

מכון הנרייטה סאלד
המכון הארצי למחקר במדעי ההתנהגות

תוקף במחקרי הערכה

Validity in Evaluation Studies

יצחק פרידמן

הרצאה במסגרת יום עיון במכון הנרייטה סאלד:

כיצד לתקף? סוגיות במדידה ובהערכה

ד' בתמוז תשס"ז (20.6.2007)

תוקף במחקרי הערכה

יצחק פרידמן

מושג התוקף (Validity) של כלי המדידה החל להתגבש לפני כ-65 שנים כמושג פרטני (פרטיקולארי) המתייחס להיבטים השונים של מאפייני כלי מדידה (מבחנים בתחום הפסיכולוגיה והחינוך בעיקר). המשגה פרטנית זאת הניבה הגדרות נפרדות וייחודיות לכל אחד מן ההיבטים של מאפייני כלי המדידה. הוגדרו סוגים רבים של תוקף כלי מדידה, כמו "תוקף נראה", "תוקף תוכן", "תוקף אמפירי", "תוקף גורמי", "תוקף ניבוי", "תוקף לוגי", "תוקף מושג-נבנה", "תוקף הגדרות" ועוד. תוקפם של כלי המדידה נבחן אפוא מזוויות ראייה שונות, ללא קשר חזק במיוחד ביניהן. אפשר היה למצוא, למשל, שתוקף התוכן של מבחן גבוה, אך התוקף האמפירי שלו נמוך, או שתוקף הניבוי שלו גבוה, אך תוקף המושג-הנבנה (Construct) שלו נמוך (או ההיפך, כמובן).

במשך השנים נתגבשה ההכרה שתוקף כלי המדידה הוא מושג יחידתי (Unitary) במובן זה שיש להתייחס אליו כאל יחידה אחת, שיש לה מספר שטחות. התפתחות נוספת שחלה במשך השנים בתפיסת מושג התוקף הייתה מעבר מהתמקדות בשאלת התוקף של כלי המדידה לתוקף של השימושים בתוצאות של כלי המדידה (ציונים, הערכות). לפי התפתחות זאת מתמקדת תשומת הלב במטרות שלשמן חובר כלי המדידה ומתייחסת לשאלת ה"אמת", שאפשר לייחס למסקנות שאפשר להסיק מן הממצאים הנובעים מכלי המדידה. רמה גבוהה (או נמוכה) של תוקף מדידה עשויה להיקבע כאשר כלי מדידה **מפיק מידע שעשוי להוביל להחלטות "נכונות", "מדויקות" ו"אמינות" (או לא).**

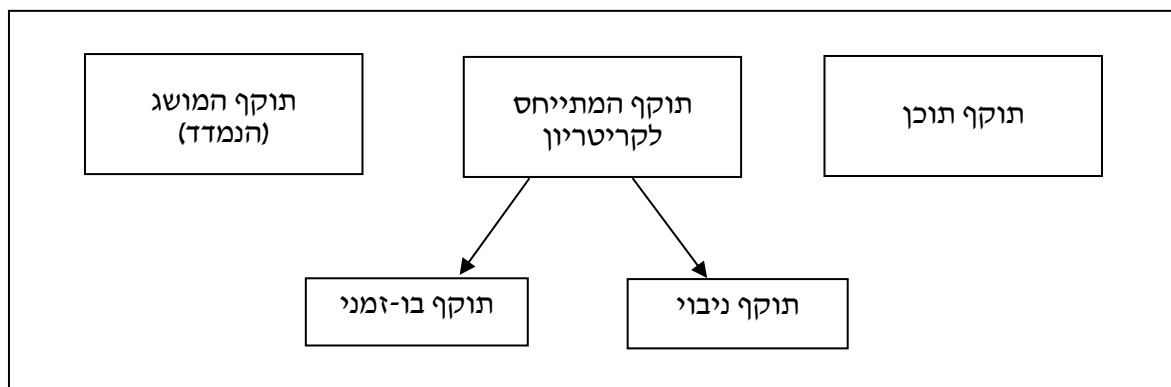
ההתפתחות השלישית שאפשר להבחין בה במשך השנים היא המעבר משיוך בלעדי של המושג תוקף לכלי מדידה, לשיוך המושג לכלי מדידה ולתהליכים כמו מחקרים או פעולות, שמיועדות לחשוף אמת אמפירית, ולבחון בעזרת המושג הזה את ערכה של האמת הנובעת מאותן פעולות.

המושג תוקף במדידה: הגדרותיו ומשמעותו

ניסיונות שיטתיים ראשונים להגדיר את מושג התוקף במדידה נרשמו במחצית שנות ה-40 של המאה הקודמת. גילפורד (Guilford, 1946) התייחס למושג התוקף של מבחנים, והציע ש"באופן כללי מבחן יכול להיחשב תקף לכל דבר שיש לו אתו מתאם" (עמ' 429). ואמנם חוקרים ותיאורטיקנים הציעו קשרים מתאמיים בין מבחנים ובין תופעות או תכונות מסוימות, וכך הגדירו סוגים שונים (ורבים) של תוקף. רק כעבור שנים לא מעטות היה זה קרונבך (Cronbach, 1980), שקבע כי "כל פעולות התיקוף הן למעשה דבר אחד" (ועל כן אין לדבר על "סוגים" שונים של תיקוף) (עמ' 99).

הדרך המקובלת לסיווג העדויות המשמשות לקביעת התוקף במדידה (של מבחנים בדרך כלל), מאז שנות ה-50 של המאה שעברה, היא לחלק אותן לשלושה סוגים מובחנים, או ליתר דיוק לשלושה סוגים, שאחד מהם נחלק לשניים, באופן הבא (ראו תרשים 2):

תרשים 2 סוגי התוקף



תוקף תוכן (Content Validity) מוערך על-ידי כך שבודקים באיזו מידה תוכן המבחן דוגם את תחום האוניברסל, המצבים או התכנים, שבנוגע להם מבקשים לקבל החלטות, או להסיק מסקנות. במלים אחרות, תוקף במדידה (מבחנים) מבוסס על שיפוטים מקצועיים על הרלבנטיות של תוכן המבחן לתוכנו של תחום התנהגות מסוים, שהוא נשוא התעניינותנו, ועל מידת הייצוגיות שבה פריט או משימה מסוימים מכסים את הנושא שבו יש לנו עניין.

תוקף שמתייחס לקריטריון (Criterion Related Validity) מוערך על-ידי השוואת ציוני המבחן עם משתנה אחד או יותר (הקריטריון), שנחשבים כמספקים מידה ישירה של מאפייני

ההתנהגות הנבחנת. תוקף המתייחס לקריטריון מבוסס על קשר אמפירי, שנמדד בדרך כלל במונחים של מתאמים ורגרסיות, בין ציוני המבחן ובין ציוני הקריטריון.

תוקף ניבוי (Predictive Validation) מראה את המידה שבה הרמה העתידה של תפקוד היחיד בתחום הקריטריון, או בהקשר של הקריטריון, מנובאות על-ידי ביצועיו של היחיד במבחנים קודמים.

תוקף בו-זמני (Concurrent Validation) מראה את המידה שבה ציוני המבחן מעריכים את רמת תיפקודו או מעמדו של היחיד בהקשר או בתחום של הקריטריון, בעת הנוכחית.

נדגיש כי תוקף שמתייחס לקריטריון מתבסס על תיאוריה רלבנטית ברורה, וגם על ממצאים אמפיריים הקשורים בתיאוריה זאת.

תוקף מושג-נבנה (Construct Validity) מוערך על-ידי הניסיון לחקור ולתת תשובות על השאלה: מהם המאפיינים או התכונות, שהמבחן מודד, כלומר על-ידי קביעה של הרמה שבה מושגים מסבירים (בלתי תלויים) מסוימים אחרים לביצוע במבחן. תוקף מושג נבנה מבוסס על שילוב של כל עדות הנשענת על הפרשנות או המשמעות של ציוני המבחן. הציונים הנובעים מכלי מדידה כלשהו אינם מושווים עם המושג שאותו הם מודדים, וגם אין הם מנסים להגדירו.

ללא קשר לסוג התיקוף או לתהליך התיקוף שנבחר לנקוט בו, חשוב להבין שתוקף הוא מושג מאוחד (Unitary Concept), במובן זה שהסוגים השונים של התוקף או של תהליכי התיקוף מהווים יחידה אחת, שיש לה מספר שטחות. נקודה זאת ראויה לציון משום שעד לפני כשני עשורים מקובל היה לדון ולדווח על התוקף של מבחנים במונחים של **סוגי** תוקף (תוקף תוכן, תוקף קריטריון ותוקף מושג-נבנה). גם חוקרים שלא קיבלו את תפיסת הסוגים השונים של התוקף נשענו על החלוקה המשולשת הנזכרת למעלה, משום שלפי גיזלי, קמפבל וצדק (Ghiselli, Campbell, & Zedeck, 1981), זה מה שאנשים עשו באופן מסורתי.

תוקף אינו מושג דיכוטומי וגם איננו מושג מוחלט (הכול או לא כלום), אלא הוא עניין של דרגה: התוקף של כלי מדידה, או של כל פעולה מחקרית אחרת, יכול להשתנות מרמה נמוכה ביותר ועד רמה גבוהה ביותר. התוקף הוא תכונה מתפתחת, והתיקוף הוא תהליך נמשך. כיוון שהעדויות, שמשמשות בסיס לתיקוף אינן תמיד מלאות או שלמות, התיקוף נשען על הסברים הגיוניים או על מקרים הגיוניים, שמאפשרים לכוון את השימוש במבחן או במחקר, כדי לקדם את הבנתנו בכל הקשור במשמעותם של ציוני המבחן.

