הוטרה הביקורת חמש-עשרה שנים רצופות, ורק בתחילת שנות ה-80, כלומר, לפני שנים ספורות בלבד, חדלו מכך. הסקירה של אנטסטי מתייחסת רק לחלק מן הנושאים השנויים במחלוקת שהעסיקו את הפסיכולוגים בקשר למבחנים. אנו ננסה לסקור פה את סוגי הטענות המרכזיות שהופנו נגד המבחנים בארצות-הברית; בצמוד לכל סוג נביא את התשובות או הפתרונות שהוצעו. נעסוק בנושאים האלה :

- \* הגנה על זכויות הפרט
- \* הטיה תרבותית במבחנים
- \* היבחות חוזרת והכנה למבחנים
- \* המבחנים כסותרים את המעשה החינוכי.

מעניין לציין, כי מרבית הנושאים הבעייתיים ומרבית התביעות הועלו קודם כל על ידי פסיכולוגים. במלים אחרות, המבקרים החריפים ביותר והרפורמטורים התקיפים ביותר בתחום המבחנים היו פסיכולוגים.

די לפתוח באופן אקראי כתבי עת כמו - American Psychologist; Journal of Educational Measurement; Journal of Applied Psychology; Psychological Bulletin

מאותה תקופה, כדי לגלות במה דברים אמורים. על כל פנים, הבה ניגש כעת לפירוט.

א. הגנה על זכויות הפרט  
מדובר כאן בקבוצה של טענות. נשתדל למנות את העיקריות ונביא אותן פחות או יותר בלשון המבקרים.

1. פתיחות במבחנים (openess in testing)  
אין הצדקה להתייחס למבחנים כאל "מכשיר רנטגן"; אין הצדקה להתייחס לציונים במבחנים כאל מושג מסתורי ואין להתייחס לפסיכולוגים כאל בעלי "כח עליון".  
באופן קונקרטי התביעות הן:  
- לפני הבחינה: הנבחנים זכאים להכנה מוקדמת, למידע לגבי מה שהם עומדים לעבור.  
- לאחר הבחינה: הנבחנים (או האפוטרופסים שלהם) זכאים לדעת את הציונים וההערכות שקיבלו על סמך המבחנים וכיצד הם "בושלו". לשם כך יש לתת פרסום לתוכן הבחינה (test disclosure).

2. חדירה לתחום הפרט (invasion of privacy)  
- יש להשיג הסכמה מפורשת של הנבחן (או האפוטרופוס שלו) בטרם ייבחן.  
- מידע אישי הנובע מן המבחנים חייב בכללי סודיות מחמירים.  
- מאגרי מידע ממוחשבים צריכים לעמוד תחת פיקוח ציבורי, על מנת למנוע שימוש בלתי-אתי.

כיצד נערך המימסד הפסיכולוגי לפתרון בעיות אלו?  
ראשית, הטענות הללו הוכרו כצודקות, כלומר, הכל הסכימו שיש לנסות בכנות לפתור את הקשיים. אלא שלא תמיד עלו המאמצים יפה.  
(Bersoff, 1981; Gordon & Molidown, 1981; Novick, 1981).  
לדוגמה, בשתי מדינות בארה"ב (ניו-יורק וקליפורניה) נתקבלו בשנות ה-70 חוקים המחייבים חשיפה של תוכני המבחנים (test disclosure).  
ETS, שהוא המכון הבונה, המעביר והמציין את ה-SAT, נאלץ על פי חוקים אלה למסור לעיונו של כל נבחן הרוצה בכך את המבחן ואת מפתח התשובות. מאחר וחשיפה כזו "שורפת" את המבחן ומונעת שימוש חוזר בפריטים, ניצבה הנהלת ETS בפני דילמה: להתעלם מן החוק ולנקוט סחבת משפטית, או, לחילופין, להזדקק להוצאה תקציבית גדולה בשל הצורך לבנות פריטים חדשים לעתים קרובות מאד. (הוצאה כספית זו עתידה היתה ליפול על כתפי הנבחנים, היות ומרבית הכנסות הארגון מתקבלות מדמי הבחינה).

כמוצא מן הסבך ננקטה דרך ביניים. ETS צמצם את מספר מועדי הבחינה בניו-יורק ובקליפורניה משישה לשניים בשנה; בו זמנית, החליט ETS לפרסם בעיתונים את הבחינה ותשובותיה (לחסוך בכך את הצורך לטפל בפניות של נבחנים). בעקבות מהלכים אלה ארעו מספר דברים. האחד - לצעירי ניו-יורק וקליפורניה התברר, כי מצבם ביחס לעמיתיהם בשאר מדינות ארה"ב הוחמר מבחינת מספר מועדי הבחינה העומדים לרשותם. השני - במידה מסוימת נפגעה רמת האמינות של ציוני הנבחנים בניו-יורק ובקליפורניה, כתוצאה מכך שבכל מועד הועבר נוסח בחינה אחד בלבד (בדרך-כלל מועברים בכל מועד 15-25 נוסחי בחינה, כך שלמעשה אי-אפשר להעתיק). צמצום אמינות זו נתפס באופן שלילי, במיוחד על ידי האוניברסיטאות. אפשר שהדבר אירע שמסובות אלו ואפשר שמסובות אחרות; מכל מקום, תהליך החקיקה בנושא חשיפת המבחנים נעצר ולא הצטרפו אליו מדינות נוספות.

דוגמה נוספת לקונפליקט בין האידיאלי לאפשרי ניתן למצוא בנושא קרוב לתחום המבחנים - הערכות והמלצות אישיות. בעקבות לחץ של התומכים בזכויות הפרט, נקבע בשנות ה-70 בארגונים רבים הכלל האומר כי כל אדם המביא עמו או המבקש שישלחו למענו המלצה לצורך קבלה לעבודה או ללימודים, רשאי לפנות למעסיק הפוטנציאלי ולדרוש לראות את ההמלצה. כך נקבע גם לגבי הערכות עובדים על ידי ממונים במפעלים אחדים. כתוצאה מכך חל פחות מיידי בערכן של ההמלצות וההערכות והן חדלו כמעט לחלוטין להשפיע על ההחלטות. עד מהרה מצאה המערכת את הפתרון המעשי: על טופסי ההמלצה החל להופיע משפט סטנדרטי מעין זה: "אני הח"מ מותר/לא מותר (מתק את המיותר) על זכותי לעיין בהמלצה זו". מבקשי ההמלצות הבינו את הרמז, וכיום הבעיה אינה קיימת עוד, היות וכולם כאחד מותקים את המיותר... נראה, אם כן, שבתחום זכויות הפרט חלה התקדמות ברורה, אם כי המציאות היומיומית מקשה לא פעם על יישום הפתרונות האידיאליים.

### ב. הטיה תרבותית

הטענה בדבר הטיה תרבותית במבחנים פסיכולוגיים בארה"ב משתלשלת בערך כך: המבחנים הפסיכולוגיים נבנים ומעוצבים בידי אנשים אשר ברובם התחנכו והתפתחו בתוך תרבות הלבנים בני המעמד הבינוני המשכיל; המבחנים משקפים את עולמם התרבותי-השכלתי של אנשים אלה. נבחנים מקבוצות אתניות שונות (שחורים, היספנים), ומעמדות נמוכים ממין אחר (נשים) נפגשים בעת היבחנם באלמנטים קוגניטיביים זרים למנטליות שלהם. חוסר היכרות עמוק זה עם מטלות המבחן ותכניו גורם לביצוע לקוי ולהישגים נמוכים. כך, במקום שהמבחנים ישקפו את היכולת הטבעית של נבחנים אלה, הם מציגים תמונת מצב כפי שהיא נתפסת ונמדדת בידי המעמד הבינוני הלבן, תמונה שהיא מטעה ומקפחת. מצב זה גורם למיסוד הפערים החברתיים הקיימים ולהנצחת הקיפוח. עד כאן לשון הביקורת.

אנסטסי כותבת על "שוני תרבותי" ועל הצורך ב"הגדרות נכונות של מושג הקיפוח" ודבריה נכתבים ונקראים בנחת. אך נושא זה הפך בשנות ה-60 "חס" מאד ובעל טעינות אמוציונלית, והוא הביא כמה וכמה כינוסים מדעיים אל סף פיצוץ.

בשנות ה-60 החלו הטענות בדבר הטיה במבחנים ללבוש צורה ממשית: התנהל מסע ציבורי ומשפטי להפסקת ברירת סטודנטים וברירת עובדים בתעשייה על סמך מבחנים פסיכולוגיים, בטענה שהמבחנים מקטינים את סיכוייהם של שחורים (שחורים בעיקר, אך גם היספנים, נשים ומהגרים) להתקבל ללימודים ולעבודה. המבחנים היוו מטרה להתקפות חריפות.

הבעיה לא היתה מתעוררת, כמובן, אלמלא נמצאו שוב ושוב פערים בממוצעי ציוני המבחנים של אוכלוסיות. הפער בין לבנים ושחורים, למשל, נע באופן עקבי למדי בין חצי סטיית תקן לסטיית תקן שלמה לרעת השחורים. הפער ב-SAT-M (מבחן חשיבה מתמטית למועמדי האוניברסיטאות) בין גברים לנשים נשאר יציב מאז שנות ה-50 ועומד על כשליש סטיית תקן.

אין תימה שהנסיונות הראשונים שעשו פסיכומטריקאים להתמודד עם הבעיה (זמן רב לפני פרוץ הסערה הציבורית) היו בכיוון של חיפוש אחר מבחנים שיאפסו את ההבדלים בין הקבוצות. מרבית הניסיונות הללו כשלו, או שההבדלים התרבותיים שבו ונתגלו גם במבחנים ה"הוגנים", או שהמבחנים הללו נמצאו חסרי תקפות כלפי קריטריונים רלוונטיים (נבו, 1976).

חרף הכישלון, הותירה התביעה הנאיבית את חותמה על המסד הטסטולוגי, עד עצם היום הזה: במרבית המכונים הגדולים למבחנים יש שאיפה לאיזון אתני ובין-מיני בגלל הפסיכומטרי העוסק בחיבור הפריטים וכן מתקיימות פעולות בקרה קבועות לסילוק פריטים בעלי תוכן תרבותי גלוי מובהק מדי.

