STATE OF THE ART AND SCIENCE
Bias in Assessment of Noncognitive Attributes
Rick D. Axelson, PhD, and Kristi J. Ferguson, MSW, PhD

Professional competence for physicians, as defined by Epstein and Hundert [1], is:

> the habitual and judicious use of communication, knowledge,
> technical skills, clinical reasoning, emotions, values, and reflection in
> daily practice for the benefit of the individual and community being
> served [2].

As implied by their definition, noncognitive traits figure prominently in Epstein and Hundert's discussion of physicians' professional competence. They cite attributes such as respect for patients, caring, emotional intelligence, teamwork, tolerance of ambiguity and anxiety, and basic communication skills as fundamental components of professional competence.

Although the importance of characteristics like those mentioned above is clear, it is quite difficult to assess the extent to which individuals possess them. Traits and skills related to providing humane medical care such as "caring" and "emotional intelligence" are much easier to recognize in practice than they are to explicitly define and measure. The crux of the problem is that traits and skills that develop over time through personal experience (i.e., learning-by-doing) in various social contexts can be difficult to express in words. They are often described as "tacit knowledge," i.e., knowledge and skills that enable one to perform certain tasks without necessarily fully knowing, or being able to explain, how one does it. Lam notes that in contrast to the type of knowledge associated with cognitive skills (explicit knowledge), tacit knowledge is personal and contextual [3]. Consequently, it is difficult to articulate, formalize, and share with others.

Although physicians' tacit knowledge enables them to recognize similar competence in others, there are two major drawbacks to overreliance upon tacit knowledge as the basis for admission and evaluation processes. First, physicians' tacit knowledge reflects their personal, perhaps idiosyncratic, understandings of the essential noncognitive traits and skills. These views may vary considerably among faculty. To be used effectively in admission and evaluation processes, these views would need to be synthesized and articulated as a shared vision of the medical school community. Secondly, because tacit knowledge develops through social interaction over time, it most likely contains outmoded beliefs and biases that hamper objective evaluation of others. A recent AAMC literature review described evidence of erroneous tacit knowledge in the form of unconscious gender or race/ethnicity bias [4]. The review

cited several studies showing that evaluators' awareness of gender or race/ethnicity caused them to mistakenly favor one equally qualified candidate over another. Thus, unchecked reliance upon tacit knowledge can result in biased recruitment and evaluation decisions.

Therefore, the central challenge in evaluating noncognitive traits is to leverage the useful portions of physicians' tacit knowledge into a common understanding of the most essential traits, while at the same time minimizing the influence of personal biases and irrelevant or mistaken information. Ultimately, if we are to select and develop physicians' capacity for requisite noncognitive skills and traits, we need reliable, valid, and transparent methods for measuring them.

In the following section, we outline strategies for refining organizations' processes to define and assess crucial noncognitive attributes. Effective use of research methods and data to move toward more explicit understanding of the desired characteristics and valid assessment of them is the guiding principle for this approach. The four steps are intended as elements of an iterative cycle to continuously improve processes for evaluating noncognitive attributes.

Improving Assessment of Noncognitive Attributes
1. *Develop more explicit definitions of the desired skills and attributes.* Oftentimes the daunting task of developing precise definitions of learning outcomes is addressed by committees or task forces. To support such work, preliminary qualitative research methods (c.f., Denzin and Lincoln [5], Giorgi [6]) can be used to describe and analyze the tacit knowledge available among medical school personnel regarding their interpretations and understandings of the noncognitive traits needed for professional competence and how these skills and attributes can be recognized in practice.

   Foundational research, like the above, can guide committees' deliberations as they seek to identify the most essential noncognitive traits and explore practical means for assessing them. Without locally developed research, members may struggle to articulate their tacit knowledge and get frustrated by the size and difficulty of their task. Under such conditions, committee members face the temptation of settling for the most easily defined and measurable traits rather than struggling to express and define the most essential ones. Research and conceptualizations can support efforts to make explicit their understanding (i.e., the "externalization" of tacit knowledge) of the desired attributes.

2. *Structure data collection to observe instances of the desired traits.* With a more explicit understanding of the desired noncognitive attributes, one can fine-tune the methods used to assess them. Assessment processes can be refined to elicit more revealing and relevant performances. Behavior-based interviewing [7] and the Multiple Mini-Interview [8] (for example, are

valuable approaches to consider for gathering more useful assessment information from applicants.

3. *Train raters/evaluators to use the system*. Better tools for assessing the desired traits will only improve outcomes if evaluators are trained and have the opportunity and resources to use those tools properly. Martell's research provides evidence of the importance of sufficient time, information, structure, and training for reducing the use of irrelevant information, including stereotypes and bias in evaluations [9]. Like most skills, practice and experience also seem to improve the quality of evaluations [10].

4. *Provide feedback to evaluators/raters*. When aided by thoughtful reflection and feedback on the accuracy of their previous decisions, one would expect that evaluators would reap even greater benefits from their experience. Toward this end, the Implicit Association Test [11] is one example of a resource that can help individuals identify sources of unconscious bias affecting their evaluations.

