



שיעור מספר 5 – דיוק בהערכה

המשך תבניות הערכה

בשיעור קודם

הדגמת חשיבות הדיוק בהערכות:

חשיפת ליקויים בתפקוד העובדים בכל רמות הארגון הגנה על הארגון מפני תביעות משפטיות.

דיוק של הערכה נמדד באמצעות המדדים הבאים:

(א) מדדים עקיפים של דיוק:

קריטריונים פסיכומטריים מסורתיים כמהימנות ותוקף. מהימנות ותוקף נמוכים מעידים על חוסר דיוק בהערכה. מדדים לטעויות (שגיאות) מעריך המשקפים הטיות תגובה מצדו. העדר טעויות מעריך מצביע על דיוק רב יותר של ההערכה. ההתפלגות האמיתית אינה ידועה – ולכן משתמשים בהתפלגות משוערת. סטייה מהתפלגות משוערת תעיד על טעות.

(ב) מדדים ישירים של דיוק בהערכה (רק במעבדות)

נמדדים על ידי השוואה בין תצפית לבין הערכת המעריכים כאשר לפעמים התצפית היא על שחקנים (ואז ידוע מה ההתנהגות המדויקת)

מהימנות

מהימנות (reliability) היא מדידה מדויקת, ללא טעויות, כאשר הערכים המתקבלים במדידה (בגיליון הערכת תפקוד) זהים לערכי האמיתיים של המשתנה (תפקוד)

את הערכים האמיתיים של המשתנה איננו יכולים למדוד ישירות, אך אנו יכולים לאמוד אותם על פי הערכים המתקבלים במדידה (=ערכים נצפים) המורכבים מ: המשתנה האמיתי וממשתנה הטעות.

מהימנות היא גם מידת היציבות בתוצאות מדידה המתקבלות בבדיקות חוזרות ונישנות, על אותם מוערכים, בתנאי מדידה דומים ככל האפשר.

גורמים לפגיעה במהימנות

גורמי מזל או גורל- משתנים הפועלים באופן אקראי ללא חוקיות, בדרך כלל חד פעמיים ואינם ניתנים לצפייה מראש או למניעה.

גורמים אישיים זמניים- שהשפיעו על תפקוד הפרט בעת שהוערך אך אינם משפיעים בדר"כ על פעילותו היומיומית. ייצוגיות הדגימה- כמות התצפיות על תפקוד העובד קטנה מכדי למצות באורח מייצג את כל תחום השתנותם.

מהימנות ועקיבות פנימית (Reliability and internal consistency) מבחן הוא מדגם של פריטים שכל אחד מהם וכולם יחד מיועדים למדוד את אותו משתנה. מבחן מהימן יותר ככל שהמתאמים בין פריטיו השונים גבוהים יותר. לא עובד ב GRS כי יש רק ייצוג אחד לכל תכונה נמדדת.

מהימנות בין מעריכים (Interrater agreement)- זהו מדד למהימנות המבוסס על מתאם בין הערכותיהם של שני מעריכים (או יותר) את אותו מוערך.

מהימנות כאקוויוולנטיות – (Equivalence of alternate forms) מקדם המהימנות מבוסס על מתאם בין שני נוסחים מקבילים (אקוויוולנטיים) של אותו מבחן. השקילות מתבטאת בכך שהנוסחים מתייחסים אל אותו מושא הערכה. לא כל כך ישים בתחום שלנו. לפעמים במקום זה משווים בין שתי תבניות דומות כמו BARS ו-BOS. **מבחן חוזר** (test-retest) – שיטה לבחינת מהימנות בה מבססים את מקדם המהימנות על חישוב המתאם שבין שתי העברות של אותו מבחן. בעייתיות – בפרק זמן ארוך, יכול להיות שינוי אמיתי, בפרק זמן קצר – המעריך זוכר מה הוא ענה.