מאמצע שנות ה-60 החלו להופיע בספרות הפסיכומטרית מודלים אשר ניסו להגדיר את תופעת ההטיה באמצעות כלים סטטיסטיים. הדגש בהם הועבר מן הפערים במבחנים לפערים בקריטריון. המשותף למודלים אלה היא הקביעה שהוגנותו של מבחן אינה יכולה להיות נושא לדיון עצמאי, אלא יש לבדוק אותה בהקשר לקריטריון כלשהו. מבחן ייחשב להוגן אם ורק אם הפער בממוצעי ציוני החזאי (המבחן) ייעלם כאשר יחושבו ציוני הנבחנים על הקריטריון. במיוחד, ייחשב מבחן כמוטה אם לנבחנים מקבוצה מסוימת יתרחש חיזוי - חסר (underprediction), כלומר, ביצועיהם האקטואליים בקריטריון יהיו גבוהים מכפי שנחזה על פי ציוניהם במבחן.

במינוח טכני, מבחן (x) יחשב כמכשיר מיון מוטה כלפי קריטריון (y) מסוים אם בהפרדת הנבחנים לשתי קבוצות (על בסיס אתני, למשל) ולאחר עריכת חישובים

סטטיסטיים נפרדים עבור כל אחת מהן, יתברר כי קיים ביניהן שוני מובהק באחד משלושת הפרמטרים הבאים או לפחות באחד מהם: שיפוע קו הרגרסיה, האינטרספט, טעות התקן של האומדן. (כתב העת Journal of Educational Measurement הקדיש למודלים אלה חוברת שלמה בשנת 1973).

במשך שנות ה-70 נערכו עשרות מחקרים שעיקרם השוואת קווי הרגרסיה של לבנים ושחורים, גברים ונשים, תושבי עיר לעומת תושבי כפר. במאמץ מונומנטלי כינס ג'נסן את כל המחקרים הללו בכרך עב כרס שנקרא - (Jensen, Bias in Mental Testing, 1980).

קשה לסכם את הספר בקיצור. אף על פי כן אפשר אולי להביא כאן משפטי מפתח אחדים המופיעים בו: "מחקרים אמפיריים של הטיה במבחנים כנגד קבוצות מיעוט (בדרך כלל שחורים) התבססו בעיקר על השוואת מקדמי התקפיות ועל השוואת שני קווי הרגרסיה... נערכו מחקרים רבים מאד בתחומים ישומיים מגוונים... העדות הכוללת ביחס לתקפיות חיזוי בקרב לבנים ושחורים היא כי התקפיות הן למעשה זהות עבור קבוצות אתניות אלו וכנראה גם עבור קבוצות אחרות... גם כאשר משווים קווי רגרסיה, העדויות הן ברורות... במקרים הספורים שנמצאו הבדלים בין לבנים ושחורים, התבטא הדבר בכך שלשחורים היה אינטרספט נמוך יותר... תופעה כזו מביאה לחיזוי-יתר עבור השחורים, כלומר, להעדפתם (לא לקיפוחם) אם הברירה המעשית מתבצעת על פי קו הרגרסיה המשותף... אפשר לקבוע בביטחון כי מרבית מבחני הכושר המצויים בשימוש במערכות החינוך, בצבא ובתעסוקה, אינם מטים את תהליך הברירה" (Jensen, 1980:515).

באמצע שנות ה-80' אנו רואים מגמה ברורה של רגיעה בנושא הטיה התרבותית. התפיסה האומרת כי המבחנים הם אך ורק כלי מדידה המשקף מציאות מסוימת, שנועד לחזות תפקוד בתוך מסגרת מציאותית נתונה ואין בכוחו לתקן עוולות חינוכיות או לשנות מסגרות - תפיסה זו הולכת ונעשית נחלתם של רבים. במקביל, מופנית המתאה החברתית כלפי המטרות ה**נכונות**, קרי: תנאים חברתיים, שיטות חינוך וסדר קדימויות פוליטי-ציבורי, ולא כלפי מכשירי המדידה.

אף כי החוקרים מצאו די עדויות התומכות בהשערת ההוגנות, אי-אפשר לומר שהסערה המקצועית-ציבורית לא הטביעה את חותמה. המודעות לנושא גרמה לכך שכמעט כל מבחן חדש עובר בדיקה השוואתית בין-תרבותית בטרם יפורסם המדריך שלו; יש הקפדה רבה על התוכן הגלוי של פריטים - שמא חלילה יש בהם רמזים תרבותיים ייחודיים מדי; בדיקת הטיה הוקדמה לשלב הטרומי של פיתוח המבחנים. לתוך ניתוח הפריטים הקלאסי נוספו מדדים של הטיה תרבותית (ברמת הפריט), המאפשרים לזהות פוטנציאל של הטיה ולהתגבר עליו באמצעות פסילת הפריט, עוד בשלב בניית המבחן, כאשר נוסחים ניסיוניים שלו מועברים למדגמי חלוץ.

## היבחנות חוזרת ואימון למבחן

מאז ומתמיד שימש נושא ההכנה למבחנים אתגר לפסיכומטריקאים. הטענה, שאפשר לשפר בצורה ניכרת את ההישגים במבחני כושר בעזרת חזרה והכנה, מעמידה שתי קושיות עיקריות:

האם כל מה שהמבחנים הללו מודדים הוא שליטה במקבץ של טכניקות צרות וייחודיות? ואם כך הדבר, האם מותר להניח קשר אמיתי כלשהו בין המבחנים הללו למשכל שהם מתכוונים למדוד ולתפקוד הקוגניטיבי שהם אמורים לשקף?

האם קיים שיוון בין הנבחנים ביחס לאמצעי ההכנה העומדים לרשותם? האם ייתכן שנבחנים ממעמד חברתי-כלכלי גבוה יותר מסוגלים לגייס אמצעי הכנה טובים יותר, להיבחן מספר רב יותר של פעמים, ולפיכך, גם להשיג ציונים גבוהים יותר במבחנים?

הגוף אשר נדרש יותר מאחרים לנושאים אלה היה Educational Testing Service (ETS), האחראי להעברת המבחן הנפוץ ביותר בעולם - Scholastic Aptitude Test (SAT). ה-ETS יזם את מרבית המחקרים בתחום. בעזרת מחקרים אלה (ורבים אחרים) מתחילה התמונה להתבהר. במסגרת המצומצמת שלנו נוכל רק לתמצת כמה מן המחקרים האחרונים.

הנתון המצוטט ביותר בספרות בהקשר לאפקט של בחינה חוזרת הוא שיפור של 20-25 נקודות (על סולם של 100;500), המדווח על ידי Angoff (1971) כאופייני להפרש ממוצעים של ציוני SAT עבוד מדגמי נבחנים חוזרים. אמנם, שיפור זה מרשים למדי, אך יש לזכור שהוא ניזון בין היתר מהשפעה של תופעת הרגרסיה אל הממוצע ומהשפעה של טיב אוכלוסיית הנבחנים אשר ניגשים לבחינה חוזרת (בדרך כלל אלה הם אנשים אשר מעריכים נכונה כי ציונם בבחינה הראשונה נמוך מכפי יכולתם האמיתית). כאשר "מנכים" את שני האפקטים הללו, נותר פער של 15-17 נקודות בלבד (Alderman, 1981). פער זה יורד במהירות בבחינה החוזרת השלישית וְהרביעית, עד כדי איפוס מוחלט.

באשר לשיטות הכנה שונות - החל מעיון בחוברת הסבר, דרך התעמקות בספרי תרגול ועד להשתתפות בקורסי הכנה אינטנסיביים - נמצא כי האפקטים הממוצעים הם הלשים ועוצמתם נעה בין אפס לחמישית סטיית תקן (Powers, 1980; Linn, 1982).



1984, Stricker; Powers & Alderman, 1983). סיכום מספר רב של מחקרים שנעשה על ידי Jungeblut & Messik (1981) בהקשר למספר שעות האימון הממוצע לשיפור ציוני SAT-V ו-SAT-M מראה כי המאמץ הנחוץ אפילו עבור שינוי חיובי קל, הוא רב ביותר. יתר על כן, באמצעות סקר צרכנים שערך ETS התברר כי השתתפות בקורסי הכנה לבחינת ה-SAT איננה קוסמת במיוחד לנבחנים ורק כ-5% מן הנבחנים טורחים ומצטרפים לקורסים כאלה.

אנו רואים, אם כן, שהמחקרים בכללותם מצביעים על כך שהנושא אינו בעייתי במיוחד, היות והאפקטים שמדובר בהם אינם חזקים מאוד. לפיכך, אין רואים כיום צורך להילחם בתופעת ההיבחנות החוזרת, או בתופעת ההתכוננות לבחינות. עדיף למסור מידע רלוונטי לנבחנים ולהשאיר בידיהם את ההחלטה. בחוברת החסבר של ETS מ-1982, נמסרים נתונים מדויקים על הסבירות של שיפור הציון בבחינה חוזרת. לדוגמה, נמסר שם כי כ-65% מהנבחנים החוזרים מעלים את ציונם. 5% אף מעלים אותו בשיעור של סטיית תקן שלמה. אך יש גם "נשחקים", אשר ציונם ירד בפעם השנייה (כ-35% מכלל הנבחנים החוזרים) ולעתים הירידה היא חזקה עד כדי סטיית תקן אחת (כ-1% מהנבחנים החוזרים). בדברי החסבר שבחוברת מושארת ההחלטה בדבר כדאיות ההתכוננות וכדאיות הבחינה החוזרת בידי הנבחן.

בנוסף על קידום המחקר ושינוי מדיניות החסברה נעשו שם פעולות מעשיות. ב-1982 הוציא ETS חוברת הסבר ובה מבחן SAT אמיתי שלם מלווה בחסברים, מפתח תשובות ונורמות. ארגונים אחרים מתכוונים ללכת בעקבותיו. הגישה העניינית הזו, של מתן מידע המבוסס על מחקרים ושל נקיטת אמצעים קונקרטיים, תרמה להרגעת הרוחות.

#### ד. המבחנים כסותרים את המעשה החינוכי

מבחני כושר סכולסטי ומבחני הישג סטנדרטיים במערכת החינוך בארצות-הברית, הלכו והיו לדומיננטיים במרוצת השנים. הידועים ביותר בהם הם אלה המשמשים לברירת סטודנטים: American College (SAT) Scholastic Aptitude Test ו- (ACT) Testing לצידם קיימים עשרות מבחנים נפוצים (כקבוצה) לא פחות. אלה הם מבחני הישג תקינים (Standardized Achievement Tests) שנבנים בכל מדינה, (state) בדומה למתכונת בחינות הבגרות אצלנו. מבחנים אלה מתרבים והולכים בעקבות הדרישה למדידה אובייקטיבית של אפקטיביות בהוראה (accountability) in education) ורצונם של הדרגים הניהוליים לקבל מידע כמותי מקיף על השגת היעדים החינוכיים במוסדות שבאחריותם. אחד הביטויים לחשיבותם של המבחנים בארצות הברית הוא יחסם של מועמדים למשרות פוליטיות לתוצאות המבחנים. ביטוי אחר הוא הדאגה שהתעוררה במערכת החינוך שם נוכח הירידה המתמדת בציונים הממוצעים של ה-SAT על פני ארבעה עשר שנתונים.