ועדה, שמונתה על-ידי אגודת הפסיכולוגים האמריקאית לקבוע סטנדרטים למדידה חינוכית ופסיכולוגית (APA, 1954) הציעה ארבעה סוגים של תוקף: תוקף תוכן, תוקף ניבוי, תוקף בו-זמני ותוקף המושג הנבנה (Construct). אלה מוינו מחדש בשנת 1966 על-ידי ועדת סטנדרטים

אחרת של אגודת הפסיכולוגים האמריקאית, וסווגו לתוקף תוכן, לתוקף קריטריון ולתוקף המושג הנבנה (APA, 1966). סוגי תוקף אלה התייחסו למטרות המוגדרות האלה של המבחנים :

1. לקבוע כיצד היחיד מתפקד בהווה במכלול של מצבים או תכנים (תוקף תוכן)
2. לחזות את מצבו העתידי של היחיד, או לאמוד את מצבו הנוכחי לפי משתנה מסוים, השונה מזה הנבדק על-ידי המבחן (תוקף קריטריון)
3. להעריך באיזו מידה יש לנבדק תכונה מסוימת, סמויה בדרך כלל, שיש קושי למדוד אותה במישרין (תוקף מושג נבנה).

הצעות נוספות לסוגי תוקף היו תוקף נראה (face validity), תוקף גורמי (factorial validity) ותוקף אמפירי (Anastasi, 1954). סוגים אחרים היו: תוקף לוגי ותוקף ההגדרות. התוקף הנראה, המתייחס למה שנראה על פניו שהמבחן בודק, לא תמיד נתפס על-ידי החוקרים כתוקף של ממש. תוקף גורמי מתייחס למתאם שבין ציוני המבחן ובין גורם משותף לקבוצה של מבחנים או למדידה אחרת של התנהגות. תוקף מסוג זה מתייחס למבנה הפנימי של המבחן, שכיום מוגדר כתוקף המושג הנבנה. במסגרת תוקף המושג הנבנה נכללו השוואות המשתנה או המשתנים שהמבחן בודק עם גילאים שונים, מתאמים עם מבחנים אחרים וניתוחי גורמים, וכמו כן השוואות עם עקיבות פנימית והבדלים בציונים שנתקבלו לאחר טיפולים שונים. תוקף אמפירי מתייחס ליחס בין ציוני (תוצאות) המבחן וקריטריון, שהוא המשתנה של מה שהמבחן מיועד לבחון.

בשנים מאוחרות יותר הוסבה תשומת הלב ממושג התוקף (של הכלי) לתהליך התיקוף, והספרות החלה לעסוק בסוגים השונים של פרוצדורות תיקוף. במהדורה החמישית של ספרו, הציע קרונבך (Cronbach, 1990) לדבר על תיקוף (Validation) ולא על תוקף (Validity), והגדיר תיקוף כ"חקירה של איתנות הפרשנות הניתנת לציונים של מבחן" (עמ' 145). הוא הדגיש שלמבחן אחד ראוי לערוך שלושה סוגי חקירה משולבים לצורך הערכת טיבו של מבחן, והם: תיקוף המכוון לקריטריון, תיקוף התוכן ותיקוף המושג הנבנה.

תיקוף המכוון לקריטריון מתרגם ציון שהופק ממבחן להצהרה בדבר משתנה אחר, שאינו נמדד באותו מבחן. הצהרה זאת היא ניבוי של מה שאפשר לצפות לו, אם המשתנה (הקריטריון) היה נמדד או נצפה. ניבוי זה מתבסס על התנסות, וזאת יכולה להתבטא במחקרי מעקב, שמשווים ציוני מבחן עם ציונים על משתנים אחרים. היא יכולה להתבטא במחקרים עכשוויים, הנעשים במקביל (Concurrent), אולם גם סוג זה של מחקרים הוא בבחינת "ניבוי", אלא שהניבוי הוא "עכשווי" מבחינת הזמן.

תהליך תיקוף התוכן הוא בדיקה קפדנית של ייצוגם של תכני מבחן המתיימר לבחון כשיריות או ידע, כדי להעריך תכניות או טיפולים מסוגים שונים, ומידת התאמתם לתכנים של אותו תוכן או של אותן כשיריות. תיקוף המושג הנבנה (construct validation) מתייחס לניסיון לאתר את

ה"גורם" או את ה"מושג" שמונח ביסוד המבחן שניתן לנבדקים, כלומר לתהליך או לתהליכים המסבירים את התגובות (או הציונים) שנתקבלו במבחן. זהו ניסיון "להסביר" את ציוני המבחן. המילה construct נובעת מן המילה construe, שפירושה ניתוח, חשיפה והבנייה של הקשר בין דברים (או ניסיון להבין או להסביר את המשמעות, או את הכוונה של דברים, ביחס למצבים נתונים מסוימים). "מושג נבנה" הוא דרך או אמצעי להבניה או של ארגון דברים, במגמה להבינם ולהסבירם. תהליך התיקוף המשולש של בחינת השערות והפרכתן (תוכן, קריטריון ומושג נבנה) הוא ארוך, ועלול להמשך כמעט ללא סוף.

אנאסטאזי ואורבינה (Anastasi & Urbina, 1997), במהדורה השביעית של ספרן "מדידה פסיכולוגית" (Psychological Testing), הגדירו "תוקף" בפשטות רבה: "התוקף של מבחן מתייחס למה שהמבחן מודד ועד כמה הוא עושה זאת היטב" (עמ' 113), כלומר, התוקף מייחד אותנו באשר למה שניתן להסיק מציוני המבחן. אנאסטאזי ואורבינה, כמו קרוונבך, התייחסו לתהליך התיקוף ולא לתוקף, והדגישו כי באופן בסיסי כל הפרוצדורות לקביעת תוקף המבחן נוגעות ליחסים שבין הביצועים במבחן ובין עובדות נצפות, שאינן קשורות למבחן, הקשורות במאפייני ההתנהגות הנבדקים (עמ' 113).

אנאסטאזי ואורבינה הגדירו תיקוף תוכן כתיקוף תיאור-התוכן (content-description validation), שבו מתקיימת בדיקה שיטתית של תוכן המבחן כדי לקבוע האם מדגם מייצג של תחום התנהגויות מסוים אמנם נכלל ("מכוסה") במבחן. הן הבחינו בין "תוקף תוכן" לתוקף נראה (face validity), שאינו מתייחס למה שהמבחן באמת בודק, אלא למה שהדברים נראים באופן שטחי שהוא בודק. התוקף הנראה מתייחס אפוא לשאלה האם המבחן נראה תקף לנבחנים בו, לאנשים המשתמשים בו ולמתבוננים בלתי מקצועיים אחרים.

תהליך תיקוף המתייחס לקריטריון מוגדר אצל אנאסטאזי ואורבינה כתהליכי ניבוי קריטריון (criterion-prediction procedures), שנועד להראות את מידת האפקטיביות של מבחן בניבוי ביצועיו של הנבחן בפעולות מוגדרות. הביצועים הנבדקים יכולים להיות עתידיים (ניבוי) או בהווה (בו זמני).

תיקוף המושג הנבנה מוגדר על-ידי אנאסטאזי ואורבינה כמידה שבה המבחן מודד מושג (construct) או תכונה (trait) תיאורטיים. גם תהליך תיקוף זה יכול להתבסס על מתאמים עם מבחנים אחרים, שמשמעות המדידה שלהם מוכרת וידועה, על-ידי ניתוח גורמים ועל-ידי בדיקת העקיבות הפנימית. תהליך תיקוף מושג נבנה צריך לענות לשתי דרישות. האחת, הוא צריך להישען על בדיקת מתאמים עם משתנים אחרים, שעמם יש צידוק תיאורטי לכך שמתאמים כאלה אמנם יתקיימו. השנייה, תהליך התיקוף צריך לכלול גם בדיקה מבחינה, כלומר להראות שציוני המבחן הנבדק אינם נמצאים במתאם עם משתנים אחרים. שתי דרישות אלו נקראות, בהתאמה, תיקוף מתכנס (Convergent Validation) ותיקוף מבחן (Discriminant Validation)

(Campbell, 1960) (Campbell & Fiske, 1959). בשנים האחרונות התווספו לתהליך תיקוף המושג הנבנה גם שיטות מחקר מורכבות יותר, כמו מודלים של משוואות מבניות (Structural Equation Modeling), המבוססים על ניתוחי רגרסיה מרובה.

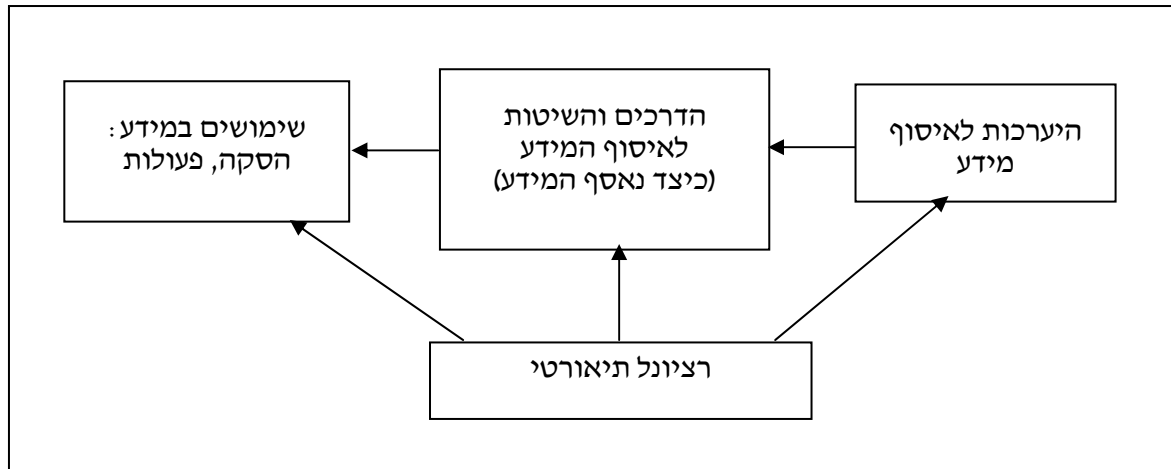
ועדה מיוחדת של אנשי מקצוע (APA, 1985), שגם היא כוועדות שקדמו לה, נטלה על עצמה משימה לפתח סטנדרטים למדידה חינוכית ופסיכולוגית, לא הגיעה להגדרה מוסכמת אחת של המושג. במקום זאת, ניסחה הוועדה את המאפיינים של המושג והציעה שתוקף "מתייחס למידה שבה העדויות, שאנו משתמשים בהן כדי להסיק מסקנות מחקריות ראויות, בעלות משמעות ושמישות" (עמ' 9). רוצה לומר, ראשית, שעדויות (מידע, נתונים), שנאספות במחקר (על-ידי כלים מסוימים, או על-ידי פעולת המחקר עצמו, הן אלה שתוקפן עומד לבחינה. שנית, הקריטריון להערכתה של רמת תוקף העדויות הוא המידה שהן ראויות, כלומר, מבוססות, אמינות, או נכונות מצד אחד, וניתן לעשות בהן שימוש ראוי.

