

WHEN REASONABLE PEOPLE DISAGREE : IMPROVING FEEDBACK SCALE PERFORMANCE

Winter, 2003

WHAT MAKES A "GOOD" RATING FORM?

One of the cornerstones of evaluation in medical education is the use of direct observation as a formative and summative measure of complex performances and competencies. To make direct observations an effective and consistent tool for providing feedback it is important to use well developed rating forms. Forms that function well are said to have a higher degree of validity and reliability. Let's explore these two important concepts and how they play out in the world of observations.



Validity is an indicator of the extent to which an instrument is actually measuring what it is designed to measure. Face validity (a.k.a. content validity) employs a panel of experts to look over an instrument and is used as an initial indicator during development. For rating scales, one of the standard reliability indicators is a correlation between ratings and an established external performance indicator, such as grades or board examinations. This type of "criterion referenced" validity is also expressed as decimal coefficient, where bigger is better.

Reliability is the degree of stability, repeatability or consistency a measure exhibits. For a rating scale, one of the key functions is the degree of agreement between multiple raters observing the same thing, known as Inter-rater reliability. This agreement is expressed as a correlation coefficient. For longer rating instruments, the effect of length can be measured using an internal consistency indicator like the split-half method.

Finally, consider the effect of Differential Validity, which is the potential for members of certain groups to be measured differently. Internal consistency measures aggregated by race and ethnicity can be indicators of this threat in addition to careful analysis of item response patterns.

WHY ANCHORS MATTER

Anchor points in a scale are the definitions placed on various levels of performance. For example, a clear definition of an acceptable and unacceptable level of culturally appropriate communication skills might be used in a scale to provide feedback for patient encounters. Taking the time to develop a rich, descriptive language around these anchor points will go far in supporting reliability and

validity in a measure. Good anchors are non-ambiguous statements that evaluators can generally describe in terms of prototypical examples and defining attributes. Developing the language for anchors is generally a team effort.

This document is about:

- Improving existing feedback measures
- Improving the process of implementing feedback measures
- Documenting the characteristics of feedback measures

GME Can provide help and support for these projects