טענות קשות נגד המבחנים הסטנדרטיים באות בעיקר מתוך מערכת החינוך עצמה. (ראה חוברת Phi-Delta-Kappa מאי 1985, אשר הוקדשה כולה לנושא). כמו כן, ראה סיכום ראיונות בנושא זה עם אנשים מרכזיים בתחום (The New York Times, 11 במרס 1986).

להלן אחדות מהן:

- \* התמקדות במבחנים מתוקננים מובילה לכך שהמורים מפסיקים ללמד, התלמידים מפסיקים ללמוד, וכולם מעוניינים אך ורק ב"תכלית", קרי, להצליח בבחינות המתוקננות המסוימות העומדות על הפרק. נושאים שאינם כלולים בתכנית המבחן מושמטים ומוזנחים.
- \* כדי "לצאת בכבוד" ממבחן כלל-ארצי יכולים מנהלי בית-ספר למנוע מלכתחילה מתלמידים הלשים ללמוד במוסדות שהם מנהלים (שכן על ידי קיצוץ ה"זנב" ניתן להעלות את הממוצע). הנפגעים העיקריים מכך הם יוצאי שכבות חלשות ובני מיעוטים.
- \* ככל שהמבחן כללי יותר ומשמש למטרות רחבות יותר וככל שהוא מיועד לאוכלוסיות נבחנים גדולות יותר, כך התכנים שלו קשורים פחות לתכנית הלימודים של כיתה מסוימת הלומדת אצל מורה מסוים.
- \* עליית משקלם של הציונים במבחנים הסטנדרטיים, בקביעת עתידו של התלמיד, באה על תשבון ירידת משקלן של הערכות המורה. אין סומכים עוד על המורה

כמי שזכאי וחיבב להשפיע על התפתחות תלמידו. לתהליך זה השפעות שליליות  
הן על תפיסת התלמיד את מורהו, הן על הדימוי העצמי של המורה.

\* כדי שהמבחנים המתוקננים יהיו יעילים, עליהם להתבסס בעיקר על שאלות  
מטיפוס בחירה (multiple choice). תפקודים קוגניטיביים חשובים אינם  
יכולים להיבדק במסגרת כזאת וממילא יזכו לתשומת-לב פחותה מדי ולמידה  
מצומצמת מדי של זמן-הוראה; הבעה בכתב, שכנוע, פירוק בעיה למרכיביה,  
הקפת נושא מכל צדדיו - כל אלה יוזנחו לחלוטין, אם המשימה היא להכין את  
התלמיד למבחן מטיפוס בחירה.

\* תוצאות המבחנים נושאים לא פעם אופי דטרמיניסטי. ה"תג" שמטביע הציון על  
הנבחן עלול להיתפס כמשהו מולד וסופי.

האמת ניתנת להיאמר, כי המבחנים שמדובר בהם בהקשר לביקורות הללו, הם בעיקר  
מבחני הישג; כמעט ואין מדובר כאן במבחנים פסיכולוגיים (פרט לנקודה  
האחרונה); לאמיתו של דבר השיקול של איכות המבחנים הוא הרבה פחות חשוב, שכן  
אם השימוש בהם לקוי, הרי שגם המבחן המושלם לא ימנע את הכישלון החינוכי.  
יתכן מאוד שמקור הביקורת החינוכית הוא בהפרזה ובדומיננטיות - היתר של  
המבחנים  
האחידים. כאשר יושג האיזון הנכון ויפותחו כלים חלופיים, תיפחת גם הביקורת.

האמריקנים לא ראו צורך להגן על העמדות התומכות במבחנים בתחום החינוכי. הם  
הותירו את ההתמודדות עם הנושא לאנשי החינוך, והללו אמנם עשו זאת. על כל  
פנים, כיום מקובל לחשוב כי הבעיות שצויינו לעיל אינן נוגעות למבחנים  
המתוקננים, אלא לדרך השימוש בהם. יש לכוון את המשתמשים להכיר שיטות הפעלה  
נכונות, להגביר את מודעותם למגבלות המבחנים ומגבלות הציונים.

#### פרק שלישי:

#### המצב בישראל

בין ישראל לארצות-הברית קיימים הבדלים בולטים בתחום המבחנים הפסיכולוגיים;  
כמה מהם מתרכזים ברמה הטכנית-מקצועית. רבים מן המבחנים הנמצאים בשימוש נרחב  
אצלנו, לוקים לפחות באחד החסרונות הבאים:

מתרכזים ברמה הטכנית-מקצועית.

מהימנות: אין די נתונים לגבי סוגים שונים של מהימנות, ומכאן שאין די מידע על גודל טעות המדידה.

נורמות: במידה וקיימות נורמות, אין הן מבוססות על מדגמים רחבים ומייצגים.

תקפות: אין די נתונים על התקפות לסוגיה.

תיקנון לקוי: אין הוראות העברה אחידות ומחייבות והנחיות הציינון אינן מפורטות וחד-משמעיות דיין.

פגמים "אסטתיים" וארגוניים: שגיאות דפוס במבחנים; הנייר מאיכות נמוכה והחומר המודפס עליו אינו קריא; סידורי איחסון מבחנים הינם לקויים ותוכנם של מבחנים מודלף.

מבחנים רבים הנמצאים בשימוש בארץ מקורם בארצות-הברית. הם תורגמו מאנגלית לעברית אך לא עובדו ולא הותאמו לארץ, היינו, לא נערכו בהם בדיקות אלמנטריות כגון בדיקת מהימנויות, תקפויות, נורמות, השוואות לשוניות וכדומה. לעתים קרובות גם התרגום עצמו לא נעשה כראוי (בשיטת ההצלבה).

שתי דוגמאות שליליות המוכרות היטב לפסיכולוגים בארץ הן מבחן ה-WAIS לבדיקת המשכל במבוגרים ושאלון ה-MMPI לבדיקת האישיות והפתולוגיה הנפשית; ואין צורך לפרט.

קיימים כמובן גם עיבודים תקינים, כגון מבחן ה-WPPSI ומבחן ה-WISC-R, שניהם מבחנים לבדיקת המשכל (בגיל הרך ובגיל הילדות). הם עובדו בידי עמית ליבליך מן האוניברסיטה העברית.

מהן הסיבות לכך שהמצב בישראל כה ירוד? נראה שהדבר אינו נעוץ בתרונן בכוח אדם מקצועי משובח בתחום המבחנים. לפחות בקצה הפירמידה יש כוח אדם הולם. אחת הבעיות היא, שאין תודעה צרכנית מפותחת. הלקוחות, הן פרטיים, הן מוסדות, אינם מודעים לכך שהשירות שהם מקבלים יכול להיות טוב יותר. בעיה אחרת היא ההשקעה הכספית הגדולה הנדרשת לביצוע פרויקטים לפיתוח מבחנים בקנה-מידה ארצי. הסיכונים של "גניבה תעשייתית" הם כל כך ברורים והשוק כל כך מצומצם, עד ששום משקיע שפוי לא ירצה להיכנס לתחום זה. (חברת "רמות" ליישום מדעי של אוניברסיטת תל-אביב היא בבחינת יוצא מן הכלל המוכיח את הכלל). מבחינת אינטרס הקידום האקדמי של מרצים, התרומה הפוטנציאלית של בניית מבחן היא אפסית: ועדות הקידום של האוניברסיטאות מתייחסות בעיקר לפרסומים מדעיים בכתבי-עת

יוקרתיים בארצות-הברית. כתבי עת כאלה אינם ששים לפרסם מאמרים שעניינם העיקרי הוא עיבוד גירסה עברית לשאלון או למבחן אמריקאי.

העתיד בתחום זה אינו ורוד. מבחנים טובים לא ייבנו רק מפני שזה נחוץ, או "ישוב לעם ישראל". באין חיזוקים חיוביים בעלי עוצמה, קשה לחזות שיפור בנושא.

באפריל 1985 הקים מנכ"ל משרד החינוך והתרבות ועדה לבדיקת השימוש במבחני כושר קבוצתיים בראשות פרופ' יוסף באשי, המדען הראשי של משרד החינוך והתרבות (ראה להלן כתב המינוי ורשימת חבריה). הועדה פירסמה ברבים את דבר קיומה והזמינה עדויות בכתב ובעל-פה. העדויות שנמסרו משתרעות על פני מאות עמודים (יש לקוות שביום מן הימים ייערכו ויפורסמו ברבים). דיוני הועדה נמשכו כארבעה חודשים וסוכמו בדו"ח קצר (יובא להלן).

הועדה בחרה למסור את המלצותיה כמעט ללא הנמקות. אחת הסיבות לכך היתה היעדר האפשרות לבטא בצורה תמציתית את שפע החומר שהצטבר. סיבה נוספת מופיעה בגוף הדו"ח, סעיף 1.6 לאמור:

"... התברר שגם אם ההמלצות להלן מוסכמות על כל חברי הועדה הרי הסכמתם של חברים שונים עלולה לנבוע מנימוקים שונים ואף מגוונים..." (אשר לגביהם יכולים להיות חילוקי דעות ב"נ).

במסגרת מאמר זה אין כל אפשרות להביא עדויות ואפילו לא את סיכומן. אפשר רק לנסות לחלק אותן לסוגים:

- א) טענות בדבר פגיעות מסוגים שונים בזכויות הפרט ("בעיות אתיות");
- ב) טענות בדבר הטיית תרבותיות במבחנים;
- ג) תפיסת המבחנים כסותרים את המעשה החינוכי;
- ד) טענות בדבר ליקויים באיכות המבחנים, ברמה המקצועית של הבוחנים ובדרך קבלת ההחלטות הקשורות בציוני המבחנים.

אפשר לומר כי ארבע הקבוצות הללו מייצגות את המלצותיה של ועדת באשי.

חלוקה זו לקבוצות מצביעה על כך ששלושה מסעיפי הביקורת משותפים לישראל ולארצות-הברית: זכויות הפרט, הטיית תרבותית וסתירת המעשה החינוכי. נושא שזכה לתשומת לב בארצות-הברית ולא טופל כלל בידי ועדת באשי, נוגע להשפעה של אמצעי הכנה שונים של נבחנים על הישגיהם במבחנים. הסיבה לכך היא שועדת באשי עסקה בתחום גילי שבו הנושא פחות רלוונטי, ואילו נושא שהעסיק את ועדת באשי, אך לא את הציבוריות האמריקנית, היה הרמה המקצועית של המבחנים, של הבוחנים ושל מקבלי ההחלטות. גם כאן ההבדלים בהדגשים מובנים מאליהם, שהרי המצב בארצות-הברית הרבה יותר תקין מאשר בישראל.