כארבע שנים אחרי שוועדת הסטנדרטים של אגודת הפסיכולוגים האמריקאית (APA, 1985) פרסמה את מסקנותיה, פרסם סמואל מסיק (Messick, 1989) את מאמרו הידוע, שנחשב אבן דרך מרכזית בהבנת המושג תוקף. הוא הגדיר תוקף כ- **"שיפוט הערכתי משולב של המידה שבה עדות אמפירית ורציונל תיאורטי תומכים במידה שבה ההסקה והפעולות המבוססות על ציוני מבחנים או צורות אחרות של הערכה, הן ראויות, הולמות ומתאימות"** (שם, עמ' 13). המלים "ציוני מבחנים" מובאות בהגדרתו של מסיק באופן גנרי, כלומר במשמעותן הכללית ביותר, לא רק בהקשר של מבחנים, כפי שהדבר נעשה בדרך כלל, אלא מתייחסות לכל אמצעי של התבוננות, או תצפית ותיעוד של התנהגות או של ייחוסים. תוקף יכול אפוא להתייחס לכלי מדידה מסוים (שאלון, מבחן) או למחקר (study) שלם.

הגדרתו של מסיק נשענת אמנם על המאפיינים שנוסחו על-ידי ועדת הסטנדרטים (APA, 1985), אלא שהיא מרחיבה את הבסיס שעליו נשען המושג. מודגש בהגדרתו של מסיק כי תוקף מוערך על-ידי **שיפוט** הן של העדויות האמפיריות והן של הרציונל התיאורטי של ההסקה, שאליה מכוון הכלי או התהליך המיועד לאסוף עדויות. התוספת החשובה ביותר של מסיק, על המאפיינים שהוצעו לפניו, היא הצורך ברציונל תיאורטי, וכן טיבם של השימושים שיעשו במידע (ראו תרשים 1).

תרשים 1

מאפייני התוקף (Validity) במדידה, לפי מסיק (Messick, 1989)



המיוחד שבמאפיינים שהוצעו על-ידי ועדת הסטנדרטים ובהגדרתו של מסיק, בהשוואה למה שהיה ידוע בעבר, הוא שהם מתייחסים אל השימושים האפשריים במידע המתקבל מכלי הערכה או מדידה, כלומר אל ההסקה והפעולות, ולא אל כלי המדידה או כלי ההערכה עצמו. במלים אחרות, אם יש בידינו שיפוט הערכתי המאפשר לנו לחשוב שהמידע שהתקבל מכלי הערכה מסוים, או מפעולת הערכה מסוימת, הוא ראוי ("נכון", "טוב"), הולם ומתאים במידה רבה – נוכל לומר שכלי זה של מדידה או הערכה "תקף" במידה רבה. חשוב להדגיש כי כלי מדידה או הערכה (מחקר הערכה) יכול להיות בעל תוקף גבוה עבור שימוש מסוים אחד, ובעל תוקף נמוך עבור שימוש מסוים אחר.

חשיבותה של הגדרתו של מסיק והמשתמע ממנה היא בכך שבבואנו לבחון את מידת התוקף של כלי מדידה או של כלי הערכה מסוימים, נבקש לבחון גם את האופן שבו הופק המידע והופעלו הכלים, ולא נתמקד רק בכלי עצמו. כמו כן חשובה לא פחות, ואולי אפילו יותר, בדיקת ההתאמה של דרך הכנת הכלי והכלי עצמו עם מטרות השימוש בכלי ובמתקבל ממנו.

הגדרתו של מסיק מביאה תועלת רבה ברמה המושגית והתיאורטית, אך אינה ללא קשיים, אופרטיביים בעיקר. מסיק נמנע מלציין מה יכול להיחשב כעדויות "ראויות", "בעלות משמעות" ו"שמיות". מושגים אלה כשלעצמם דורשים הגדרות, גבולות ותנאים, כדי שניתן יהיה ליישם באופן ברור בהגדרת התוקף. אולם הדבר החשוב ביותר, הן בהצעותיהן של ועדות הסטנדרטים של אגודת הפסיכולוגים האמריקאית והן בהגדרתו של מסיק, הוא שהמושג תוקף, או פעולת התיקוף, אינה מתייחסת כל כך למדידה, אלא להסקה או להשלכות, או ליישומים המתבססים על הציונים המתקבלים מן המדידה. במלים אחרות, החוקר אינו "מתקף" את כלי המדידה, אלא את הפרשנות של הנתונים המתקבלים בתהליך מסוים (Cronbach, 1971). כיוון שכך, וכדי למנוע אי בהירות, ראוי להבהיר ולהגדיר מראש מהן מטרות המדידה, מהם השימושים הצפויים

בממצאים שיעלו ממנה, וכל זאת כדי לתחום את תנאי היסוד של המושג תוקף, כלומר את מידת היות ההסקות מממצאי המדידה "ראויות", "בעלות משמעות" ו"שימושיות".

הדברים שנאמרו עד כה מסכמים בקצרה את עיקרי התיאוריה ה"קלאסית" של מושג התוקף במדידה. אפשר אפוא לסכם ולומר כי הרעיון המרכזי במושג התוקף בשנים האחרונות הוא המידה הראויה, המשמעות והשמישות של ההסקה (המסקנות) והפעולות המבוססות על ציוני המבחן. כלי מחקר בעל תוקף גבוה הוא זה ש"תוצריו" הם בעלי משמעות, ושניתן להשתמש בהם בביטחון. כדי שדבר כזה יקרה, על המבחן לייצג היטב ובמדויק את התכנים שאותם הוא מבקש לבחון (תוקף תוכן), עליו להיות מסוגל "לקיים את הבטחתו" (ניבוי או התאמה – תוקף המתייחס לקריטריון) ועליו להתבסס על מבנה או על יסוד תיאורטי איתן (תוקף מושג נבנה). חוקרי תוקף המדידה הציעו דרכים לשכנע בדבר המידה שבה למבחן אמנם יש התכונות האלה, שמבטיחות שימוש, הסקת מסקנות וקיום פעולות בביטחון.

תוקף במחקר

הבה ננסה עתה להכליל את מושג התוקף, או התיקוף, מכלי מדידה למחקר. כשמדובר בתיקוף כלי מדידה אנחנו מתעניינים בשימושים, או במטרות, של כלי המדידה. המטרה של מחקר היא מתן תשובה לשאלה מוגדרת. במלים אחרות, ה"שימוש" או ה"מטרה" של מחקר גלומים בשאלת המחקר או נובעים ממנה. באופן כללי ביותר אפשר לומר שתוקף המחקר מתייחס למטרת (מטרות) המחקר או לשאלת (שאלות) המחקר. מחקר "תקף" הוא, כאמור, זה שממצאיו משקפים "אמת" אמפירית, הם ראויים, וניתן בהתבסס עליהם לקבל החלטות "נכונות".

כדי להבחין בסוגיית התוקף בין אסטרטגיות המחקר השונות יש שמגדירים את "ערך האמת" של המחקר בהתאם לאסטרטגיית המחקרית, כמוצג להלן (Newman & Benz, 1998):

- **"תוקף" (Validity)** הוא המונח שמשמשים בו לציון "ערך האמת" במחקר **כמותי** (Campbell & Stanley, 1963).
- **"אמינות", או "היותו ראוי לאמון" (trustworthiness)** הוא המונח שמשמשים בו לציון "ערך האמת" במחקר **איכותני** (Lincoln & Guba, 1985).
- **עקביות (קונסיסטנטיות) (Consistency)**, הוא המונח שמשמשים בו לציון "ערך האמת" במחקר **שיטות מעורבות** (mixed-methods research).

מה, אם כן, אפשר ללמוד ממאפייני התוקף במדידה (מבחנים), שיתאים להבנת המושג תוקף במחקר? אציע כאן 5 נקודות עיקריות.

1. מחקר בעל תוקף גבוה הוא זה ש"תוצריו" (ממצאיו) הם בעלי משמעות, ושניתן להסיק מהם מסקנות ולקבל בביטחון החלטות בדבר פעולות נדרשות.

2. כדי שמחקר יהיה תקף ברמה גבוהה, "תוכן" המחקר, כלומר שאלת (או שאלות) המחקר והנושאים שהוא מתמקד בהם, צריכים להיות מדגם מייצג של הנושאים, הסוגיות או השאלות, שבנוגע אליהן מבקשים לקבל החלטות או להסיק מסקנות. במלים אחרות, שאלת המחקר או מטרתו צריכים לייצג שאלות רחבות וכלליות יותר באותו תחום דעת או עניין.

3. שיטת המחקר (המתודולוגיה), צריכה להיות בהלימה עם השאלה המחקרית.

4. כלי המחקר צריכים להיות מכוונים לכך, שהם בודקים באופן אמין את הרכיבים והמשתנים הכלולים בשאלת המחקר או נובעים ממנה.

5. המחקר עונה ל-6 תנאי היסוד של חקירה מדעית, כמפורט להלן:
- א. תהליכים מאורגנים ושיטתיים, המבוצעים ברצף, כאשר שלבי הרצף עשויים להשתנות בפרדיגמות מחקר שונות
 - ב. תהליכי המחקר ותוצאותיהם ניתנים לבדיקה ולבחינה חיצונית
 - ג. אפשר לחזור על ממצאי המחקר (רפלקציה)
 - ד. קיים תהליך משוב פנימי, המאפשר תיקונים, כנדרש.
 - ה. המחקר מכוון להסביר תופעות, לפעמים תוך כדי חיפוש קשרים סיבתיים
 - ו. תהליך המחקר מאפשר לבטוח בתוצאות המושגות באמצעותו.

