TITLE: RESIDENT EVALUATION OF STAFF TEACHING

AUTHORS: J.H. Lecky, M.D., F.L. Murphy, M.D., D.E. Greenhow, M.D., E.B. Hutchins, Ph.D.

AFFILIATION: Department of Anesthesia and Office of Educational Research and Development,
University of Pennsylvania, Philadelphia, Pennsylvania 19104

A credible evaluation of postgraduate medical teaching improves resident education via feedback to the staff and documents teaching performance for promotion purposes. We have devised an objective system for collecting residents' opinions of staff performance as teachers of clinical anesthesia.

Methods: Content analysis of a pilot study suggested five important categories: "Availability"(A), "Personal Attributes"(PA), "Case Related Teaching Ability"(CRT), "Didactic Teaching"(DT), and "Overall Contribution to your Learning"(OC). We defined levels of performance in each of these categories to render residents' responses more susceptible to objective analysis(Example 1) and also provided a "comments" section for amplification and qualification.

(Example 1) PERSONAL ATTRIBUTES
PLEASE RATE EACH STAFF PERSON ON THOSE PERSONAL ATTRIBUTES WHICH MAKE HIM/HER AN EFFECTIVE ROLE MODEL FOR YOU. THESE MIGHT INCLUDE: POISE, RAPPORT WITH PATIENTS AND CO-WORKERS, JUDGEMENT, ABILITY TO MAKE DECISIONS, INTELLECTUAL AND MORAL INTEGRITY, KINDNESS, PERSONAL APPEARANCE, ETC.

Rate personal attributes with this scale:
A=OUTSTANDING The staff person possesses those desirable attributes which enhance professional performance, with many outstanding positive characteristics
B=GOOD The staff person falls short of an ideal model by virtue of a major degree of inconsistency or absence of some desirable characteristic, but still remains an anesthetist you can look up to.
C=ADEQUATE The staff person may suffer serious lapses as a role model, but over time performs well enough to satisfy you, most patients, and co-workers.
D=INADEQUATE The staff person has enough serious, consistent flaws in professional and personal behavior that it prevents your learning from him.
E=NO CONTACT

55 residents rated 69 teachers (100% response from current residents) averaging 1.5 hours to complete the form. Evaluations were anonymous, and results went only to the staff person rated (upon request) and to the chairman of the department in all cases.

Results: The means and standard deviations were (A) 3.03±.42, (PA) 2.93±.40, (CRT) 2.93±.46, (DT) 2.89±.44, (OC) 2.69±.47, using a scale of 1=inadequate, 2=adequate, 3=good, 4=outstanding. The means approximated a "good" rating and the standard deviations were quite small.
Rating consistency, assessed using the split-half method corrected for attenuation by the Spearman-Brown prophecy formula,[1] was excellent with estimates of reliability of .86, .90, .91, .89, and .90 respectively for the five categories. (Reliabilities for such systems are usually .6 to .7.)
Correlation coefficients between individual categories ranged between .41 and .81. "Availability" vs. "Didactic Teaching" had the least overlap (.41). (i.e., availability in the O.R. bore little relationship to didactic teaching ability).

Step-wise linear regression of the first four categories on the fifth category (OC) gave: R^2 change[2] (an evaluation of the incremental increase in correlation as a function of adding new variables) values of (DT)=.09, (A)=.01, (PA)=.002, (CRT)=.60. "Case-Related Teaching Ability", then, was the major factor in "Overall Contribution".

We assessed the validity of the evaluation in several ways. A numerical rating based on a tally of the undirected comments as "positive", "mixed", or "negative" gave the same results as did the objectified ratings A through D. Aggregate profiles for agreed with common sense impressions.

Figure 1   Aggregate Teaching Profiles
(A Z score of 0=the average rating, which approximated a "good" in all categories.)
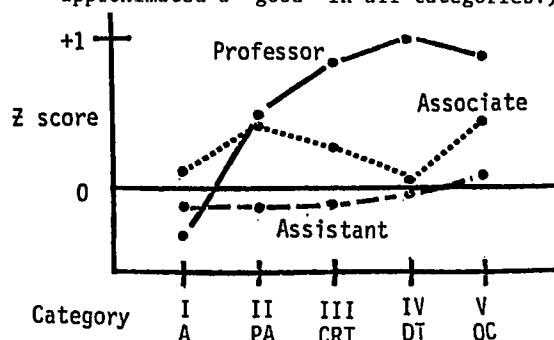
Full professors were less available, but their teaching performance was superior to that of the associate professors, which in turn was superior to that of the assistant professors. (PA) ratings, were similar for full and associate professors but lower for assistant professors. (CRT), (DT), (OC) ratings increased with rank. Confidence in the system increased when the chairman of the department associated correctly 95% of the staff with their unidentified evaluations.

Our procedure for evaluating anesthesia teaching offers several advantages: all residents are willing to participate; data are inherently consistent and reliable by statistical tests; characterizations of staff teaching performance agree well with recognized performance as judged by academic rank and the chairman's opinion; and objective data document teaching ability and allow the chairman to counsel those who should improve. Annual repetition will examine changing opinions of graduate residents and document improved staff teaching for academic management decisions.

References:
1. Noll, VH: Introduction to Educational Measurement. ED II, Houghton-Mifflin Co., Boston, 1965.
2. Nie, NH: Statistical Package For The Social Sciences. ED II, McGraw-Hill Book Co., N.Y., 1975.