הועדה שמעה, כמובן, מספר רב של עדויות חיוביות אשר תמכו בהמשך השימוש במבחנים ואף בהרחבתו. הדבר בא לידי ביטוי בראש סעיף 2.3 של הדו"ח. אולם ההיבטים השליליים ריכזו את מרבית תשומת-הלב (הטייה זו, תידון יותר מאוחר).

נציג כעת את דוח הועדה כלשונו ואת חוזר המנכ"ל שיצא בעקבותיו.

**דו"ח ועדת מנכ"ל משרד החינוך בנושא מבחני כושר קבוצתיים  
(ועדת באשי)**

1. הקדמה

- 1.1 הועדה לבדיקת השימוש במבחני כושר קבוצתיים נתמנתה על ידי המנהל הכללי של משרד החינוך והתרבות ביום כ"ד בניסן תשמ"ה, 15 באפריל 1985 (העתק כתב המינוי - ראה נספח א').
  - 1.2 על הועדה הוטל לבדוק את מטרות השימוש במבחני כושר קבוצתיים, את התאמת המבחנים למטרות אלה ואת ההיבטים האתיים, המשפטיים והכספיים הכרוכים בעריכת המבחנים; כמו כן נתבקשה הועדה להמליץ על נהלי השימוש במבחנים אלה.
  - 1.3 על פי כתב המינוי, הגבילה הועדה את דיוניה למבחני כושר קבוצתיים. הועדה לא דנה במבחני כושר אינדיבידואליים, במבחני אישיות ובכלים הבודקים עמדות, נטיות וקשרים חברתיים.
  - 1.4 הועדה פעלה בהרכב כדלקמן: (ראה נספח א').
  - 1.5 הועדה קיימה עשר ישיבות ושלושה ימי דיונים מלאים, בתקופה שבין מאי לספטמבר, 1985. הועדה נפגשה עם נציגי הורים ומומחים מטעמם, עם מנהלי שירותים פסיכולוגיים חינוכיים, עם מומחים בתחום הפסיכומטריקה ובתחום המשפטים ועם נציגי האגודה לזכויות האזרח. חברי הועדה ראינו בנפרד כמה מנהלים של בתי-ספר על-יסודיים. הועדה פרסמה בעיתונות פנייה לציבור, בבקשה להתייחס לנושא. בעקבות פנייה זו התקבלו תזכירים וחוות דעת מהורים, גופים ציבוריים, ממורות יועצות וממנהלי בתי-ספר. כמו כן, פנתה הועדה למומחים בתחום החינוך, הפסיכולוגיה והסוציולוגיה וקיבלה את חוות דעתם. (רשימות שמיות של העדים, הפונים והמומחים - ראה נספח ב'). דיוני הועדה והמלצותיה התבססו על תחקירי העדים שהופיעו בפניה וכן על חוות הדעת והתזכירים שהוגשו בכתב.
  - 1.6 במהלך הדיונים התברר, שגם אם ההמלצות להלן מוסכמות על כל חברי הועדה, הרי שהסכמתם של חברים שונים עלולה לנבוע מנימוקים שונים ואף מגוונים. הוחלט על כן, למקד את הדו"ח בהמלצות, מבלי לפרט בכל מקרה את הנימוקים שהביאו להמלצות אלה. יחד עם זאת, הועדה מציעה לשקול את האפשרות לפרסם מסמכים אחדים הקשורים לעבודת הועדה, לרבות קטעים מהפרוטוקולים של הדיונים.
  - 1.7 מתוך מגמה להציע מדיניות העונה על הצרכים והבעיות שהתעוררו בהקשר של מבחני כושר קבוצתיים, גיבשה הועדה המלצות שונות, אשר את חלקן אפשר ליישם מיד ואילו את האחרות אפשר יהיה ליישם אם וכאשר ימצאו המשאבים הדרושים.
  - 1.8 על מנת לאפשר לגופים השונים להיערך על פי הנהלים החדשים מבעוד מועד, ממליצה הועדה לקבוע תקופת מעבר ליישום ההמלצות.
  - 1.9 הועדה דנה בנושא מבחני הכושר הקבוצתיים לפי שימוש בשלבי החינוך השונים. כמו-כן, התייחסה הועדה להיבטים האתיים הרלבנטים לכל שלבי החינוך. סעיפי הדו"ח תואמים הלוקה זו.
  - 1.10 הועדה מודה לכל האישים והגורמים שטרחו וסייעו לה, אם במתן עדות וחוות דעת, בכתב או בעל-פה ואם בהמצאת נתונים ומסמכים שונים שנדרשו לעבודתה.
- תודתנו נתונה במיוחד לגב' ניצה ננר, מרכזת הועדה, שבצעה את המשימות שהוטלו עליה ואשר נטלה על עצמה - ביעילות ובנאמנות.

2. כ ל ל י

2.1 מבחני הכושר הקבוצתיים משמשים היום במערכת החינוך לארבע מטרות מובחנות:

- (א) אבחון - לגבי רמת הכושר של הפרט, תוך התייחסות לפער בין רמה זו והישגיו הלימודיים.
- לגבי רמת הכושר הממוצעת של מסגרות (כיתות, בתי"ס וכיו"ב).
- (ב) שבוע - בהקבצות, בכיתות אם הטרוגניות ובכיתות אם בעלות מידה זו או אחרת של הומוגניות.
- (ג) ייעוץ והכוון - הניתנים במידה זו או אחרת כשירות לתלמיד עם סיום חטיבת הביניים או ביה"ס היסודי השמונה-שנתי.
- (ד) ברירה - בחירת מועמדים למסגרות סלקטיביות.

2.2 מן הראוי להדגיש, כי השימוש במבחני כושר קבוצתיים למטרות אלה אינו כללי: רשויות מקומיות ובתי-ספר נבדלים זה מזה בעצם השימוש במבחנים, במטרות לשמן הם עורכים אותם, ברמות הכיתה הנבדקות וכדומה. למרות זאת, מצאה הועדה לנכון להתייחס לארבעת השימושים, גם אם חלקם אינם נפוצים במערכת. בטבלה שלהלן מצוינות מטרות השימוש במבחני כושר קבוצתיים, לפי ארבעת שלבי החינוך, כפי שהן מקובלות כיום.

| מטרת השימוש שלב החינוך | א ב ה ו ן                       | שבוע                       | ייעוץ והכוון          | ברירה                         |
|------------------------|---------------------------------|----------------------------|-----------------------|-------------------------------|
| גן ילדים               | מבחן מוכנות                     | --                         | --                    | --                            |
| יסודי                  | (1) תלמידים חלשים<br>(2) מסגרות | הקבצות                     | סיום יסודי שמונה שנתי | קבלה לתכניות לתלמידים מוכשרים |
| חטיבת ביניים           | (1) תלמידים חלשים<br>(2) מסגרות | (1) הקבצות<br>(2) כיתות אם | סיום חטיבת ביניים     | --                            |
| חטיבה עליונה           | --                              | מגמות                      | --                    | קבלה לביה"ס                   |



2.3 בדיוניה עמדה הועדה על התועלת שעשויה לצמוח משימוש נכון ואתי במבחני כושר, לרבות מבחנים קבוצתיים. עם זאת, עמדה הועדה גם על הסכנה שמבחנים אלה עלולים, במצבים מסוימים, לשרת מטרות הנוגדות את מטרות החינוך. למשל, הרכבת כיתות הומוגניות בחטיבות הביניים על-פי מבחני כושר, בניגוד לעקרונות הרפורמה, או, למשל, הצגת ציון במבחן כושר כאפיון קבוע של הפרט, המעיד על יכולת מולדת. הועדה סבורה אפוא, שיש להגביל את השימוש במבחני הכושר הקבוצתיים בכפוף להמלצות שיובאו בהמשך. לאור תפיסה זו סיכמה הועדה את עמדתה הבסיסית באשר למטרות השימוש המקובלות במבחני כושר קבוצתיים בנקודות שלהלן:

2.3.1 אבחון של תלמידים אשר תפקודם הלימודי לקוי, וכן אבחון תלמידים לצורך ייעוץ והכוון, אינם מחייבים עריכת מבחני כושר לכלל התלמידים.

2.3.2 השיבוצים למיניהם נערכים על מנת להתאים את חומר הלימודים ודרכי ההוראה לרמת הישגיו העכשוית של התלמיד. משום כך, סביר לשבץ את התלמידים על פי הישגיהם בפועל בתחומי התוכן הרלוונטיים.

2.3.3 בתהליך הברירה (בניגוד לשיבוץ) המתקיים על-פי הלוח המובא לעיל, בשלב הקבלה לרטיבה העליונה ובברירת תלמידים מוכשרים, יש ניסיון לחזות את הצלחתו העתידית של התלמיד. יש במקרה זה הצדקה להשתמש בכל מידע התורם לתוקף החיזוי וליעילות הייעוץ וההכוון (בתנאי שהשימוש במידע זה אינו פסול מבחינה ערכית), ובמבחני כושר בכלל זה.

תפיסות אלו הינחו את הועדה בבואה להציע את המלצותיה.

### 3. בית-הספר היסודי וחטיבת הביניים

3.1 מאחר ועל-פי מדיניות משרד החינוך והתרבות ועל-פי החוק, מתקבלים התלמידים לבית-הספר היסודי ולחטיבת הביניים על-פי אזורי/רובעי רישום, אין צורך במבחני ברירה כלשהם לקראת כניסתם למוסד. יש להתנגד על כן, לעריכת מבחני כושר, דרך שיגרה, לכלל התלמידים בגן הילדים (1) ולקראת קבלה לחטיבת הביניים.

3.2 לרשות חטיבות הביניים הקולטות עומד כל המידע המועבר מבית-הספר היסודי (ציונים, הערכות וכיו"ב). יחד עם זאת, מאחר וחטיבת-הביניים קולטת תלמידים מבתי-ספר ברמות שונות, היא עשויה להזדקק לבחינה אובייקטיבית אחידה של כלל התלמידים הנקלטים, וזאת לצורך:  
(1) חלוקה לכיתות אם בעלות מידה זו או אחרת של הומוגניות  
(ב) חלוקה להקבצות (מיון זה קיים גם בבית-הספר היסודי)  
(ג) התאמת חומר הלימודים ודרכי ההוראה לרמת הישגיהם ההתחלתית של התלמידים.