בהמשך הדברים אסקור את העקרונות והכללים, שבהם מקובל בספרות המתודולוגית להתייחס אל התוקף של מחקר. נסקור תחילה את הספרות (הותיקה במידה רבה) העוסקת במחקרים כמותיים ואיכותניים, ואחר כך אתייחס לסוגיית התוקף של מחקרים (חדשים יחסית) בשיטות מעורבות (Mixed-Methods Research).

תוקף במחקר כמותי

הדיון השיטתי בתוקף במחקר מתבסס בדרך כלל על עבודותיהם ומחשבותיהם של דונלד קמפבל ושותפיו (Campbell, 1957, 1969; Campbell & Stanley, 1963; Cook & Campbell, 1979). בעבודותיהם המוקדמות הבחינו קמפבל, וקמפבל וסטנלי (Campbell, 1957; Campbell & Stanley, 1963) בין שני סוגים של תוקף במחקר: **תוקף פנימי ותוקף חיצוני**. בשנים מאוחרות יותר הרחיבו קוק וקמפבל (Cook & Campbell, 1979) את סיווג התוקף, וכללו בו 4 סוגים:

1. **תוקף המסקנה, או ההסקה הסטטיסטית** (Statistical Conclusion Validity)

2. **תוקף המושג-הנבנה** (Construct Validity)

3. **תוקף פנימי**

4. **תוקף חיצוני**

ארחיב מעט את הדיבור על כל אחד מסוגי התוקף.

1. תוקף המסקנה (או ההסקה) הסטטיסטית

הערכת טיבו של מחקר, שעניינו חקר השתנות משותפת, נשענת על נתונים מדגמיים המצויים בידינו, שבהתבסס עליהם אנחנו מבקשים לקבל שלוש החלטות:

- (1) האם המחקר רגיש דיו כדי לאפשר מתן אמירות סבירות בדבר השתנות משותפת?
- (2) אם המחקר אמנם רגיש דיו, האם ישנן עדויות סבירות, שבהתבסס עליהן אפשר להסיק שהקשר הסיבתי המשוער אמנם קיים?
- (3) אם יש עדות כזאת, באיזו מידה של עצמה המשתנים הנחקרים משתנים במשותף?

קמפבל ועמיתיו ביססו את תוקף המסקנות הסטטיסטיות, או ההסקה הסטטיסטית, על מבחני מובהקות סטטיסטיים. תפיסת התוקף של מחקר הושתתה אפוא על סוגיות כמו גודל האפקט, עצמת (power) המבחן הסטטיסטי, טעות מסוג I ו-II ושאלת הקבלה או הדחייה של השערות האפס.

2. תוקף המושג-הנבנה

תוקף המושג הנבנה מתעניין במה שבמחקר ניסויי קרוי ערפול או ערבוב (Confounding) של מושגים. הוא מתייחס לאפשרות שהפעולות או המושגים שגורמים את יחסי הסיבתיות (קשר סיבתי בין משתנים), יכולים להיות מורכבים ממספר מושגים, שקשה להפריד ביניהם. הערפול (Confounding) פירושו שמה שחוקר אחד מפרש כיחסים סיבתיים בין מושגים תיאורטיים (A ו-B למשל), חוקר אחר עשוי לפרש כיחסים סיבתיים בין שני מושגים אחרים, נוספים ושונים (מ-A ו-B). אולם בקרב החוקרים בשנים האחרונות מקובלת ההבנה שתוקף המושג הנבנה במחקר קרוב יותר לתפיסת המשמעות של הסוג הזה של תוקף כמו בכלי מדידה, כלומר: המידה שבה מושג מסדר גבוה אמנם מיוצג במחקר מסוים (Johnson & Christensen, 2004). לפי הגדרה זאת, תוקף המושג הנבנה מתגבש על-ידי כך שהחוקר משתמש במחקר בהגדרות טובות, בהירות ומדויקות, של המושגים שנעשה בהם שימוש ובהסברים משכנעים בדבר המשמעות של המושגים האלה.

תוקף המושג הנבנה, בכלי מדידה כמו גם במחקר, מצביע על הקושי שבשימוש במושגים מסדר גבוה ("אינטליגנציה", "שחיקה", "אלימות", כדוגמה), משום שככל שהמושג מופשט יותר, כך יש לו רכיבים אפשריים רבים יותר, שלא קל לזהותם. דרך אחת להתגבר על קושי זה היא לייצג את המושגים במחקר באופן אופרציונלי. אופרציונליות פירושה שהמונחים או המושגים הכלולים במחקר מיוצגים בו על-ידי מקבץ ייחודי של צעדים ופעולות. ככל שנכללים במחקר היבטים רבים (ומשלימים) יותר של המושג הנחקר, כך גדלים הסיכויים שמושג זה יהיה מיוצג באופן מדויק יותר (תוקף המושג הנבנה יהיה גבוה יותר). השימוש בהיבטים, בהגדרות ובאמצעי מדידה רבים

יותר של מושג אחד, נקרא "אופרציונליזם מרובה" (Multiple Operationalism), ושימוש כזה מומלץ ביותר במחקר (Campbell, 1988).

3. תוקף פנימי (Internal Validity)

מתייחס למידה המשוערת שבה אנו יכולים להסיק מקשר אמפירי בין שני משתנים, שהקשר הזה הוא סיבתי, או שהיעדר קשר ביניהם רומז על היעדר גורם סיבתי בין שני המשתנים האלה. בניסוח מעט שונה: התוקף הפנימי הוא "תוקף מוערך, שבו אנחנו מסיקים שקשר בין שני משתנים הוא סיבתי" (Cook & Campbell, 1979, p.37).

בצורה הרחבה ביותר אפשר לנסח את התוקף הפנימי כך: אם מה שהתרחש (התופעה הנצפית) אמנם התרחש בגלל המשתנים שהחוקר סבור שהם אלה המשפיעים, או שאפשר לשייך את מה שהתרחש למשתנים אחרים? למעשה, התשובה לשאלה זאת תלויה בתשובות חלופיות, כלומר בהסברים חלופיים אפשריים. יוצא אפוא שככל שמחקר נעשה תחת בקרה טובה יותר (כמו למשל בניסוי מבוקר) כך הוא תקף יותר, כשכל התנאים האחרים נשארים ללא שינוי. חשיבותו של התוקף הפנימי רבה, משום שמטרה מרכזית ביותר במחקר (במדעי החברה במיוחד) היא לקבוע קשרים סיבתיים בין משתנים או מושגים.

4. תוקף חיצוני (External Validity)

"תוקף חיצוני" (External Validity) מתייחס להשערה, שלפיה הקשר הסיבתי שנמצא במחקר יכול להיות מוכלל (generalized) הן על משתנים אחרים הקשורים באותה סיבה ומסובב, והן בנוגע לאנשים אחרים, למצבים ולזמנים שונים. תוקף חיצוני הוא מושג שנטבע על-ידי קמפבל וסטנלי (Campbell & Stanley, 1963), והורחב על-ידי שדיש, קוק וקמפבל (Shadish, Cook & Campbell, 2002). תוקף חיצוני מתייחס להכללה של הממצאים ל – אוכלוסיית יעד או לרוחב (across) אוכלוסיות יעד, לזירות שונות, או לזמנים שונים.

"הכללה ל..." מתייחסת להכללה ממדגם או ממדגמים לאוכלוסייה, שהמדגם אמור לייצג אותה. כיוון שכך, ההכללה קשורה, או נעוצה, בתהליך בחירת המדגם. כיוון שקשה לאמוד טעויות דגימה, הרי קשה לאמוד את רמת התוקף החיצוני של מחקר, כשהכוונה היא הכללה לאוכלוסייה.

"הכללה לרוחב" (generalization across) מתייחסת להכללה החוצה אוכלוסיות. הכללה מעין זאת יכולה לכלול מחקר מדגמי מאוכלוסייה אחת, שיש כוונה להכליל את ממצאיו על אוכלוסייה אחרת, או מחקר שנעשה בהקשר אחד (בית ספר), שמבקשים להכליל את ממצאיו להקשר אחר (תעשייה).

הקשר בין ארבעת סוגי התוקף במחקר

קיים קשר בין ארבעת סוגי התוקף במחקר. הסקת מסקנות לא נכונות בדבר הקשר הסיבתי בין משתנים או מושגים במחקר היא לב לבו של התוקף הפנימי. הסקה כזאת (נכונה או מוטעית) נשענת היטב על היבטים סטטיסטיים (כמו חוסר מהימנות של מדידה, דגימה בלתי נכונה או לא יציבה, חוסר יציבות במדידות חוזרות וכיו"ב). כמו כן ייתכנו במחקר השפעות כמו הטיה (bias), המתייחסת לגורמים המשפיעים באופן שיטתי על ערכים של ממוצעי המשתנים במחקר, וכמו טעות (error) שמגדילה את ההשתנות ומפחיתה את הסיכוי לקבל אפקטים סטטיסטיים מובהקים. תוקף ההסקה הסטטיסטית יכול אפוא להיחשב כמקרה פרטי של התוקף הפנימי במחקר. הכללה היא לב ליבם הן של התוקף החיצוני והן של תוקף המושג הנבנה במחקר. תוקף גבוה של מושג נבנה מקל מאוד על יכולתנו להכליל את ממצאי המחקר (שנעשה במדגם) לאוכלוסייה הרחבה, או לאוכלוסיות אחרות, בזמנים ובהקשרים אחרים.

מושגים נוספים הקשורים בתוקף חיצוני

ספרות המחקר הענפה, שעניינה תוקף במחקר כוללת מושגים שונים הקשורים בתוקף חיצוני. אזכיר אותם כאן בקצרה.

תוקף אוכלוסייה (Population Validity). תוקף המתייחס ליכולת להכליל מתוך מדגם של יחידים, שעליהם נעשה המחקר, לאוכלוסיית יעד רחבה של אנשים (הכללה ל...), ולרוחב תת-אוכלוסיות שונות, במסגרת אוכלוסיית היעד (הכללה לרוחב). לתוקף אוכלוסייה יש אפוא שני רכיבים (האוכלוסייה הרחבה, ותת אוכלוסיות במסגרת האוכלוסייה הרחבה) (Bracht & Glass, 1968).