- לסיכום - קיים קושי להחליט באיזו הגדרת מהימנות להשתמש כדי לבחון דיוק הערכות תפקוד:
- מהימנות כעקיבות פנימית- מודל זה אינו ישים לתבניות הטיפוסיות של הערכת תפקוד.
- מהימנות בין מעריכים- מעריכים שונים צופים בהיבטים שונים של תפקוד אותו עובד ולעיתים בכנות לא מסכימים בהערכותיהם.
- מהימנות כאקוויוולנטיות- קיים קושי לפתח שני גיליונות הערכה מקבילים.
- מבחן חוזר- במודל זה קיימת בעיה מושגית של בלבול תוצאות עקב טעות המדידה עם תוצאות עקב שינויים אמיתיים בתפקוד. לכן בפועל אין היום כמעט שימוש במדידת מהימנות

סוגי תוקף

- תוקף תלוי קריטריון** (Criterion-related validity) בוחן את הקשר (המתאם) בין ציוני מכשיר המדידה המנבא לבין ציון במשתנה אחר-קריטריון. יש בעיה בתחום זה שהקריטריון קשה למדידה, לכן לא נעשה כל כך שימוש בתוקף זה.
- תוקף מבנה** – (Construct Validity) בוחן הלימה בין משתנה תצפיתי מסוים לבין משתנה תיאורטי. נחלק לשתי תת בדיקות:
- תוקף מתכנס** – (Convergent Validity) בודק באיזו מידה המשתנה התצפיתי ממצה את מרכיבי המשתנה התיאורטי.
- תוקף מבחין** – (Discriminant Validity) בודק באיזו מידה המשתנה התצפיתי בודק רק את מרכיבי המשתנה התיאורטי.

תוקף מתכנס

באיזו מידה המשתנה ממצה את מרכיבי המשתנה התיאורטי. בתחום שלנו זה יבדק על ידי השוואה למספר מדדי תפקוד שונים – הערכות תפקוד, GRS,

תוקף מבחין

באיזה מידה המשתנה התצפיתי בודק רק את המשתנה התיאורטי ולא משתנה אחר. בתחום שלנו זה יבדק על ידי השוואה למשתנים קרובים שיתכן בלבול איתם ונצפה למתאמים נמוכים.

קיים קושי לקבוע תוקף הערכות תפקוד:

הקריטריון שכנגדו ציוני ההערכה מוערכים הוא לעיתים קרובות בעייתי, מבנה התפקוד לא מתואר בפירוט ונדיר למצוא מחקר אמפירי בנושא.

יחד עם זאת, קיימות עדויות מחקריות עקיפות לתוקף מבנה של הערכות תפקוד, לדוגמא:

- הסכמה בין הערכות ממקורות שונים בהתייחס לתפקוד היחסי של מוערכים שונים. מקורות שונים – ממונים, עמיתים, לקוחות.
- השפעה עקבית וחיובית של משוב על תפקוד העובדים. אם המשוב נעשה היטב אנחנו אמורים לראות שיפור עקבי בביצועים.
- מחקרים המשתמשים בהערכות הממונים כקריטריון לתיקוף מבחני בחירה.

שגיאות מעריך

טעויות התפלגות

טעויות התפלגות – (Distributional errors) נחשבות כקיימות כאשר התפלגות הערכות התפקוד שונה מההתפלגות (המשוערת) של תפקוד בתפקיד. בדרך כלל נשער שההתפלגות היא נורמלית (עקומת גאוס).

בית הלל - (Leniency) מתבטאת בשימוש בערכים מהטווח שמעל לנקודת האמצע בתבנית (סולם) ההערכה .
בית שמאי - (Strictness) מתבטאת בשימוש בערכים מהטווח שמתחת לנקודת האמצע בתבנית (סולם) ההערכה .
הנטייה למרכז - (Central tendency) נטייה קבועה אצל מעריך להימנע ממתן הערכות קיצוניות. הערכותיו נוטות להתקבץ סביב מרכז התבנית (סולם) - נקבל עקומה נורמלית תלולה יותר
צמצום טווח (Reducing range) - מתבטאת במתן הערכות מאוד דומות זו לזו עם השתנות מועטה בין ההערכות.