הועדה סבורה שהמבחנים הרלוונטיים לכל שלוש המטרות שלעיל, הם מבחני הישגים ולכן אין להשתמש במבחני כושר. הועדה ממליצה שבית-הספר הקולט יערוך במסגרתו מבחנים פנימיים אחידים בתכנים הרלוונטיים לשיבוץ, מתוך מגמה להתאים את חומר הלימוד ודרכי ההוראה (כגון מתמטיקה, אנגלית, הבנת הנקרא). החלוקה לכיתות אם הטרוגניות אינה מצריכה כלי מיון כלשהו, היות ואפשר לחלק את התלמידים באופן אקראי.

3.3 למרות שהחלוקה לכיתות אם ולהקבצות אינה נושא השייך לתחום סמכותה של הועדה, ביקשו חברי הועדה להדגיש את הצורך להימנע משיבוץ חד-פעמי של תלמידים בהקבצות/בכיתות אם (אם הן לא הטרוגניות לחלוטין) למשך שלוש שנות שהותם בחטיבה. יש לחזור ולשבץ את התלמידים בהמשך תקופת לימודיהם בחטיבה, וזאת על פי הישגיהם הלימודיים.

3.4 הועדה ערה לצורך לאבחן תלמידים בודדים על מנת לבחור דרכי טיפול מתאימות. במקרה זה אפשר לערוך מבחני כושר לתלמידים אלה.

---

(1) אמנם מבחני המוכנות בגן הילדים נערכים במתכונת אינדיבידואלית, אך יש לכלול אותם במסגרת מבחני הכושר הקבוצתיים, כאשר הם ניתנים לכלל התלמידים.

3.5 מקובל על הועדה השימוש במבחני כושר קבוצתיים לצורך אבחון של מסגרות (כיתות, בתי-ספר, איזורי פיקוח וכיו"ב) או לצורך מחקר. המבחן ייערך אז באופן אנונימי כך שאי-אפשר יהיה לזהות את תוצאותיו של כל תלמיד. משרד החינוך והתרבות ימנה גוף מוסמך אשר יהיה רשאי, על פי פנייה אליו, לאשר חריגה מכלל זה ובלבד שתוצאות המבחן לא יישמשו לקבלת החלטה כלשהי הנוגעת לתלמיד ואף לא ירשמו בכרטיס התלמיד.

3.6 ניתן לערוך מבחני כושר קבוצתיים לשם איתור תלמידים מוכשרים למסגרות מיוחדות בתנאים הבאים:

- (א) להודיע להורים שמבחנים אלה אינם מבחני חובה
- (ב) המבחנים יתקיימו רק לאחר שניתנה להורים זכות התנגדות
- (ג) תוצאות המבחן לא יישמשו בסיס להחלטה אחרת
- (ד) בתי-הספר יקבלו רק רשימות שמיות של תלמידים שאותרו לתכניות המיוחדות. בכל מקרה, לא יועברו לבתי-הספר תוצאות מפורטות.

3.7 הועדה ממליצה בפני משרד החינוך והתרבות להקים בנק פריטים בתחומי תוכן שונים, אשר בעזרתם יוכלו חטיבות-הביניים לבנות מבחני הישגים לצורך שיבוץ.

#### 4. המעבר לחטיבה העליונה

4.1 בחלק מבתי-הספר העל-יסודיים הקולטים תלמידים בכיתות ט'-י', מתקיימת ברירה מתוך מגמה לבחור את המועמדים המתאימים ביותר.

4.2 לצורך ברירת התלמידים, יוכל בית-הספר העל-יסודי להיעזר בכל פרטי המידע שימסרו על התלמיד על-ידי חטיבת הביניים/בית-הספר היסודי (למעט תוצאות של מבחני כושר אינדיבידואליים או קבוצתיים שנערכים לתלמידים בחטיבת הביניים/בבית-הספר היסודי לצורך טיפול, הכוון וייעוץ), וכן, להשתמש בכלים נוספים לפי צרכיו (כגון מבחני הישג).

4.3 בית-ספר על-יסודי שירצה להסתמך גם על מבחני כושר לצורך ברירת מועדיו, יפנה למשרד החינוך והתרבות לקבלת אישור. בפנייתו יציין בית-הספר מהם הקריטריונים שעל פיהם נבחרים התלמידים, מיהו הגוף שיערוך את המבחן, מהו כוח-האדם שיבצע, אילו מבחנים כלולים בסוללה, למי ובאיזה אופן יימסרו התוצאות, באיזה אופן יוכנו הנבחנים לקראת הבחינה וכדומה. אישור זה יינתן בתנאי שבית-הספר יתחייב למסור את תוצאות המבחן להורים על-פי בקשתם (ראה סעיף 5.2 בהמשך).

4.4 בהליכי ייעוץ והכוון יש מקום להיעזר גם בתוצאות של מבחני כושר. יחד עם זאת, הועדה לא שוכנעה שתהליך הייעוץ וההכוון מחייב בחינת כלל התלמידים במבחני כושר. יש אפוא מקום לערוך מבחני כושר לתלמידים בודדים, כאשר המידע הקיים עליהם הוא סותר או בלתי מספק. הנאמר בסעיף זה כפוף לנאמר בסעיף 5 להלן.

## 5. החיבת האתי

5.1 בחינת תלמיד במבחני כושר מחייבת הודעה להורים, בה יובהר מה מטרת המבחן, איזה שימוש יעשה בו ולמי תועברנה התוצאות. כמו כן, יובהר בהודעה, שלהורים הזכות להתנגד לעריכת המבחן. בית-הספר יאשר בכתב ביצוע ניהול זה. התנגדו הורים - הענין יובא לדיון והכרעה בפורום מקצועי מחייב. יש להוציא מכלל זה את האמור בסעיף 3.5. תוצאות המבחנים לא יועברו לכל גורם אחר שלא דווח להורים, אלא אם כן התקבלה הסכמתם בכתב.

5.2 ההורים רשאים לקבל את תוצאות המבחנים בליווי הסבר. צורת ההסבר וניסוחו יקבעו על ידי ועדה מוסמכת.

תנומים: חברי הועדה  
ספטמבר 1985

נספח א' : כתב המינוי

אל: פרופ' יוסף באשי  
פרופ' עוזר שילד  
מר אריה בריק  
דר' ברוך נבו  
מר יוסף יונאי  
מר יוסף קולודנר  
דר' עמירם רביב  
גב' נורה כהן  
גב' חנה שחורי

לאחרונה אנו עדים לקיומם של מבחנים רבים במערכת החינוך שאינם מהווים חלק אינטגרלי מתהליך ההוראה.

מבחנים אלה נעשים על ידי גורמים שונים ומטרתם: אבחון, ייעוץ, בדיקת רמת הישגים כללית ואישית, ועוד.

במטרה לבדוק נושא זה, מיניתי בחדש מאי 1984, ועדה בראשותו של פרופ' קוגלמס, המדען הראשי דאז. בשל פרישתו של פרופ' קוגלמס מתפקידו, הופסקה זמנית פעילות הועדה. עם כניסתו של פרופ' באשי לתפקיד המדען הראשי, ברצוני לחדש את פעילות הועדה במתכונת מורחבת ולמנותכם כחברים בה.

מסמכותה של הועדה יהיה לקבוע כללים המסדירים את השימוש במבחנים במערכת החינוך, תוך התייחסות לנקודות הבאות:

א. מטרת השימוש במבחנים

ב. התאמת המבחנים למטרות אלה, תוך התייחסות ליתרונותיהם ולחסרונותיהם ככלי מדידה

ג. ההיבטים האתיים, המשפטיים והכספיים הכרוכים בהעברת המבחנים, בשימוש ובמסירת התוצאות המתקבלות מהם.

עוד מתבקשת הועדה לקבוע נהלים ליישום השוטף של הכללים וההנחיות שיאומצו על ידה.

על הועדה לסיים את עבודתה עד 30.9.85.

בראש הועדה יעמוד פרופ' יוסף באשי.

אני מודה לכם על היענותכם ומאחל לכם הצלחה בעבודתכם.

ב ב ר כ ה ,

אליעזר שמואלי

כ"ד ניסן תשמ"ה  
15 אפריל 1985

רשימת העידים והמומחים שחוות דעתם נמסרה בכתב ו/או בעל-פה

פרופ' חיים אדלר, ביה"ס לחינוך, האוניברסיטה העברית  
דר' גינה אורתור, ביה"ס לחינוך, האוניברסיטה העברית  
הה"כ שולמית אלוני  
פרופ' יהודה אמיר, המחלקה לפסיכולוגיה, אוניברסיטת בר-אילן  
דר' קלמן בנימיני, לשעבר מנהל השירות הפסיכולוגי חינוכי, עיריית ירושלים  
ד"ר גרשון בן-שחר, המחלקה לפסיכולוגיה, האוניברסיטה העברית  
פרופ' רות גביזון, הפקולטה למשפטים, האוניברסיטה העברית  
מר אברהם דגני, ביה"ס הריאלי, חיפה  
גב' סילביה זילברמן, מנהלת השירות הפסיכולוגי חינוכי, עיריית ירושלים  
פרופ' מיכאל חן, ביה"ס לחינוך, אוניברסיטת תל-אביב  
גב' גאולה יפה, נציגת ועד ההורים הארצי  
מר אריה לית, רשם מאגרי מידע, משרד המשפטים  
גב' רינה לוי, השירות הפסיכולוגי חינוכי, עיריית ירושלים  
פרופ' אלחנן מאיר, מנהל מחלקת המחקר, מכון הדסה לייעוץ  
דר' עמוס ניר, מנהל השירות הפסיכולוגי חינוכי, עיריית תל-אביב  
דר' מרדכי ערן, הסתדרות הפסיכולוגים  
פרופ' קרל פרנקנשטיין, ביה"ס לחינוך, האוניברסיטה העברית  
דר' סורל קאהן, ביה"ס לחינוך, האוניברסיטה העברית  
פרופ' שלמה קוגלמס, המחלקה לפסיכולוגיה, האוניברסיטה העברית  
פרופ' זאב קליין, המחלקה לפסיכולוגיה, האוניברסיטה העברית.  
ד"ר שמואל קנלד, ביה"ס התיכון שליד האוניברסיטה העברית  
ד"ר אליעזר קרני, ביה"ס לחינוך, אוניברסיטת תל-אביב  
ד"ר יוסף שביט, המחלקה לסוציולוגיה, אוניברסיטת חיפה  
הורים שעתרו לבג"צ בעניין עריכת מבחני מילתא ע"י עיריית ירושלים  
מכון הדסה לייעוץ (מר יצחק גרתי, מר זאב סרדי)  
מכון סאלד (ד"ר רחל זורמן, גב' לאה שפטיה)  
מרכז לחינוך טכנולוגי (ד"ר אליעזר פישמן)

**נספח ג' : חוזר המנכ"ל**

אל: כל מוסדות החינוך

**חוזר מיוחד  
תשמ"ו - שפול/16**

**מבחני כושר קבוצתיים**

ועדה, בראשותו של פרופ' יוסף באשי, המדען הראשי של משרד החינוך והתרבות, דנה בשימוש במבחני כושר קבוצתיים במערכת החינוך והגישה את הדו"ח המצורף המכיל את המלצותיה להסדרת השימוש במבחנים אלה. המלצות הועדה אומצו במלואן על ידי המנהל הכללי.