תוקף אקולוגי (Ecological Validity). מתייחס ליכולת להכליל את התוצאות של מחקר למצבים ולמערכות שונים. מחקר שנעשה באזורים כפריים, שהתבסס על כיתות קטנות, והעלה ממצאים מסוימים. האם אפשר להכליל תוצאות אלו לבתי ספר עירוניים, מבוססים, שבהם מספר תלמידים רב יותר בכל כיתה? תוקף אקולוגי אפשרי כאשר תוצאות מחקר אינן קשורות בסביבה או בתנאים סביבתיים מסוימים (Bracht & Glass, 1968).

תוקף זמנים (Temporal Validity). מתייחס למידה שבה תוצאות של מחקר ניתנות להכללה לזמנים שונים. נתונים נכונים לנקודת זמן מסוימת יכולים להיות, ויכולים לא להיות, נכונים לנקודת זמן אחרת.

תוקף השתנות הטיפול (Treatment Variation Validity). מתייחס ליכולת להכליל תוצאות לאורך (או לרוחב) השתנות בטיפול שניתן לנבדקים. דבר זה חשוב משום שמתן טיפול (הוראה, הדרכה), יכול להשתנות מזמן לזמן.

תוקף תוצאות (Outcome Validity). מתייחס ליכולת להכליל את התוצאות של מחקר על משתנים תלויים שונים, אך דומים.

תוקף המחקר במחקרים איכותניים

סוגיית התוקף במחקר טופלה במשך השנים כשהיא מתייחסת למחקרים הנשענים על איסוף נתונים בכלי מדידה כמותיים, ונעשים בהם עיבודים סטטיסטיים. בשנים האחרונות, עם התרחבות השימוש בשיטות מחקר איכותניות, התעוררה השאלה איך אפשר להתייחס לסוגיית התוקף (וסוגיית המהימנות) גם בנוגע למחקרים איכותניים. חוקרים המדברים על "תוקף" בהקשר של מחקרים איכותניים, גם הם מתייחסים למחקר אמין, שניתן לסמוך עליו, ועל כן הוא בר-הגנה (Smith, 1984).

אפשר לומר באופן כללי שהתוקף של מחקר איכותני הוא המידה שבה המידע המופק מן המחקר הוא אמין ומדויק, ושהפרשנות שמתבססת על הממצאים שלו היא אמנם "נכונה" ואמינה. אנשי המחקר האיכותני ממעטים לעסוק בהגדרות המושג תוקף, ומרבים לעסוק בדרכים לשמירה על תוקף המחקר ולהעלאתו. על אלה ארחיב את הדיבור בהמשך.

אפשר להציע דיון על היבטי תוקף כמו תוקף פנימי ותוקף חיצוני בהקשר של המחקר האיכותני. אולם בדרך כלל המחקר האיכותני מכוון **לתיאור** תופעות ולהסבירן, ולא לחיפוש קשרים סיבתיים, שהם המהות והתוכן של התוקף הפנימי. כאשר חוקרים במחקר איכותני מזהים יחסים אפשריים של סיבתיות, הם נדרשים לנקוט באסטרטגיות הדומות ביותר לאלו הנקטות במחקר הכמותי. כך הדבר גם בנוגע לתוקף החיצוני. בדרך כלל מחקר איכותני אינו רואה עצמו מכוון להכליל את ממצאיו לא אוכלוסייה הרחבה ולא לתת-אוכלוסיות.

תוקף המחקר במחקרים במתודולוגיה מעורבת

בשנים האחרונות גוברים הקולות השוללים את ההפרדה בין שימוש במחקר בשיטות כמותיות ובין שימוש במחקר בשיטות איכותניות, ומצדדים בשילוב של שתי האסטרטגיות המחקריות. קולות אלה מצביעים על האפשרות לנקוט באסטרטגיה שלישית, והיא המתודולוגיה המעורבת (Mixed-Methodology Research). מחקר במתודולוגיה מעורבת נוקט בשתי שיטות בסיסיות: מחקר **בשיטות** מעורבות ומחקר **במודל** מעורב. במחקר בשיטות מעורבות החוקר משתמש בפרדיגמות המחקר האיכותני בשלב אחד של המחקר ובפרדיגמת המחקר הכמותי בשלב אחר של אותו מחקר. במחקר במודל מעורב החוקר משתמש בשתי השיטות, כמותיות ואיכותניות במשולב, באותו שלב של המחקר. כיוון שכך, מתעוררת שאלת התוקף במחקרי המתודולוגיה המעורבת. כיצד נתייחס לתוקף של מחקר בשיטות מעורבות, או לחילופין, כיצד נתייחס לתהליך תיקופו של מחקר בשיטות מעורבות?

העיקרון הבסיסי של התוקף במחקר בשיטות מעורבות הוא עקביות (קונסיסטנטיות), ועקביות זאת צריכה להתבטא בכל רכיביו ושלביו של המחקר. כדי לקבוע את התהליך הנכון של תיקוף מחקר בשיטות מעורבות, או כדי להעריך את רמת התוקף שלו, מוצע לנקוט בשלושה שלבים אלה:

- א. קביעה והבהרה של מטרות המחקר ושאלות המחקר
- ב. קביעת השילוב הנכון של אסטרטגיות המחקר (כמותי, איכותני)
- ג. שימוש בסטנדרטים להשגת רמה גבוהה של עקביות במחקרי מתודולוגיה מעורבת.

א. קביעה והבהרה של מטרות המחקר ושאלות המחקר

הצעד הראשון בקביעת "ערך האמת", או פשוט ה"אמת", של מחקר, היא הבהרה וקביעה של מטרות המחקר ושאלות המחקר. מטרת המחקר היא הסיבה, או המניע, לכך שהמחקר נעשה, או הרציונל של המחקר. מטרת המחקר מנסה לזהות למה, בדיוק ככל האפשר, החוקר עושה את המחקר שלו. שאלת המחקר, או שאלות המחקר, מזהות ומגדירות את תחום החקירה ותחום העניין שבו עוסק המחקר: מה, בדיוק, מנסה החוקר למצוא? ישנם מחקרים שמתחילים בהגדרה ברורה ומפורטת של שאלת (או שאלות) המחקר, שעליה מנסה המחקר להשיב, וישנם מחקרים שיוצאים לדרך רק ברמז עמום בדבר שאלת המחקר.

מהן מטרות המחקר האפשרויות במדעי החברה והחינוך? להלן רשימה של 9 מטרות מחקר עיקריות (לפי: Newman, Ridenour, Newman & DeMarco, 2003):

1. **ניבוי** (לצורך יצירת חוקים כלליים).
2. **תוספת לידיע הקיים** (אימות ממצאים, רפליקציות, פירוש חוזר של ממצאי מחקרים קיימים, הבהרת מבנה וקשרים, חיזוק בסיס ידע).
3. **השפעה או תרומה מוסדית או אישית** (הפרכת תיאוריות או טיעונים, השפעה על יצירת שינוי, קידום שינוי, קביעת מדיניות, הצבת סדרי עדיפויות).
4. **מדידת שינוי או הבדלים** (מדידת השפעה של צרכי פעולה, בחינת השפעות של טיפול, מדידת תוצאות).
5. **הבנת תופעות מורכבות** (הבנת תופעות, הבנת תרבות, הבנת שינוי, הבנת אנשים).
6. **בחינה של רעיונות חדשים** (בחינת חידושים, בדיקת השערות, בדיקת רעיונות חדשים, בחינת פתרונות חדשים).
7. **יצירת רעיונות חדשים** (חקירת תופעות, יצירת השערות, יצירת תיאוריה, חשיפת קשרים, חשיפת תרבות).
8. **אספקת מידע** (מידע לקהל הרחב, העלאת מודעות, יחסי ציבור, תיאור מצב קיים).
9. **בחינת העבר** (פירוש, או פירוש חוזר, של העבר, בחינה מחודשת של הבנות ותיקות, חקר מקורותיהן של סוגיות חברתיות קיימות).

ב. קביעת השילוב הנכון של אסטרטגיות המחקר

ההחלטה על אסטרטגית המחקר קשורה בקשר אמיץ במטרת המחקר ובשאלה המחקרית. רק כאשר מטרות המחקר ושאלות המחקר מובהרות היטב, אפשר להצדיק את השימוש הראוי בשיטת המחקר, או בשילוב של שיטות המחקר. קיימים שלושה מודלים עיקריים לשילוב שיטות מחקר במתודולוגיה המעורבת: מודל לא אינטגרטיבי, מודל של תכנון מעורב סימולטאני ומודל הרצף האינטגרטיבי: כמותי-איכותני. המודלים מאורגנים על פני רצף מן הפחות חזק מושגית לחזק יותר מושגית, או במלים אחרות – מן ה"חלש" יותר מדעית ל"חזק" יותר מדעית.

המודל הלא אינטגרטיבי

בדפוס הלא אינטגרטיבי מבוצע בשלב ראשון מחקר איכותני, ואחריו – מחקר כמותי, או בסדר הפוך, מבלי שפרדיגמה אחת "מדווחת" לזולתה. שתי הפרדיגמות משמשות באופן בלתי תלוי, מבלי שנעשה ניסיון לקשור את שתיהן למטרה או למטרות משותפות.

מודל התכנון המעורב הסימולטאני

בדפוס זה של מחקר החוקר מנסה לעשות מחקר איכותני וכמותי באופן סימולטאני, כשלשני סוגי המחקרים ישנה אותה מטרה או מטרות. דפוס עבודה כזה יוצר בעיות אפיסטמולוגיות קשות, משום שהנחות היסוד של המחקר הכמותי והאיכותני שונות בתכלית. למשל, מחקר כמותי מניח סוג מסוים של "מציאות" אובייקטיבית, שממנה אפשר להכליל ממדגם לאוכלוסייה, באומדן כלשהו של ביטחון בעשייה כזאת. זאת בעוד שבמחקר איכותני אין "מציאות אובייקטיבית" מצויה, לפי ההנחה. המציאות היא ייחודית לכל יחיד, ואין אפשרות להכליל ממדגם לאוכלוסייה במידה סבירה של אומדן בטחון, וגם אין צורך או רצון לעשות כן. המחקר האיכותני והמחקר הכמותי אינם הולכים אפוא ביחד, והנחות מתנגשות אינן יכולות לחיות בכפיפה אחת.

