EVALUATION OF HUMANISTIC/ETHICAL DIMENSIONS OF MEDICAL STUDENTS' ATTITUDES, VALUES, AND INTERPERSONAL BEHAVIOR

Hi everyone. I'm sorry I can't be with you today. The issue of how to assess student attitudes and behaviors for evidence of humanistic values (Swick's list of honesty, integrity, caring, compassion, altruism, empathy, respect for others, and trustworthiness works for me); as well as adherence to high ethical and moral standards is obviously of great interest to me. Like Swick, I also believe in the critical importance of self-reflection, both in terms of improving clinical decision-making and performance, as well as personal and interpersonal dynamics, as one of the crucial tools in the physician's commitment to a process of life-long learning. But when we move beyond definition to evaluation, things become murky. I've had a chance to review the articles Mike kindly provided, and below I share a few brief and, I'm afraid, somewhat disjointed thoughts.

The first thing that strikes me is how incredibly daunting a task it is to even *attempt* assessment in this area. I was particularly struck by the following issues:

1) How difficult it is to get "experts" to agree on what terms like compassion, altruism, empathy look like when they are translated into practice (Ginsburg, Regehr, & Lingard study). The implication is that it is very difficult to identify a reliable "gold standard." This study also found that, across situations, even the same *individuals* were inconsistent in their opinions. This does not engender confidence in the reliability and validity of assessment in these areas!
2) How much ratings of student professionalism differ depending on which profession (nurses, residents, attendings) is doing the rating (Arnold article). This again speaks to different criteria held by different professions; or differing opportunities to observe students resulting in evaluation of substantially different aspects of their behavior.
3) How many raters it requires to obtain a reliable rating (about 11, according to the Arnold review). This issue vastly complicates peer or supervisor rating systems. A related issue is the number of observations needed to make reliable judgments (between 20 and 50!). What faculty or patient has observed a student 50 times in order to make an informed judgment about their "compassion" or "respect"? I shudder to think that I often rendered such judgments of family practice residents based on 2-3 observations (although of 2-3 hours each).
4) How evaluation itself changes behavior. Arnold makes this point in her excellent article, and I think it's an important one. I'm pretty sure students would not say the things they do in the humanities sessions if they knew I would turn around and fill out an evaluation on the level of their compassion or empathy.
5) How problematic it is to bridge the gap between what people say they would do (or even report what they did) and what actually transpired. I was often struck when observing residents in exam rooms how little their presentation to the attending conformed to what happened. I don't think this is intentional deception so much as creating an acceptable "story," but it raises troubling doubts about relying on written or oral self-reports (critical incident techniques) in evaluation.
6) How difficult it is to get agreement across "humanism" measures. This point was brought home to me with particular force in the study I participated in with Mike,

Desiree, and John Boker, where student essays demonstrating empathy toward a physician whose patient dies showed no correlation with standardized patient assessments of student communication skills and professionalism (which were so closely correlated that the authors question the value of having a separate scale for professionalism). Setting aside the distinct possibility of instrument limitations, it is worrisome at the least that constructs such as these, which seem intuitively related, may in fact operate independently of each other.

7) A related concern arising from this same study is that expression of professionalism etc. may be highly case-dependent. If that is so, it might be possible for students to receive low humanistic evaluations in direct observation situations simply because they did not have the good fortune to have the "right" kinds of patients to evoke empathic, compassionate responses.

To the above research-based concerns I must add a philosophical reservation of my own. I believe strongly that individuals in the helping professions, such as medicine, have a moral obligation to regularly reflect on their own attitudes, values, and behavior, with the goal of making themselves the most moral, humane, and compassionate people they can be professionally and personally. Further, I very much endorse the concept of transparent, authentic relations among colleagues (and teachers and students) where there is an honest exchange of constructive feedback and criticism through a process of mutual reflection. However, I see both of these endeavors (personal and interpersonal) as more horizontal, collegial processes than formally evaluative ones. It seems problematic at best and presumptuous at worst to me to formally judge another person's moral compass, except perhaps in the rather extreme situations referred to in the Papadakis, Loeser, and Healy article. Even in these cases, it was not clear that all the students actually benefited from the write-ups; and certainly not evident that these students became more ethical or more professional as a result of the reports issued about them.

That having been said, if we are to embark on an evaluation process for the humanistic and ethical aspects of professionalism, despite the limitations above, I would endorse multiple strategies that can be triangulated, and therefore achieve some measure of trustworthiness. For example, the following seem to me to have some potential, and if they were conducted simultaneously, might produce outcomes which we could regard with some confidence (the only problem is that they would be so time-consuming most of the other activities of the institution would grind to a halt!).

1) 360 degree evaluation. I am persuaded that this method of evaluation has much to recommend it; and that it could be applied to the ethical and humanistic domains as well. Students, for example, observe each other with an intimacy not available to any other evaluators. But residents and faculty also bring the perspective of clinical experience and can place students within a larger context. However, rather than Likert-type evaluations, I think narratives tell us a lot more about humanistic/ethical dimensions, because they provide specific detail (stories rather than numbers).
2) Critical incident essays. Despite concerns about written evidence versus real-time observation, I continue to believe that we can learn a lot from how students

describe clinical situations, how they define ethical dilemmas, what they believe are right and wrong actions, and how they position themselves within these situations. Therefore, I feel some type of evaluative self-report might provide useful information.

3) Patient feedback. I do not believe patient feedback trumps other sorts of information, because patients themselves may not always see their interactions with others clearly and dispassionately. For example, they may be swayed by pain or by past encounters with medical personnel. Conversely, they may have a positive bias toward young, nice (although not necessarily ethical or humanistic) students. However, a system that would provide patient feedback on humanistic dimensions of caring, compassion, empathy etc. could be extraordinarily valuable.
4) Standardized patients. I have certain qualms about the extent to which standardized patients can adequately evaluate humanistic qualities. However, in deference to the article by Arnold which asserts that OSCEs are useful in assessing ethical behavior, I can see using this format as part of an overall assessment.