מאחר וחלק מהמלצות מחייבות גיבוש נהלים מעשיים והקמת גופים להחלטה ולאישור, תידרש תקופת מעבר אשר במהלכה יגובשו הנהלים, ימונו הגופים ויקבע המועד להחלת הנהלים החדשים.

למרות זאת, סעיפים א', ב', ו-ג' להלן יכנסו לתוקף עם פרסום הודעה זו. בעניינים אחרים, ניתן לנהוג לפי הנוהל הקיים במשך כל תקופת המעבר.

(א) תיפסק לאלתר עריכת מבחני כושר קבוצתיים, לכלל התלמידים בכיתות א'-ז' החל מיום פרסום הודעה זו. אין הוראה זו מתייחסת למבחני כושר אינדיבידואליים שימשיכו להתקיים בהתאם לצורך (ראה סעיף 3.4 בדוח הועדה).

(ב) המבחנים לאיתור תלמידים מוכשרים יערכו בהתאם לנוהל המפורט בסעיף 3.6 בדו"ח הועדה. הוראה זו תכנס לתוקף עם פרסום הודעה זו.

(ג) מותרת עריכת מבחני כושר קבוצתיים בבתי-הספר היסודיים ובחטיבות הביניים לצורך אבחון של מסגרות (כיתות, בתי-ספר וכיו"ב) או לצורך מתקן. המבחן ייערך אז באופן אנונימי (ראה סעיף 3.5 בדוח הועדה). המבקשים לחרוג מכלל זה יפנו לפסיכולוג המחוזי. הפונים במחוז תל-אביב יפנו למנהל השירות הפסיכולוגי ייעוצי, רח' דוד המלך 8, ירושלים.

(ד) הנהלים הנוגעים לעריכת מבחני כושר לכלל התלמידים בגני הילדים ומועד החלתם, יפורסמו בחודשים הקרובים בחוזר המנכ"ל.

במסגרת זו, תותר בתקופת המעבר בחינת תלמידים הלומדים בכיתות ח' ו-ט' לצורך ברירה ושיבוץ כנ"ל.

ב ב ר כ ה,

אליעזר שמואלי

י"ד טבת תשמ"ו  
26 דצמבר 1985

## עבודת הועדה - בראייה לאחור

מחבר מאמר זה היה חבר בוועדת באשי וחתם בשעתו על נוסח הדו"ח שלה. זמן מה לפני שסיימנו את עבודתנו, התחלתי לחוש בחוסר נוחות עמום ביחס לכיוון הכללי שנקטה הועדה בשלבי ניסוח ההמלצות הסופיות. תחושה זו גברה לאחר פרסום הדו"ח (פרסום חפוז מדי ולא מתוכנן כראות) ולשמע התגובות שחלו להגיע מן השדה. במשך הזמן התגבשו הפקפוקים לכדי הכרה, כי מערך חשיבה קולקטיבי מסויים הנחה אותנו, חברי הועדה, לכיוון לא נכון. המערך הזה התבטא בכך שתשומת לבנו התמקדה בהשלכות השליליות של הפעלת מבחנים במערכת החינוך. התמקדנו בחיפוש אחר דרכים להקטנת הנזקים החינוכיים והמוסריים, אשר נראו לנו כקשורים להפעלת מבחנים. לכן, המלצנו על מגבלות ותיחומים ולא השכלנו להביא בחשבון את ההתפתחויות החלופיות השליליות אשר עתידות לצמוח אם יופעלו ההמלצות כלשונן ובמלואן. במלים אחרות, ניתחנו את החשפעות הצפויות על המערכת באופן צר מדי.

הבה נעייין בהתפתחויות אפשריות אחדות כאלו:

\* סעיף 3.1. הערת שוליים (1): המלצה היא שלא לקיים מבחני מוכנות לכיתה א' כאשר הם מועברים במסגרת גן הילדים, לכלל הילדים. המלצה זו משאירה בידי הגננת את ההחלטה אם להפנות ילד מסוים לבדיקה מוכנות. במשך שנים רבות שררה החשקפה, כי גננות מתקשות באיתור ילדים בלתי בשלים ללימודים וזאת משום שאין להן קני-מידה כלליים (נורמות), משום שהן נוטות לזהות חוסר בשלות ללימודים עם הפרעות התנהגות ומשום שבעיות ספציפיות בתחום הפסיכומטורי ובתחום התפיסה החזותית אינן באות לידי ביטוי במטלות הגן.

חוסר אמצעים הוא שמנע ברשויות מקומיות מסוימות העברת בדיקות מוכנות אנטנסיות כוללות, ככלי עזר לגננת, להורים ולבית הספר הקולט.

כתוצאה מהמלצות הועדה, ילד אשר לא יופנה על ידי הגננת לבדיקה פרטנית, יכנס לכתה א' ללא בקרה נוספת. ביטלנו, אפוא, טעות מסוג אחד (זיהוי ילד באמצעים פסיכודיאגנוסטיים כבלתי מוכן לבית-ספר בעוד שהוא מוכן כדבעי), והגדלנו את הסבירות לטעות מסוג שני ("המיקח" של ילד לא מוכן אל בית-הספר). איזה משני סוגי הטעויות חמורה יותר?

\* המלצות 3.2, 3.7, 4.2 מדגישות את ערכם של מבחני הישג למטרות ברירה ושיבוץ, וזאת כמשקל נגד למבחני כושר. אם אמנם מבחני הישג יוצגו וישמשו כחלופה מושלמת למבחני כושר, מה יהיו התוצאות החברתיות? החלחה במבחני הישג תלויה במשתני רקע סוציו-אקונומיים, בתכנית הלימודים



המסוימת שדרכה עבר התלמיד ובנכונות של ההורים להשקיע בספרי לימוד ובמורים פרטיים. מבחני כושר מנסים לפחות להיות הוגנים מבחינה תרבותית. מבחני הישג, על פי טיבם, אינם מכירים כלל בבעיה!

\* המלצות 2.3.1, 3.4, 4.4 מתייחסות להבחנה בין העברת מבחנים לכיתות שלמות, לבין העברת מבחנים לפרטים לצורך ייעוץ, הכוון, אבחון קשיי לימודים, וכיו"ב. הועדה מכירה בנחיצותם של המבחנים עבור פרטים, ופוסלת, כידוע, את העברתם הטוטאלית. הבעיה היא שזימון תלמיד יחיד מתוך כיתה שלמה, לתהליך דיאגנוסטי, מעורר מחדש את בעית הסטיגמה אשר ממנה, כזכור, ניסה הדו"ח לברוח (סעיף 2.3).

\* המלצה 4.3 מרמזת באופן ברור מאד, כי השלב הבא של עבודת ועדת באשי (או של ועדה ממשיכה) הוא הגדרת קריטריוני איכות למבחנים פסיכולוגיים, שישמשו לברירת תלמידים לחטיבה העליונה. ללא קריטריונים כאלה, לא ניתן לבצע את ההמלצה האומרת שהאישור להפעיל מבחנים פסיכולוגיים ינתן רק לאחר בדיקה. ובכן, קביעת סטנדרטים גבוהים לאיכות מבחנים היא מטרה חשובה מאוד, במיוחד נוכח המצב העלוב בהווה. עם זאת, כדאי לבדוק מה יקרה אם אמנם יוגשמו ההמלצות, קרי, אם אמנם ייקבעו סטנדרטים מחמירים למתן אישור שימוש במבחנים פסיכולוגיים. ובכן, התפתחות כמעט בלתי נמנעת במקרה זה תהיה העברת הדגש בתהליך הברירה למכשירי ברירה ומיון אחרים. אלה יכולים להיות מבחני הישג מקומיים, ראיונות אישיים וציוני מורים משנים קודמות. מאחר ועל כל המשתנים הללו לא יחולו כל הגבלות חוקיות, יכול להיווצר מצב שבו תורע איכותו של תהליך הברירה לחטיבה העליונה, בכך שיתבצע מעבר למשתני ברירה הרבה פחות איכותיים והרבה פחות מעובדים, מאשר מבחני הכושר.

נכון הוא כי השימוש במבחני כושר במערכת החינוך כיום, כפי שהצטייר מן העדויות, הוא בעייתי ביותר. קשה היה לקבל בשוויון נפש את אותם קטעי עדויות אשר עסקו ברמה המקצועית הירודה של פיתוח המבחנים, תחזוקתם, רמת הבחונים ואופן השימוש בציונים, אולם, חשוב לזכור כי יש מצבי הכרעה שבהם הבחירה אינה בין חלופה טובה לגרועה, אלא בין חלופה גרועה לגרועה יותר. ביטול המבחנים יפתור אמנם בעיות אחדות, אך עשוי ליצור אחרות, לא פחות קשות.

## לקראת שנות ה-90

קיימת סבירות גבוהה לכך שהביקורת הקשה שהוטחה בארצות-הברית בשנות ה-60 וה-70 נגד מבחנים פסיכולוגיים ונגד הפסיכולוגים המפעילים אותם, תגיע בעתיד הקרוב לישראל. הדברים כבר מתרחשים לנגד עינינו. בווריאציות שונות, ועל במות אלו ואחרות, עולים כבר כיום הנושאים הקלאסיים אשר היוו אבני פינה ב-  
anti-test movement: זכויות הפרט, הטיות כנגד קבוצות מיעוטים, פגיעה במעשה החינוכי, בהכנות לבחינות, רמתם הטכנית-מקצועית של המבחנים ושל המחקרים המלווים אותם.