מודל הרצף האינטראקטיבי: כמותי-איכותני

מודל זה שונה משני קודמיו במספר היבטים. שני המודלים הראשונים מתייחסים לשתי הפרדיגמות המחקריות באופן דיכוטומי, בעוד שהמודל השלישי דוחה דיכוטומיה זאת ונשען על רצף, שבו המחקר יכול להיות בעיקרו כמותי או בעיקרו איכותני.

בכל שלושת המודלים שיטות המחקר נובעות משאלות המחקר, שהן קשורות למטרות המחקר.

כדי לקבוע האם מחקר בשיטות מעורבות הוא האסטרטגיה הראויה ביותר, אפשר להציע, בהתבסס על עבודתה של מורס (Morse, 2003) 4 עקרונות תכנון, או סטנדרטים של תכנון, שראוי להתחשב בהם, ואלה הם:

1. **היכרות עם הבסיס או המניע התיאורטי של נושא המחקר.** אם מטרת המחקר, לשם הדוגמה, היא לתאר או לחשוף תופעה, למצוא משמעות או לחקור תופעה מוכרת, הרי הכיוון התיאורטי למחקר זה יהיה אינדוקטיבי. שיטת המחקר הראויה ביותר עשויה להיות במקרה כזה איכותנית, והתוצאה תהיה תיאור "עבה" של מידע פנומנולוגי או נרטיבי, יחד עם רמה

מסוימת של תיאוריה, שעשויה להתקבל ממידע נרטיבי או מתיאוריה מעוגנת שדה. אסטרטגיה איכותנית יכולה לשמש גם למטרות חקר ניסיוני, כשאינן עדיין הגדרה ממוקדת של יעדי המחקר. כשמטרת המחקר היא לאשש (השערות או תיאוריה), או לקבוע את התפלגותה של תופעה באוכלוסייה, הכיוון התיאורטי יהיה דדוקטיבי, והשיטה שראוי לנקוט בה היא בדרך כלל כמותית.

2. **הכרת ההיבטים השונים של נושא המחקר.** מחקרים רבים מורכבים הם וכוללים מטרות שונות ושאלות מחקר רבות. במחקרים "דלי" מטרות או שאלות, או במחקרים שבהם ישנה רק מטרה אחת או שאלת מחקר אחת בלבד, מטרה בודדת זאת מכתובה את הכיוון התיאורטי של המחקר. אולם במחקרים אחרים, שכיחים ביותר, יש להתבונן בהיבטים השונים של המחקר ולסווג את הכיוון התיאורטי שנובע מכל אחד מהם. זיהוי רכיבי המחקר והכרתם יסייעו לגבש את הכיוון התיאורטי או הכיוונים התיאורטיים של המחקר הנדון.

3. **הצמדות להנחות המתודולוגיות של שיטת (או שיטות) המחקר שנבחרו.** לדוגמה, כאשר נוקטים בשיטת מחקר איכותני, ישנם חוקרים שמתפתים "לספור", כלומר לבקש לדעת "כמה", דבר המעניק תחושה של דיוק ו"הליכה על קרקע בטוחה". אולם זהו פיתוי "מסוכן", משום שספירה אינה המטרה או הצד החזק של המחקר האיכותני. לא תמיד הנבדקים בראיונות נשאלים אותה שאלה, ועל כן אין זה נכון "לספור" את תשובותיהם של הנבדקים לשאלה מסוימת (שייתכן, כי בראיונות נוסחה באופן שונה מנבדק לנבדק). לעומת זאת, חוקרים "כמותיים" נחשפים להערות או למידע מילולי (שלפעמים נכתב בשולי כלי המדידה), ויש שמתפתים לעשות ניתוח תוכן של ההערות האלה, דבר שעלול להיות מטעה. הערות רבות מסוג זה מעידות יותר על חולשתו של כלי המדידה ואינן מספקות מידע חשוב אחר.

4. **יצירת השילוב ההולם ביותר (דפוס לא אינטגרטיבי, תכנון מעורב סימולטאני ורצף אינטראקטיבי).** ככל שהדבר אפשרי, יש לשלב את המידע מאסטרטגיות מחקריות שונות, ליחידת מידע אחת. צמצום מערכי המידע פירושו יצירת שילוב אינטגרטיבי של מידע כמותי ומידע איכותני ליחידת מידע מצומצמת אחת.

לוח 1 מתאר קשרים משולבים בין מטרות המחקר לאסטרטגיות המחקר (ראו לוח 1).

לוח 1

הקשר בין מטרת המחקר לאסטרטגיה המחקרית

מטרת המחקר	אסטרטגיה בגישת המחקר המסורתי	אסטרטגיה בגישת המחקר בשיטות מעורבות
1. ניבוי	מחקר כמותי ("השיטה המדעית", פוזטיביסטי)	מחקר כמותי
2. תוספת לידע	מחקר כמותי (הכללה)	מחקר כמותי ואיכותני
3. השפעה אישית ומוסדית	מחקר איכותני	מחקר איכותני בתוספת כמותי
4. מדידת שינוי	מחקר כמותי	מחקר כמותי
5. הבנת בעיות מורכבות	מחקר איכותני	איכותני בתוספת כמותי
6. בחינת רעיונות חדשים	מחקר איכותני	איכותני בתוספת כמותי
7. יצירת רעיונות חדשים	מחקר כמותי (בחינת השערות)	כמותי, בתוספת איכותני
8. מסירת מידע	מחקר כמותי או איכותני	מחקר מעורב
9. בחינת העבר	מחקר איכותני	מחקר איכותני

המקור: Newman, et al., (2003)

משהחליט החוקר על אסטרטגית המחקר (המודל) בשיטות המעורבות, עליו להביא בחשבון את הדברים האלה:

- הנחות היסוד ומגבלות התוקף של המחקר האיכותני והמחקר הכמותי
- האופן שבו הנחות היסוד של המחקר הכמותי והאיכותני משתלבים
- תרומתם של המחקר האיכותני והכמותי למחקר הנדון
- מטרות המחקר עצמו והקשרן לשילוב שיטות המחקר.

ג. שימוש בסטנדרטים להשגת רמה גבוהה של עקביות במחקרי מתודולוגיה משולבת

ריידנור וניומן (Ridenour & Newman, 2001) וכן ניומן ובנץ (Newman & Benz, 1998) הציעו

8 סטנדרטים לקביעת העקביות בתכנון מחקרים, ואלה הם (ראו לוח 2):

לוח 2

סטנדרטים ליצירת עקביות גבוהה במחקר

- כל מטרות המחקר צריכות להיות ברורות, מוסכמות ומנוסחות היטב
- כל שאלות המחקר צריכות להיות ברורות, מוסכמות ומנוסחות היטב
- עקביות צריכה להתקיים בין שאלת(ות) המחקר ומטרת(ות) המחקר
- עקביות צריכה להתקיים בין מטרת(ות) המחקר ובין שיטת(ות) המחקר
- שיטת(ות) המחקר צריכה(ות) להוכיח שהן מסוגלות להשיג את מטרות המחקר
- שיטת(ות) המחקר צריכה(ות) להוכיח שהן מסוגלות לטפל בשאלות המחקר
- ההנחות האפיסטמולוגיות צריכות להיות עקביות עם המטרה(ות), השאלה(ות) והשיטה(ות)
- שיטות המחקר צריכות להוכיח שהן הולמות את ההנחות האפיסטמולוגיות

סטנדרטים לפירוש ממצאי המחקר

- הממצאים צריכים להיות עקביים עם שיטות המחקר
- מטרות המחקר צריכות להיות עקביות עם המסקנות
- ההשלכות של ממצאי המחקר צריכות להיות עקביות עם מטרות המחקר.

דרכי השמירה על רמת גבוהה של תוקף המחקר

תוקף פנימי במחקר כמותי

ישנם שלושה סוגי עדויות הנדרשות כדי להגיע למסקנה בדבר קשר סיבתי, ואלה הם :

קיומו של מתאם

יש צורך בעדות על כך שהמשתנה התלוי והמשתנה הבלתי תלוי אמנם קשורים בקשר מתאמי כלשהו. במלים אחרות, יש צורך להיווכח בכך ששינויים במשתנה הבלתי תלוי מתייחסים לשינויים במשתנה התלוי. אם אין הדבר כך, כמובן שאין מה לחשוב על קשר סיבתי בין שני המשתנים.

סדר הזמנים

המקור השני של העדות הנדרשת כדי לשער קשר סיבתי הוא סדר הזמנים של התרחשות האירועים הקשורים במשתנים, זאת משום שהגורם המשפיע (הסיבה) צריך להתרחש או לפעול לפני שהשפעה ניכרת.

אי תלות במשתנה שלישי

הסוג השלישי, שנדרש כדי להגיע למסקנה בדבר סיבתיות, מתייחס לכך שהמשתנים הנחקרים קשורים אחד בשני בקשר סיבתי, ואינם קשורים במשתנה שלישי, חיצוני. קיומו של משתנה

שלישי, ששני המשתתפים (התלוי והבלתי תלוי) נמצאים אתו במתאם, יגרום לכך שהם יימצאו במתאם אחד עם השני, גם אם אין לכך כל קשר סיבתי.

הניסיונות לשמור על תוקף פנימי של מחקר כוללים תהליך דדוקטיבי, שבו החוקר הוא המבקר החריף והמעמיק ביותר של עצמו. להלן מספר "איומים" אפשריים על התוקף הפנימי של המחקר (כלומר על יכולתנו לקבוע באופן אמין את הימצאותם של קשרים סיבתיים בין משתנים או מושגים במחקר) (לפי Cook & Campbell, 1979; Maxwell & Delaney, 1990).