Although there are numerous types of feedback that could be provided to evaluators, here we describe an analysis that provided evaluators feedback on bias in their evaluations.

In a recent study at the University of Iowa, we analyzed 5 years of clinical performance evaluation forms for evidence of unconscious gender bias in the ratings of our medical students [12]. Our method involved examining whether the meaning of adjectives was affected by the gender of the student being rated. Within a factor analysis framework, highly intercorrelated groups of adjectives are interpreted as having a similar meaning; the common meaning for a given adjective grouping is represented by an underlying factor. If raters use the same meaning of the adjective regardless of the student's gender, then the expected pattern of intercorrelations and underlying factors among adjectives would be the same for men and women students. This hypothesis was tested statistically using Multigroup Confirmatory Factor Analysis (CFA). (See Brown [13] for an accessible description of this technique.)

From this analysis, we found that raters did, in fact, interpret the adjectives (i.e., "measurement models") differently based on the gender of the student being rated. These different measurement models resulted in gender-biased evaluations. Women were given more credit than comparable men for being "compassionate," "sensitive," and "enthusiastic," and men were given more credit than comparable women for being "quick learners." Thus, this type of analysis enabled us to raise evaluators' awareness of an unconscious bias evident in the pattern of their ratings.

In sum, physicians' tacit knowledge of vital noncognitive attributes provides invaluable raw data for developing, implementing, and refining assessment processes. As outlined in the steps above, qualitative and quantitative research

methods can facilitate efforts to externalize tacit knowledge, improve measurement processes, and correct implicit biases in judgments based upon tacit knowledge. Ultimately, however, it is physicians' reflective and judicious use of such research that will enable them to create increasingly meaningful and accurate processes for assessing noncognitive attributes.

References
1. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287(2):226-235.
2. Epstein, Hundert, 226.
3. Lam A. Tacit knowledge, organizational learning and societal institutions: an integrated framework. *Organ Stud*. 2000;21(3):487-451.
4. Association of American Medical Colleges. Analysis in brief: unconscious bias in faculty and leadership recruitment: a literature review. *AAMC*. 2009;9(2):1-2. https://www.aamc.org/download/102364/data/aibvol9no2.pdf. Accessed November 2, 2012.
5. Denzin NK, Lincoln YS, eds. *The SAGE Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications; 2000.
6. Giorgi A. The Descriptive Phenomenological Method in Psychology: A Modified Husserlian Approach. Pittsburgh, PA: Duquesne University Press; 2009.
7. Altmaier EM, Smith WL, O'Halloran CM, Franken EA Jr. The predictive utility of behavior-based interviewing compared with traditional interviewing in the selection of radiology residents. *Invest Radiol*. 1992;27(5):385-389.
8. Eva KW, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. *Med Educ*. 2004;38(3):314-326.
9. Martell RF. Sex bias at work: the effects of attentional and memory demands on performance ratings of men and women. *J Applied Social Psychol*. 1991;21(23):1939-1960.
10. Ferguson KJ, Kreiter CD, Axelson RD. Do preceptors with more rating experience rate medical student performance more reliably? *Teach Learn Med*. 2012;24(2):101-105.
11. Project Implicit web site. https://implicit.harvard.edu/. Accessed November 2, 2012.
12. Axelson RD, Solow CM, Ferguson KJ, Cohen MB. Assessing implicit gender bias in medical student performance evaluations. *Eval Health Prof*. 2010;33(3):365-385.
13. Brown TA. *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford; 2006.

Further Reading
Eva KW, Reiter H. Where judgment fails: pitfalls in the selection process for medical personnel. *Adv Health Sci Educ Theory Pract*. 2004;9(2):161-174.

Wood PS, Smith WL, Altmaier EM, Tarico VS, Franken EA Jr. A prospective study of cognitive and noncognitive selection criteria as predictors of resident performance. *Invest Radiol*. 1990;25(7):855-859.

Rick D. Axelson, PhD, is an assistant professor in the Department of Family Medicine and a program evaluation consultant for the office of consultation and research in medical education at the University of Iowa Carver College of Medicine in Iowa City. Dr. Axelson has developed and directed academic program assessment and institutional research offices at the University of Missouri-Kansas City, the University of South Alabama, and Riverside Community College. His research interests include program evaluation (theory, methods, and practice), learning outcomes assessment, and the development of practical methods for assessing the social, cognitive, and psychological factors affecting students' engagement in learning activities and environments.

Kristi J. Ferguson, MSW, PhD, is a professor of general internal medicine, director of the office of consultation and research in medical education, and director of the master's in medical education program at the University of Iowa Carver College of Medicine in Iowa City. Her research interests include assessing the validity of measures of student performance, assessing the predictive value of a small group experience during the admissions process, and evaluating students' ability to recognize team behaviors in a simulation environment.

Related in VM
Expectations of Gender in Medical Education, December 2012

Sex Discrimination in Selection for Residency, July 2008

The State of Medical Education Research, April 2007

Standardizing and Improving the Content of the Dean's Letter, December 2012

Assessing Noncognitive Attributes: The Primary Care Orientation Scale, December 2012