אולם במידה והביקורת בישראל תושמע בעוצמה ובחריפות שבה הוטחה בארצות-הברית, עלולים להיווצר נסיגה מקצועית חמורה והפסד בלתי-נסבל לחברה כאן. הממסד הפסיכולוגי בישראל הוא עדיין חלש, הביקוש למבחנים מצומצם ומספר העוסקים בתחום מוגבל. אם תתרחש בשנים הקרובות התקפה משולבת של כלי-התקשורת והממסד הפוליטי, בסיוע עמיתים "מבפנים", על המבחנים, הרי המעט שנוצר יכול ללכת לאיבוד.

האם אפשר למנוע התפתחויות שליליות כאלו? מה יש לעשות?  
יש לנו סיכוי טוב להתמודד בהצלחה עם הקשיים הצפויים. אנו יודעים מה קרה בארצות אחרות ולפיכך יכולים להתכונן לקראת הבאות, מה גם שהעשייה הפסיכולוגית בארץ מתבצעת במסגרת מצומצמת, משפחתית כמעט. הדבר מאפשר תקשורת נוחה, שיתוף פעולה ותיאום מאמץ.

האם צריך לעשות משהו בנידון?  
ברור שצריך לטפל בבעיה, וזאת כדי למנוע, בעיקר מן הצרכנים, קרי, מן הציבור, את הנזקים הנובעים משיתוק ונסיעה בתחום המבחנים.

מה יש לעשות?  
ראשית, עלינו להיות קשובים יותר לביקורת. חשוב שנדע בפירוט ולעומק, מה מרגישים כלפינו ומה מבקשים מאיתנו הנבחנים, מקבלי ההחלטות ואנשי התקשורת. יתכן שכדאי להקים מערכות משוב (פשוטות יחסית) במיוחד למטרה זו (ראה לדוגמה: Nevo Sfej, 1985, 1985). שנית, עלינו להעלות את הרמה הטכנית-מדעית של הכלים שלנו. המובילים בתחום התכנון והבנייה, צריכים להיות הגופים הגדולים והחזקים יותר בשדה המקצועי שלנו: צה"ל, מכון הדסה לייעוץ בבחירת מקצוע, המרכז הארצי לבחינות ולהערכה, מכונים פרטיים מבוססים. אלה צריכים להיות החלוץ ההולך לפני המחנה ועליהם להעניק מהידע שלהם לאחרים.

שלישית, עלינו לגייס את המתודה המחקרית כך שנוכל להתמודד עם השאלות הקריטיות של הטיה במבחן ושל האימון למבחן. עד כה בוצעו בישראל מעט מדי מחברים. ביסוס מדעי של עבודתנו הוא ממש הכרח עבורנו.

רביעית, חייבת להימצא דרך להגדיל את מעורבות האוניברסיטאות בנושא.

חמישית, עכיננו לנקוט פעולות מיידיות, כאשר הן מתבקשות מאליהן על פי השכל הישר, גם כאשר נגרמת אי-נוחות ארגונית או הוצאה כספית. הדברים אמורים בעיקר לגבי היבטים אתיים שונים והיבטים של זכויות הפרט. ייתכן, ואולי אף ודאי, שהמלצות אלו יביאו להגדלת העלות הכוללת של הפעלת מבחנים פסיכולוגיים בישראל, וייקרו את השירות לצרכנים. אך מהי האפשרות האחרת?

- Anastasi, A. (1948). The nature of psychological "traits". Psychological Review, 55, 127-138.
- Anastasi, A. (1954a). Psychological testing (1st ed.). New York: Macmillan.
- Anastasi, A. (1954b). Tested intelligence and family size: Methodological and interpretive problems. Eugenics Quartely, 1, 155-160.
- Anastasi, A. (1956a). Age changes in adult test performance. Psychological Reports, 2, 509.
- Anastasi, A. (1956b). Intelligence and family size. Psychological Bulletin, 53, 187-209.
- Anastasi, A. (1958a). Differential psychology (3rd ed.). New York: Macmillan.
- Anastasi, A. (1958b). Heredity, environment, and the question "How?" Psychological Review, 65, 197-208.
- Anastasi, A. (1962). The longitudinal study of populations. Indian Psychological Bulletin, 7 (Part II), 25-28.
- Anastasi, A. (1970). On the formation of psychological traits. American Psychologist, 25, 899-910.
- Anastasi, A. (1981). Sex differences: Historical perspectives and methodological implications. Developmental Review, 1, 187-216.
- Anastasi, A. (1982). Psychological testing (5th ed.). New York: Macmillan.
- Anastasi, A. (1983a). Evolving trait concepts. American Psychologist, 38, 175-184.
- Anastasi, A. (1983b). Traits, states, and situations: A comprehensive view. In H. Wainer, & S. Messick (Eds.), Principals of modern psychological measurement: A Festschrift for Frederic M. Lord (pp. 345-356). Hillsdale, NJ: Erlbaum.
- Anastasi, A. (1985a). Psychological testing: Basic concepts and common misconceptions. In A.M. Rogers, & C.J. Scheirer (Eds.), The G. Stanley Hall Lecture Series (Vol. 5, pp. 87-120). Washington, DC: American Psychological Association.
- Anastasi, A. (1985b). Reciprocal relations between cognitive and affective development - with implications for sex differences. In T.B. Sonderegger (Ed.), Psychology and gender (Nebraska Symposium on Motivation), Vol. 32, pp. 1-35). Lincoln, NB: University of Nebraska Press.

- Angoff, W.H. (Ed.) (1971). The College Board admissions testing program: A technical report on research and development activities relating to the Scholastic Aptitude Test and achievement tests. New York: College Entrance Examination Board.
- Azuma, H. (1984). Psychology in a non-Western country, International Journal of Psychology, 19, 45-55.
- Bennett, G.K., Seashore, H.G., & Wesman, A.G. (1947-1984). Differential Aptitude Tests. Cleveland, OH: The Psychological Corporation.
- Bradford, E.J.G. (1925). Can present scholastic standards be maintained? Forum of Education, 3, 186-198.
- Bruner, F.G. (1908). The hearing of primitive peoples. Archives of Psychology, No. 11.
- Bruner, J.S. (1966). On cognitive growth II. In J.S. Bruner, R.R. Olver, & P.M. Greenfields (Eds.), Studies in cognitive growth, (pp. 30-67). New York: Wiley.
- Burt, C. (1946). Intelligence and fertility. London: Hamilton.
- Burt, C. (1949). The structure of the mind; a review of the results of factor analysis. British Journal of Educational Psychology, 19, 100-111; 176-199.
- Burt, C. (1954). The differentiation of intellectual ability. British Journal of Educational Psychology, 24, 76-90.
- Carroll, J.B., & Horn, J.L. (1981). On the scientific basis of ability testing. American Psychologist, 36, 1012-1020.
- Cattell, R.B. (1937). The fight for our national intelligence. London: King.
- Cattell, R.B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. Journal of Educational Psychology, 54, 1-22.
- Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. Journal of Educational Measurement, 5, 115-124.
- Coan, R.W. (1964). Facts, factors, and artifacts: The quest for psychological meaning. Psychological Review, 71, 123-140.
- Cole, M., & Bruner, J.S. (1971). Cultural differences and inferences about psychological processes. American Psychologist, 26, 867-876.
- Cole, M., Gay, J., Glick, J.A., & Sharp, D.W. (1971). The cultural context of learning and thinking. New York: Basic Books.
- Cole, M., & Scribner, S. (1974). Culture and thought. New York: Wiley.
- Dawson, S. (1932). Intelligence and fertility. British Journal of Psychology, 23, 42-51.

- Demming, J.A., & Pressey, S.L. (1957). Tests "indigenous" to the adult and older years. Journal of Counseling Psychology, 4, 144-148.
- Dockrell, W.B. (1965). Cultural and educational influences on the differentiation of ability. Proceedings of the 73rd Annual Convention of the American Psychological Association, 317-318.
- Donlon, T.F. (Ed.). (1984). The technical handbook for the College Board Scholastic Aptitude Test and achievement tests. New York: College Entrance Examination Board.
- Doppelt, J.E., & Wallace, W.L. (1955). Standardization of the Wechsler Adult Intelligence Scale for older persons. Journal of Abnormal and Social Psychology, 51, 312-330.
- Dwyer, C.A. (1974). Influence of children's sex role standards on reading and arithmetic achievement. Journal of Educational Psychology, 66, 811-816.
- Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests (3rd ed.). Princeton, NJ: Educational Testing Service.
- Entwisle, D.R., & Baker, D.P. (1983). Gender and young children's performance in arithmetic. Developmental Psychology, 19, 200-209.
- Filella, J.F. (1960). Educational and sex differences in the organization of abilities in technical and academic students in Colombia, South America. Genetic Psychology Monographs, 61, 115-163.
- Fitzpatrick, J.L. (1978). Academic underachievement, other-direction, and attitudes toward women's roles in bright adolescent females. Journal of Educational Psychology, 70, 645-650.
- Fleishman, E.A. (1975). Toward a taxonomy of human performance. American Psychologist, 30, 1127-1149.
- Fleishman, E.A., & Quaintance, H.K. (1984). Taxonomies of human performance: The description of human tasks. New York: Academic Press.
- Garrett, H.E. (1946). A developmental theory of intelligence. American Psychologist, 1, 372-378.
- Garth, T.R. (1931). Race psychology: A study of racial mental differences. New York: McGraw-Hill.
- Ghiselli, E.E. (1959). The generalization of validity. Personnel Psychology, 12, 397-402.
- Ghiselli, E.E. (1966). The validity of occupational aptitude tests. New York: Wiley.
- Ghiselli, E.E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

- Goodenough, F.L., & Harris, D.B. (1950). Studies in the psychology of children's drawings: II. 1928-1949. Psychological Bulletin, 47, 369-433.
- Greenfield, P.M. (1966). On culture and conservation. In J.S. Bruner, R.R. Olver, & P.M. Greenfield (Eds.), Studies in cognitive growth, (pp. 225-226). New York: Wiley.
- Guilford, J.P. (1967). The nature of intelligence. New York: McGraw-Hill.
- Guilford, J.P. (1981). Higher-order structure-of-intellect abilities. Multivariate Behavioral Research, 16, 411-435.
- Guilford, J.P., & Hoepfner, R. (1971). The analysis of intelligence. New York: McGraw-Hill.
- Hakstian, A.R., & Cattell, R.B. (1978). Higher-stratum ability structures on a basis of twenty primary abilities. Journal of Educational Psychology, 70, 657-669.
- Hayes, K.J. (1962). Genes, drives, and intellect. Psychological Reports, 10, 299-342.
- Horn, J.L. (1968). Organization of abilities and the development of intelligence. Psychological Review, 75, 242-259.
- Humphreys, L.G. (1962). The organization of human abilities. American Psychologist, 17, 475-483.
- Irving, J., Perl, H., Trickett, E.J., & Watts, R. (1984). Minority curricula or a curriculum of cultural diversity? Differences that make a difference. American Psychologist, 39, 320-321.
- Jones, H.E., & Conrad, H.S. (1933). The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. Genetic Psychology Monographs, 13, 233-298.
- Jones, L.V. (1984). White-black achievement differences: The narrowing gap. American Psychologist, 39, 1207-1213.
- Kelley, T.L. (1928). Crossroads in the mind of man: A study of differentiable mental abilities. Stanford, CA: Stanford University Press.
- Khan, S.B. (1970). Development of mental abilities: An investigation of the "differentiation hypothesis." Canadian Journal of Psychology, 24, 199-205.
- Khan, S.B. (1972). Learning and the development of mental ability. Educational Research Journal, 9, 607-614.
- Klineberg, O. (1935). Race differences. New York: Harper.
- Knox, H.A. (1914). A scale based on the work at Ellis Island for estimating mental defect. Journal of the American Medical Association, 62, 741-747.