1. איום על האפשרות להסיק הסקות סטטיסטיות אמינות:
 - עצמה סטטיסטית נמוכה (של מבחנים סטטיסטיים), העשויה להביא להחלטה לא נכונה בדבר היעדר הבדל (טעות מסוג II), שנובעת מכך שהמדגם קטן מדי, או שרמת ה- α נקבעה נמוך מדי
 - הפרה של הנחות סטטיסטיות הקשורות במבחנים סטטיסטיים
 - ריבוי השוואות של הפרשי ממוצעים המעלה את ההסתברות שכל אחת מן ההשוואות תיחשב "מובהקת" בטעות (טעות מסוג I)
 - מהימנות נמוכה של הציונים שהופקו מכלי המדידה; מהימנות הטיפול שניתן לאנשים שונים על-ידי מטפלים שונים
2. **היסטוריה.** אירועים שהתרחשו בשעת ביצוע המחקר, שהיה בהם כדי לשנות את תנאי התכניות או התופעות הנבדקות, בהשוואה למצב ולהנחות היסוד בתחילת ביצוע המחקר. ככל שהמחקר נמשך זמן רב יותר, כך יש לרכיב ה"היסטוריה" השפעה רבה יותר על התוקף הפנימי של המחקר.
3. **זליגה של מידע או של טיפול.** דבר זה מתרחש כאשר אנשים בקבוצת הביקורת או בקבוצת הטיפול מקבלים מידע או ניסיון, שלא היו צריכים לקבל לפי התכנית ושלא היה מיועד להם. מידע שזלג באופן כזה עשוי להפחית באופן ניכר ביותר את תוקף המסקנות מן המחקר.
4. **הבשלה.** רכיב זה מתייחס לשינויים המתרחשים באנשים המשתתפים במחקר. הם מתבגרים, רוכשים ניסיון, מתעייפים, נעשים רעבים וכיו"ב. החשש הוא שתגובותיהם של האנשים (למידה, הניעה, תוקפנות, ריכוז), שאפשר לייחס אותן לטיפול (שנמצא במוקד המחקר), עלולות להיות קשורות במלואן או בחלקן בתהליכי ההבשלה האלה. אפשר גם שההבשלה תהיה באינטראקציה עם הטיפול עצמו, כלומר שהטיפול עצמו ישתנה, כדי להתאים עצמו להבשלתם של המשתתפים במחקר, מבלי שהחוקרים יבחינו בהשלכות הנובעות מכך.
5. **הצבה.** מתייחסת לתהליך שבו מציבים אנשים לקבוצות טיפול ולקבוצות ביקורת. כשאין נעשה שימוש בהצבה אקראית, קיים חשש של הרכבים לא שווים בין הקבוצות.

6. **היבחנות.** כאשר אנשים נמדדים כמה פעמים על אותו משתנה, הביצועים שלהם עשויים להיות מושפעים, בין השאר, מרכישת ניסיון, מזכירה של פריטים ותשובות קודמות, מהבנה של מטרות המחקר וציפיות החוקרים וכיו"ב.

7. **שימוש בכלי מחקר.** תיקוף פנימי עשוי להיפגע כאשר הבדלים בתוצאות של טיפולים שונים מיוחסים, במידה זאת או אחרת, להיבטים שונים של כלי המחקר. דבר זה עשוי לקרות כאשר סבורים ששני כלים הם שווי-ערך (או דומים במידה רבה), בעוד שאין הדבר כך. דוגמה נוספת היא כאשר הכלים באמת שווים, אך דרך העברתם משתנה (לפעמים מבלי שהחוקרים שמים לבם לכך).

8. **הגרסיה אל הממוצע.** חוקרים (Furby, 1973; Wolins, 1982; Roberts, 1980) דנו בכך שתופעה זאת של שינויים בממוצעי תופעה נחקרת "מתקדמים" במשך הזמן אל הערך הממוצע של ההשתנות שלהם. התופעה מתרחשת (במקרים רבים ביותר) בכל פעם ששני משתנים נמצאים במתאם חיובי, אך בלתי מושלם.

יש להדגיש שתוקף פנימי הוא תנאי הכרחי, אם כי לא מספיק, לתוקף חיצוני. כיוון שתיקוף חיצוני מחייב הסקה אינדוקטיבית, הוא בעייתי יותר מאשר תוקף פנימי, שבבסיסו הוא יותר דדוקטיבי (Cook & Campbell, 1979).

תוקף חיצוני במחקר כמותי

מספר היבטים קשורים בפגיעה אפשרית בתוקף החיצוני של מחקר, במיוחד כשמדובר בתוקף החוצה אוכלוסיות (הכללה לרוחב) (Pedhazur & Pedhazur-Schmelkin, 1991). נציין כאן כמה מהם:

מדגם מייצג. חוקרים צריכים להיות מודעים לכך שהם מבקשים להכליל את תוצאות מחקריהם לאוכלוסיות יעד של אנשים, להקשרים או לזמנים שונים. כאשר אוכלוסיות יעד כאלה מוגדרות, יש צורך שמדגמי המחקר יהיו "מייצגים" בהתאמה, כלומר יכילו את רכיבי המאפיינים של אוכלוסיות היעד המיועדות להכללה. אולם ישנם מקרים, שבהם אין לחוקר ידיעה ברורה בדבר אוכלוסיית היעד המיועדת להכללה של תוצאות המחקר. דבר מסוג זה עשוי להתרחש כשהמטרה היא לגבש תיאוריה כללית, וגם במחקר יישומי, כאשר מחברים מבחנים לדרגת כיתה מסוימת, מבלי להתכוון לאוכלוסייה, למקומות או לזמן מוגדרים וידועים מראש. אולם כאשר אוכלוסיית היעד מוגדרת נכון הדבר – אם ניתן – לשרטט מסגרת דגימה ולדגום מתוכה נבדקים, כך שהמדגם ייצג את האוכלוסייה ככל האפשר. תהליך דגימה כזה צריך להתחיל בדגימה מקרית, והנבדקים שייבחרו בדגימה כזאת יוצבו לקבוצות המחקר והביקורת באופן אקראי.

דגימה המכוונת להטרוגניות (Model of Deliberate Sampling for Heterogeneity). במודל זה יש עניין להגדיר קבוצות יעד של אנשים, הקשרים וזמנים, ולהבטיח שטווח רחב של מקרים בתוך כל קבוצה יהיו מיוצגים בתכנון המחקר. צורת דגימה כזאת אינה מחייבת דגימה מקרית בכל אחד משלבי הדגימה. לכן אין אפשרות להכליל את הממצאים שהתקבלו בדגימה כזאת לאוכלוסייה אחרת. אולם אפשר לתאר בהרחבה את הממצאים ולהדגיש שאין ודאות שממצאים אלה תקפים גם לאוכלוסיות אחרות, ואולי אפילו לא לאותה אוכלוסייה שממנה נלקח המדגם. דגימה מסוג זה, שלעיתים קרובות ביותר היא המעשית ביותר, מחייבת לדגום מקרים שהם שונים מאוד במאפייניהם אחד מן השני, כלומר הטרוגניים ביותר, במיוחד במאפיינים שבנוגע אליהם יש כוונה להכליל.

מודל השכיח ההתרשמותי (Impressionistic Modal Instance Model). לפי מודל זה החוקר מציין את סוג האנשים, ההקשרים, או הזמנים שבנוגע אליהם היה רוצה להכליל, ואז לבחור לפחות מקרה אחד מכל קבוצה שבאופן התרשמותי הוא דומה לשכיח הקבוצתי. השלב הראשון בשיטת דגימה זאת הוא הגדרה ופירוט של המאפיינים שיש בהם עניין, והשלב השני הוא לחפש ולאתר את המקרים השכיחים באוכלוסייה, לתאר אותם, ולבחור מספר מקרים בכל קבוצה.

אינטראקציה בין טיפול ומאפיינים (מין, דרגת השכלה, מאפייני אישיות). כאשר המטרה היא להכליל לרוחב רמות משתני-מאפיינים, יש צורך שמשנתנים אלה ייכללו במחקר. הכללות לרוחב רמות של מאפיינים אפשרית כאשר אין אינטראקציה בין סוגי הטיפול ובין המאפיינים.

אינטראקציה בין טיפולים למערכים שונים (בית ספר ותעשייה). מערכים מתייחסים לסביבות שבהן המחקרים נעשים. מערכים אלה יכולה להיות צרים (מעבדה) או רחבים (שדה). בדרך כלל מערך הטיפול מציב מגבלות על יכולת ההכללה של ממצאי מחקר (הכללה ממחקר שנעשה במעבדה למצבי שטח).

טיפולים מעורבים. אלה מתרחשים כאשר קבוצה אחת של אנשים נחשפת למספר טיפולים או למספר משתנים. כאן הטיפולים עצמם עשויים להפריע אחד לזולתו, רצף מתן הטיפולים עשוי להשתנות, ועמו גם האפקט.

איום מרכזי על התיקוף חיצוני הוא אפוא ההטרוגניות והייצוגיות של מדגם המחקר. מן המפורסמות היא שחלק לא מבוטל של מחקרים נעשה בהתבסס על מדגמים לא הסתברותיים, אלא על מדגמים שנבחרו מטעמי נוחות. באופן כזה ישנה סכנה מוחשית שהמחקרים שנעשו נשענו על מדגמים לא מייצגים את האוכלוסייה הראשונית, שהחוקרים כיוונו אליה, שלא לדבר על כך שאין כל ודאות שמדגם לא הסתברותי יצליח "לייצג" אוכלוסיות שהוא מתבקש לייצג.

תוקף המחקר האיכותני

כדי לשמור על רמת תוקף גבוהה פיתחו אנשי המחקר האיכותני במשך השנים מספר אסטרטגיות חשובות, שנועדו לשפר את איכותו ואמינותו של המחקר. להלן מספר אסטרטגיות לדוגמה (Maxwell, 1992, 1996; Johnson & Christensen, 2004):

טריאנגולציה (בדיקה צולבת של מידע ומסקנות בעזרת תהליכים או מקורות מרובים).

