

Faculty Teaching Scores: Validating Evaluations, Evaluating Validation

Armin Schubert, MD, MBA

There are just under 140 Accreditation Council for Graduate Medical Education accredited anesthesiology training programs in the United States. In 2007, these programs encompassed a combined total of 5472 residents.¹ Accreditation Council for Graduate Medical Education requires at least an annual evaluation of faculty by residents. In many programs, each faculty member is evaluated more than once each year. Academic anesthesia departments have an average of 44 faculty members.² Even if faculty were evaluated only once annually by their residents, this would amount to more than 33 million evaluations by residents for 6000 anesthesiology faculty in the United States alone and many more around the world. Despite this enormous effort, anesthesia educators still do not enjoy the benefits of a validated format for faculty assessment by trainees.

In this month's issue of *Anesthesia & Analgesia*, a team of educators from Brazil, led by Dr. de Oliveira Filho, report the development of an instrument for evaluation of faculty supervision by anesthesiology residents.³ These educator investigators are to be congratulated for the high quality of their development effort. The investigative team first established appropriate content through systematic text analysis of responses from faculty and residents. Four academic programs in South America contributed to an Internet-based query about behaviors associated with good or poor supervision. Nine dimensions of supervision quality were identified and built into a questionnaire. This questionnaire was then used by 19 residents to evaluate 39 instructors at one institution, generating more than 900 individual scores. The psychometric analysis showed excellent internal consistency, high reliability, and face validity. The greatest component of score variance was related to instructors' supervisory abilities, indicating that the rating instrument was able to distinguish well between instructors of different quality. Another large component of score variance was related to the interaction between residents rating and faculty being rated. This finding, typical for teaching evaluations, points to a substantive halo effect; residents frequently thus rated faculty well or poorly across the board regardless of the specific domain, likely based on unique perceptions such as sympathy or antipathy.

A well-validated instrument for faculty teaching ratings is certainly welcome in the anesthesia education community. However, a number of questions are worth considering:

Does use of validated faculty supervision scores paint an all too reductionist picture of teaching faculty? How well does this assessment tool compare against a "gold standard" of educator quality? Since none exists, this question could be easily dismissed. True, the authors compared their instrument against their own global rating scale and obtained good correlation. Nevertheless, to further solidify construct validity, future research can compare their questionnaire against other validated scales such as Irby's clinical teaching scale which also encompasses attributes such as clarity, enthusiasm, stimulation, knowledge, rapport, instructional skill, clinical competence, and professional characteristics.⁴ Furthermore, both peer-evaluation and teaching portfolios have a substantive role in faculty assessment.^{5,6}

From the Department of General Anesthesiology, Cleveland Clinic, Cleveland Clinic Lerner College of Medicine, of Case Western Reserve University, Cleveland, Ohio.

Accepted for publication May 15, 2008.

Address correspondence and reprint requests to Armin Schubert, MD, MBA, Department of General Anesthesiology, E-31, Cleveland Clinic, 9500 Euclid Ave., Cleveland, OH 44195. Address e-mail to schubea@ccf.org.

Copyright © 2008 International Anesthesia Research Society

DOI: 10.1213/ane.0b013e318182fbf1

Does validation in one South American teaching hospital suffice for universal adoption of this instrument? To be more widely accepted, future validation would ideally occur across more cultures and languages. This is especially true because characteristics of learner raters affect instrument validity and neutralizing their effects in a single institution is nearly impossible because of smaller samples and the need for anonymous evaluation.

What precautions should prevail because of the substantive halo effect observed in this and other⁷ faculty evaluation tools? Perhaps this is the most interesting area for further investigation. Monotonic response patterns (similar ratings of a faculty member across all items of a questionnaire) are known to be affected by questionnaire content presentation format.⁸ Certainly, one can conclude that differentiating faculty along individual components of the assessment instrument would be fraught with error, as would be evaluative rank-ordering of faculty for high stakes decisions such as promotion or compensation adjustment.

Despite these questions, the instrument developed by de Oliveira Filho et al. presents an important advance for many anesthesiology residency training programs. It is easily understood, intuitively appropriate and contains essential components of educator quality such as the ability to provide feedback, which have previously been highlighted in reports from other specialties.^{9,10} Although some teaching programs may have "internally validated" their own teacher rating scales, the validation characteristics of the scale developed by de Oliveira Filho et al. far

exceed those of the typically used in-house teaching assessment instruments. Residency programs may therefore wish to adopt the teacher rating scale reported in this issue of *Anesthesia & Analgesia*. To gain the full benefit of validation, however, the summated rating scale must be used in its entirety rather than adopting pieces and parts.

REFERENCES

1. Schubert A. Anesthesiology resident class sizes and graduation rates. *ASA Newsletter* 2007;71:24-9
2. Tremper KK, Shanks A, Morris M. Five-year follow-up on the work force and finances of United States anesthesiology training programs: 2000 to 2005. *Anesth Analg* 2007;104:863-8
3. de Oliveira Filho GR, Dal Mago AJ, Soares Garcia JH, Goldschmidt R. An instrument designed for faculty supervision evaluation by residents and its psychometric properties. *Anesth Analg* 2008;1316-22
4. Steiner IP, Franc-Law J, Kelly KD, Rowe BH. Faculty evaluation by residents in an emergency medicine program: a new evaluation instrument. *Acad Emerg Med* 2000;7:1015-21
5. Appling SE, Naumann PL, Berk RA. Using a faculty evaluation triad to achieve evidence-based teaching. *Nurs Health Care Perspect* 2001;22:247-51
6. Green ME, Ellis CL, Fremont P, Batty H. Faculty evaluation in departments of family medicine: do our universities measure up? *Med Educ* 1998;32:597-606
7. Bierer SB, Hull AL. Examination of a clinical teaching effectiveness instrument used for summative faculty assessment. *Eval Health Prof* 2007;30:339-61
8. Stratton TD, Witzke DB, Jacob RJ, Sauer MJ, Murphy-Spencer A. Medical students' ratings of faculty teaching in a multi-instructor setting: an examination of monotonic response patterns. *Adv Health Sci Educ Theory Pract* 2002;7:99-116
9. Maker VK, Lewis MJ, Donnelly MB. Ongoing faculty evaluations: developmental gain or just more pain? *Curr Surg* 2006;63:80-4
10. Irby DM. Evaluating clinical teaching in medicine. *J Med Educ* 1981;56:181-6