- Lancy, D.F. (1983). Cross-cultural studies in cognition and mathematics. New York: Academic Press.
- Lewontin, R.C. (1974). The analysis of variance and the analysis of causes. American Journal of Human Genetics, 26, 400-411.
- Linn, R.L., & Werts, C.E. (1971). Considerations for studies of test bias. Journal of Educational Measurement, 8, 1-4.
- Mackenzie, B. (1984). Explaining race differences in IQ: The logic, the methodology, and the evidence. American Psychologist, 39, 1214-1233.
- Mann, C.W. (1940). Mental measurements in primitive communities. Psychological Bulletin, 37, 366-395.
- McCormick, E.J., Jeanneret, P.R., & Meacham, R.C. (1972). A study of job characteristics and job dimensions as based on the Position Analysis Questionnaire (PAQ). Journal of Applied Psychology, 56, 347-368.
- Miles, C.C., & Miles, W.R. (1932). The correlation of intelligence scores and chronological age from early to late maturity. American Journal of Psychology, 44, 44-78.
- Milton, G.A. (1957). The effects of sex-role identification upon problem-solving skills. Journal of Abnormal and Social Psychology, 55, 208-212.
- Mischel, W. (1968). Personality and assessment. New York: Wiley.
- Mischel, W. (1977). On the future of personality measurement. American Psychologist, 32, 246-254.
- Mischel, W., & Peake, P.K. (1982). Beyond deja vu in the search for cross-situational consistency. Psychological Review, 89, 730-755.
- Nash, S.C. (1979). Sex role as a mediator of intellectual functioning. In M.A. Wittig, & A.C. Petersen (Eds.), Sex-related differences in cognitive functioning: Developmental issues (pp. 263-302), New York: Academic Press.
- Olson, D.R. (1976). Culture, technology, and intellect. In L. Resnik (Ed.), The nature of intelligence (pp. 189-202). Hillsdale, NJ: Erlbaum.
- Olson, D.R. (1977). From utterance to text: The bias of language in speech and writing. Harvard Educational Review, 47, 257-281.
- Pearlman, K. (1980). Job families: A review and discussion of their implications for personnel selection. Psychological Bulletin, 87, 1-28.
- Pearlman, K., Schmidt, F.L., & Hunter, J.E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406.



- Petersen, N.S., & Novick, M.R. (1976). An evaluation of some models for culture-fair selection. Journal of Educational Measurement, 13, 3-29.
- Peterson, D. (1968). The clinical study of social behavior. New York: Appleton-Century-Crofts.
- Pressey, S.L., & Jones, A.W. (1955). 1923-1953 and 20-60 age changes in moral codes, attitudes, and interests, as shown by the "X-0 Tests". Journal of Psychology, 39, 485-502.
- Rebelsky, F., & Daniel, P.A. (1976). Cross-cultural studies of infant intelligence. In M. Lewis (Ed.), Origins of intelligence: Infancy and early childhood (pp. 279-297). New York: Plenum.
- Reed, H.J., & Lave, J. (1979). Arithmetic as a tool for investigating relations between culture and cognition. American Ethnologist, 6, 568-582.
- Rosenberg, B.G., & Sutton-Smith, B. (1960). A revised conception of masculine-feminine differences in play activities. Journal of Genetic Psychology, 96, 165-170.
- Russell, R.W. (1984). Psychology in its world context. American Psychologist, 39, 1017-1025.
- Schaie, K.W. (1965). A general model for the study of developmental problems. Psychological Bulletin, 64, 92-107.
- Schaie, K.W., & Labouvie-Vief, G. (1974). Generational versus ontogenetic components of change in cognitive behavior: A fourteen-year cross-sequential study. Developmental Psychology, 10, 305-320.
- Schmid, J., & Leiman, J. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.
- Schmidt, F.L., & Hunter, J.E. (1977). Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 62, 529-540.
- Schmidt, F.L., Hunter, J.E., Pearlman, K., & Shane, G.S. (1979). Further tests of the Schmidt-Hunter Bayesian validity generalization model. Personnel Psychology, 32, 257-281.
- Scottish Council for Research in Education. (1949). The trend of Scottish intelligence. London: University of London Press.
- Scottish Council for Research in Education. (1953). Social implications of the 1947 Scottish mental survey. London: University of London Press.
- Scribner, S. (1974). Developmental aspects of categorizing recall in West African society. Cognitive Psychology, 6, 475-494.
- Scribner, S., & Cole, M. (1973). The cognitive consequences of formal and informal education. Science, 182, 553-559.

- Scribner, S., & Cole, M. (1978). Literacy without schooling: Testing for intellectual effects. Harvard Educational Review, 48, 448-461.
- Scribner, S., & Cole, M. (1981). The psychology of literacy. Cambridge, MA: Harvard University Press.
- Sharp, D., Cole, M., & Lave, C. (1979). Education and cognitive development: The evidence from experimental research. Monographs of the Society for Research in Child Development, 44 (1, 2, Serial No. 178).
- Spearman, C. (1904). "General intelligence" objectively determined and measured. American Journal of Psychology, 15, 201-293.
- Spearman, C. (1914). The theory of two factors. Psychological Review, 21, 101-115.
- Spearman, C. (1927). The abilities of man. New York: Macmillan.
- Spence, J.T. (1974). The Thematic Apperception Test and attitudes toward achievement in women: A new look at the motive to avoid success and a new method of measurement. Journal of Consulting and Clinical Psychology, 42 427-437.
- Thurstone, L.L. (1938). Primary mental abilities. Psychometric Monographs, No. 1.
- Thurstone, L.L. (1948). Psychological implications of factor analysis. American Psychologist, 3, 402-408.
- Thurstone, L.L., & Thurstone, T.G. (1941). Factorial studies of intelligence. Psychometric Monographs, No. 2.
- Thurstone, L.L., & Thurstone, T.G. (1946-1965). SRA Primary Mental Abilities. Chicago: Science Research Associates.
- Tuddenham, R.D. (1948). Soldier intelligence in World Wars I and II. American Psychologist, 3, 54-56.
- United States Department of Labor, Employment and Training Administration. (1977). Dictionary of Occupational Titles (4th ed.). Washington, DC: U.S. Government Printing Office.
- United States Department of Labor, Employment and Training Administration. (1983). Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery (USES Test Research Report No. 45). Washington, DC: U.S. Government Printing Office.
- United States Employment Service. (1946-1947). USES General Aptitude Test Battery. Washington, DC: U.S. Government Printing Office.
- Urberg, K.A. (1979). Sex role conceptualization in adolescents and adults. Developmental Psychology, 15, 90-92.

- Urberg, K.A., & Labouvie-Vief, G. (1976). Conceptualizations of sex roles: A life span developmental study. Developmental Psychology, 12, 15-23.
- Vernon, P.E. (1961). The structure of human abilities (Rev. ed.). London: Methuen.
- Wechsler, D. (1944). The measurement of adult intelligence (3rd ed.). Baltimore: Williams & Wilkins.
- Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale - revised. New York: Psychological Corporation.
- Wirtz, W. (Chair), (1977). On further examination: Report of the Advisory Panel on the Scholastic Aptitude Test score decline. New York: College Entrance Examination Board.
- Woodworth, R.S. (1910). Race differences in mental traits. Science, 31, 171-186.

רשימת מקורות נוספת

- Alderman, D.L. (1981). Measurement error and SAT score change. (College Board Report No. 81-9), (ETS RR No. 81-39) New York.
- Alderman, D.L. (1981). Student self-selection and test repetition. Educational and Psychological Measurement, 41, 1073.
- Anastasi, A. (1985). Some emerging trends in psychological measurement: The last fifty years. Applied Psychological Measurement, 9, 121-138.
- Anastasi, A. (1986). Evolving concepts of test validation. Annual Review of Psychology, 37, 1-15.
- Anastasi, A. (1986). Mental measurement: Some emerging trends. In D. Buros (Ed.), Mental measurement yearbook (9th ed). NJ: Gryphon Press.
- Angoff, W.H. (Ed.), (1971). The College Board Admissions Testing Program. New York: CEEB.
- Bersoff, N.D. (1981). Testing and the law. American Psychologist, 36, 1047-1056.
- Buros, D. (Ed.), (1978). Mental measurement yearbook (8th ed.). NJ: Gryphon Press.
- Gordon, W.E., & Tercei, D.M. (1981). The changed social context of testing. American Psychologist, 36, 1047-1056.
- Jensen, A. (1980). Bias in mental testing. New York: Free Press.
- Linn, R.L. (1982). Admissions testing on trial. American Psychologist, 37, 279-291.
- Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89, 191-216.
- Nevo, B. (1976). Fair tests (in Hebrew). Israeli Journal of Psychology and Counseling in Education (formerly Chavat Da'at), 6, 47-58.
- Nevo, B. (1985). Face validity revisited. Journal of Educational Measurement, 22, 287-293.
- Nevo, B., & Sfez, J. (1985). Examinees' Feedback Questionnaires. Assessment and Evaluation in Higher Education, 10, 236-249.
- Novick, R.M. (1981). Federal guidelines and professional standards. American Psychologist, 36, 1035-1046.

- Powers, D.E. (1980). Students' use of and reactions to alternative methods of preparing for the SAT. College Board Research and Development Reports, (RDR 80-81, No. 3).
- Powers, D.E., & Alderman, D.L. (1983). Effects of test familiarization on SAT performance. Journal of Educational Measurement, 20, 71.
- Stricker, L.J. (1984). Test disclosure and retest performance on the SAT. Applied Psychological Measurement, 8, 81-87.