- **משוב שוטף של הנחקרים** שנקרא גם **תוקף הפרשנות (Interpretive Validity)**, המתייחס להצגה המדויקת של משמעות המיוחסת על ידי משתתפי המחקר (הנחקרים) לאירועים שבהם הם נוטלים חלק. סוג זה של תוקף מתייחס למידה שבה עולמם הפנימי של הנחקרים, כלומר מחשבותיהם, רגשותיהם, כוונותיהם וחוויותיהם הובנו כהלכה על-ידי החוקר האיכותני, והם מוצגים כראוי בדוח המחקר.

אסטרטגיה חשובה להשגת תוקף פרשנות גבוהה היא משוב משתתפים. באסטרטגיה זאת פרשנותו של החוקר את הממצאים וכן מסקנותיו, מוצגים בפני הנחקרים, לשם אימות והשגת תובנות. בכתיבת דוח המחקר נעשה שימוש בדיווחים הקרובים ככל האפשר לאופן שבו הוצגו הדברים על-ידי המשתתפים, תוך שימוש מזערי בהכללות. ציטוטים מדויקים של דברי הנחקרים הם דוגמה בולטת לצורה זאת של דיווח המובא בדוחות המחקר האיכותני.

- **שיתוף עמיתים**, שנקרא גם **תוקף התיאורים (Descriptive Validity)**, מתייחס לדיוק העובדתי של הדיווחים המתקבלים במחקר. אסטרטגיה יעילה, שנעשה בה שימוש כדי לחזק את התוקף התיאורי, היא טריאנגולציה של חוקרים. אסטרטגיה זאת נשענת על מספר חוקרים המתעדים ומתארים את התנהגותם של המשתתפים במחקר ואת ההקשר שבו מתרחשת התנהגות זאת. השתתפותם של חוקרים רבים מאפשרת בחינה ובדיקה צולבת של התצפיות, כדי להשיג הסכמה על הדברים שהתרחשו. כתוצאה מכך המחקר האיכותני אמין יותר ובר הגנה.

- **בדיקות חיזוניות** שנקרא גם **תוקף תיאורטי (Theoretical Validity)**, שמתייחס לדרגה שבה הסברים תיאורטיים, שנובעים ממחקר, מתאימים לנתונים, ועל כן הם אמינים ובני הגנה. כידוע, תיאוריה מתייחסת לדיון בדבר השאלה **כיצד** מתרחשת תופעה מסוימת, **ומדוע** היא מתרחשת באופן הזה? למחקר האיכותני ארבע אסטרטגיות המכוונות לחזק את התוקף התיאורטי של המחקר.

האסטרטגיה האחת היא **עבודת השדה הממושכת**, שלפיה החוקר מבלה זמן ניכר בשדה הנחקר ללימוד הנחקרים וההקשר שבו הם פועלים. באופן כזה ישנו בטחון רב יותר באשר להבנת אופן התרחשותן של התופעות הנחקרות ובאשר לסיבות, שהדברים מתרחשים.

אסטרטגיה שנייה, הקרויה **טריאנגולציה של תיאוריה** (Denzin, 1989), מכוונת לבחון תופעה מסוימת לאורך של מספר תיאוריות ונקודות מבט.

אסטרטגיה שלישית, הקרויה **התאמת דפוסים**, מכוונת לכך שעל בסיס התיאוריה שגיבש החוקר הוא מנסח מספר ניבויים ובודק את מידת הדיוק של הניבויים האלה. ניבויים אלה מגבשים דפוסים של תוצאות אפשריות, והנתונים שנאספו במחקר נבחנים לאורם של דפוסים התוצאות האלה.

האסטרטגיה הרביעית, שאינה ייחודית למחקר האיכותני, היא **בקרת עמיתים**. לפי אסטרטגיה זאת התיאוריה שגובשה נמסרת לעמיתים, לצורך חוות דעת, הערות וביקורת.

- **בחינה של מקרים הפוכים למקרים, שהם מטרות המחקר העיקרית**
- **רפלקסיביות** (מודעות עצמית ורפלקציה ביקורתית של החוקר על ההטיה האפשרית העשויה להשפיע על תוצאות המחקר)

דברי סיכום

בהרצאה זאת העליתי את נושא התוקף במדידה, במחקר בכלל ובמחקרי הערכה בפרט. הצגתי את סוגיית התוקף, המוכרת לחוקרים במדעי החברה, מתוך שתי זוויות ראייה. האחת – התפתחותית, והשנייה – מתייחסת למושג זה באופן שונה בהתאם לאסטרטגיה המחקרית הנבחרת. ההיבט ההתפתחותי של הדיון בסוגיית התוקף הראה שהתוקף הפך במשך השנים ממושג פרטני (שלפיו נרשמו סוגים שונים – נפרדים ובלתי תלויים) למושג יחידתי (שלפיו התוקף במחקר הוא מושג רב-שטחות, המהווה יחידה מורכבת אחת). התפתחות נוספת הייתה, שבעוד שבתחילה התמקדו החוקרים בתוקף כלי המחקר עצמו, וניסחו לכך הגדרות שונות, במשך השנים התייחסו החוקרים לשימושים ולתוצרים של כלי המחקר, וניסו להגדיר את "תוקף השימושים", "תוקף התועלת" ו"תוקף היישומים". לבסוף, העלתה סקירת התפתחות לימוד המושג שבעוד שבראשית לימוד סוגיית התוקף התייחסו החוקרים אליה כרלבנטית רק לכלי מדידה (מבחנים, שאלונים וכלי מחקר אחרים), בהמשך הדברים הורחבה ההתעניינות במושג זה גם למחקר ולתהליכי מחקר.

נמצא בספרות המחקר כי המושג תוקף (validity) הוא המונח שמשמש בו לציון "ערך האמת" במחקרים כמותיים, המושג "אמינות" (trustworthiness) משמש לציון "ערך האמת" במחקר האיכותני, ועקביות (consistency) משמש לקביעת "ערך האמת" בשיטות מעורבות. בעוד שהספרות הרבתה לעסוק ב"תוקף" וב"אמינות", הוצגו בהרצאה זאת סטנדרטים ליצירת עקביות גבוהה במחקר בכלל ובשיטות מעורבות בפרט.

מקורות

- Anastasi, A. (1954). *Psychological testing*. New York: Macmillan.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall.
- APA: American Psychological Association (1954). Technical recommendations for psychological tests and diagnostic techniques. *Psychological Bulletin*, 5 (2, pt. 2).
- APA: American Psychological Association (1966). Standards for educational and psychological tests and manuals. Washington D.C: Author.
- APA: American Psychological Association, American Educational Research Association & National Council on Measurement in Education (1985). *Standards educational and psychological testing*. Washington, D. C: American psychological Association.
- Bracht, G. H. & Glass, G. V. (1968). The external validity of experiments. *American Educational Research Journal*, 5, 437-474.
- Campbell, D. T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bulletin*, 54, 297-312.
- Campbell, D. T. (1960). Recommendations for APA test standarts regarding construct, trait, and discriminant validity. *American Psychologist*, 15, 546-553.
- Campbell, D. T. (1969). Prospective: artifact and control. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 351-382). New-York: Academic Press.
- Campbell, D. T. (1988). Definitional versus multiple operationism. In E. S. Overman (Ed.) *Methodology and epistemology for social science: selected papers* (pp. 31-36). Chicago University of Chicago Press.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. (pp. 171-246). Chicago: Rand McNally.
- Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation*. Boston: Mifflin Company.
- Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd Ed.), pp. 443-507. Washington, D. C: American Council on Education.
- Cronbach, L. J. (1980). Validity on parole: how can we go straight? New directions for testing and measurement: measuring achievement of a decade. *Proceedings of the 1979 ETS Invitational Conference* (pp. 99-108). San Francisco: Jossey Bass.
- Cronbach, L. J. (1990). *Essential of Psychological testing* (5th ed.). New-York: Harper & Row.
- Denzin, N. K. (1989). *The research act: a theoretical introduction to sociological methods*. Englewoods Cliffs, N. J: Prentice-Hall.
- Furby, L. (1973). Interpreting regression toward the mean in developmental research. *Developmental Psychology*, 8, 172-179.
- Ghiseli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco: Freeman and Company.
- Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439.

- Johnson, B., & Christensen, L. (2004). *Educational research*. Boston: Pearson.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational review* 62(3). 279-299.
- Maxwell, J. A. (1996). *Qualitative research design*. Newbury Park, CA: Sage.
- Maxwell, S. E., & Delaney, H. D. (1990). *Designing experiments and analysis data*. Pacific-Grove, CA: Brooks/Cole Publishing Company.
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). pp.13-103, New-York: Macmillan Publishing company.
- Morse, J. M. (2003). Principles of mixed methods and multimethod research design. In A. Tashakkori & C. Teddlie (Eds.). *Handbook of mixed methods research in social & behavioral research* (pp. 189-208). Thousand Oaks, CA: Sage.
- Newman, I., & Benz, C. R. (1998). *Qualitative-quantitative research: exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.
- Newman, I, Ridenour, C. S., Newman, C., & DeMarco, G. M. P. Jr. (2003). Typology of purpose: a tool for thinking about research questions, designs, and mixed methods. In A. Tashakkori, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research*. (pp. 167-188). Thousand Oaks, CA: Sage.
- Pedhazur, E. J. & Pedhazur-Schmelkin, L. (1991). *Measurement design, and analysis*. Hillsdale, N. J: Lawrence Erlbaum Associates, Publishers.
- Ridenour, C. S., & Newman, I. (2001 June). *Themes are not variables, and mixed methods are not a "panacea" for educational researchers*. Paper presented at the

annual meeting of the Midwestern Educational Research Association, Columbus, OH.

Roberts, A. O. H. (1980). Regression toward the mean and the regression-effect bias. In G. Echternacht (Ed.), *Measurement aspects of Title I evaluation* (pp. 59-82). San Francisco: Jossey-Bass.

Shadish, W. R., Cook, T. D, & Campbell D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Smith, J. K. (1984). The problem of criteria for judging interpretive inquiry and other case studies. In L. Shuman (Ed.), *review of research in education*, Itasca, IL: Peacock. Vol. 6. 316-377.

Wolins, L. (1982). *Research mistakes in the social and behavioral sciences*. Ames, IA: Iowa State University Press.