בעיה אופיינית (ראה כתבה על חברת חשמל בגלובס בתאריך 25-12-2003)
 מתן הערכות גבוהות באופן עקבי כדי לשמור על שלום בית

בעיות בהסקת דיוק הערכה מטעויות (שגיאות) מעריך:

- ההתפלגות האמיתית של תפקודים כמעט תמיד לא ידועה. לכן, אין הצדקה אמפירית להנחה שההתפלגות נורמלית ומרוכזת סביב נקודת האמצע של התבנית ושכל הקבוצות מתפקדות באותה רמה. יתרה מכך – ארגונים משקיעים זמן רב ומאמצים בשיפור העובדים לכן אין לצפות שהממוצע בתפקוד יהיה זה ממוצע התבנית. בנוסף – לא ניתן להניח שכל הקבוצות הן באותה רמה.
- הערכות בעלות טעויות (שגיאות) מעריך עדיין יכולות להיות מדויקות והערכות ללא טעויות יכולות בכל זאת להיות לא מדויקות.

טעויות מתאם

טעויות מתאם (Correlational errors) - נתמקד בטעות ההילה (Halo error).

טעות ההילה

הילה (Halo error) מתוארת בדרך כלל במונחים של
 (א) נטיית המעריך לתת להערכה כוללת (גלובלית), "לצבוע" הערכות תפקוד לפי מימדים ספציפיים של תפקוד
 או

(ב) חוסר רצון המעריך להבחין בין היבטים שונים בתפקוד המוערך. (Saal et al, 1980)

איך נראית טעויות הילה: נראת כמו מקבץ של ציונים במימדים שונים באותו אזור (גבוהים, נמוכים ובינוניים)
 מצד שני, תופעה כזו לא בהכרח מעידה על טעות הילה, יכול להיות קשר אמיתי בין הציונים (כמו למשל מוטיבציה, שמשפיעה על כלל הקריטריונים התפקודיים)

פרדוקס הילה-טעות

פרדוקס הינו סידרה של טענות, שמוכיחה כי ידיעותיו של האדם סותרות זו את זו .
 פרדוקס הילה-טעות כלומר, טעות הילה גבוהה יותר קשורה עם הערכות מדויקות יותר .
 הפרדוקס הילה-דיוק עלול לנבוע מהגדרות אופרציונליות בעייתיות הן של טעות ההילה והן של דיוק .

כדי להסביר את המתאם הגבוה בין ממדים שונים של הערכת תפקוד, קופר (Cooper, 1981b) חילק את ההילה לשני מונחים:
 הילה אמיתית (= true halo) ההתנהגויות שדורגו, למרות שהן נבדלות קונספטואלית, אכן במציאות מתואמות זו עם זו.
 הילה כוזבת (= illusory halo) = הממדים מתואמים זה עם זה בגלל עיוותים קוגניטיביים של המעריך.

האם תבנית מדויקת היא בהכרח כדאית יותר?

כעיקרון, תבנית מדויקת היא אפקטיבית יותר.
 אבל יש שני פרמטרים נוספים המשפיעים על כדאיות תבנית:
תגובות המעריכים והמוערכים לתהליך ההערכה (עמדות מעריכים ומוערכים + שינוי תפקוד בפועל)
 האם התקשורת טובה

האם נבדקו הדברים החשובים
האם יש שינוי של תפקוד בפועל
היעילות הכלכלית - אם התועלת מהשימוש בתבנית גבוהה מהעלות, המערכת תחשב כיעילה כלכלית

4 מתוך 4

נכתב על ידי גל

על בסיס תוכן ההרצאות במדעי ההתנהגות – מכללת נתניה
<http://behavioral-sciences-nac.blogspot.com>