Factors that might undermine the validity of patient and multi-source feedback.


Author Institutions: Peninsula College of Medicine & Dentistry, University of Plymouth, Drake Circus, Plymouth, UK.

Tags

**Clinical domain**
- Professional
- General

**Educational domain**
- Assessment and Evaluation

Background

Multi-source feedback (MSF; aka 360') is an approach to assessment borrowed from the business world and increasingly popular in medicine. However, MSF has been criticized with respect to several key issues, including:
- ability to discriminate between poor & acceptable performance
- impact of allowing assesees to choose their favourite assors (Ramsey et al's work says this is fine)
- the role of assessor characteristics impacting scores
- the utility of MSF vs direct patient ratings

Purpose

Archer & McAvoy from the UK conducted a study of a cohort of MDs referred for further assessment due to suspicion of dyscompetence in order to answer the following questions...
- 1. Can MSF or patient ratings identify poor performers vs a reference group?
- 2. Does the selection of assessors impact scores?
- 3. What characteristics of assessors account for differences in scores?

Type of paper

Research: Observational
**Key Points on the Methods**

The authors, who are involved in a national physician assessment program in the UK, assembled a cohort of 68 MDs referred for assessment of poor performance and compared them to a national reference cohort of 123 MDs. Using standardized MSF (SPRAT) & patient rating (SHEFFPAT) instruments, they then compared the ratings by assessor chosen by the MD vs those assigned by an institutional body. Scores on MSF vs patient ratings were compared. They also analyzed the effect of rater characteristics (age, gender, seniority, etc) on scores.

**Key Outcomes**

The MSF instrument was able to distinguish the "cohort of concern" from the general reference group. Mean scores were lower (4.22 vs 5.27 / 6), as were the % 1 or 2 SDs below a national mean (90%, 68%). 28% had a mean score below "satisfactory". 76% had at least 1 comment on health or integrity impacting practice vs 6% in the national cohort.

Assessors chosen by the MD were significantly more lenient and less discriminating.

Assessor characteristics only accounted for 3.5% of the variance.

Patient ratings did not discriminate among performance ratings and were lenient.

**Key Conclusions**

The authors conclude that the SPRAT MSF tool is useful in identifying MDs potentially performing poorly. Importantly, raters chosen by the MD and patients were both more lenient and less discriminating.

This is an interesting study that challenges some previous findings about MSF.

It is also a good paper to use as a springboard to discuss threats to validity and reliability, with multiple issues arising from the methods and constructs.

Perhaps we should all be using institutionally-assigned raters and MSF instruments more often to assess MD competence.

**Spare Keys – other take home points for clinician educators**