

# Stimulating and Evaluating Acquired Knowledge of Addiction Among Residents Through Repeat Testing: A Pilot Study

Maher Karam-Hage, MD,<sup>1,2</sup> Fangqian Ouyang, BS,<sup>3</sup> Jihane Ghorayeb, BA,<sup>4</sup> Patricia Mullan, PhD,<sup>5</sup> Kirk Brower, MD,<sup>6</sup> Larry Gruppen, PhD<sup>7</sup>

<sup>1</sup>Department of Behavioral Science, University of Texas MD Anderson Cancer Center, Houston, Texas

<sup>2</sup>Department of Psychiatry, University of Texas MD Anderson Cancer Center, Houston, Texas

<sup>3</sup>University of Texas, School of Public Health, Houston, Texas

<sup>4</sup>University of Leeds, Leeds, United Kingdom

<sup>5</sup>Department of Medical Education, University of Michigan Medical School, Ann Arbor, Michigan

<sup>6</sup>Department of Psychiatry, University of Michigan, Ann Arbor, Michigan

<sup>7</sup>Department of Medical Education, University of Michigan, Ann Arbor, Michigan

**Background:** Addictive disorders receive little attention in medical school and residency program curricula.

**Objective:** To evaluate an innovative learning approach encouraging and stimulating residents to focus on key competencies by testing before and after their addiction psychiatry rotation.

**Methods:** We developed a 50-item test on substance use disorders. Twenty-six general psychiatry residents, post-graduate year I (PGY-I) and II (PGY-II), participated in the pilot study and were divided into three groups. PGY-I residents were divided into Group 1, who were tested the last day of the rotation and again 2 months later, and Group 2, who were tested on the first and the last day of the rotation. Eight of 11 PGY-II residents agreed to participate as controls (Group 3), as they had previously completed their 2-month addiction psychiatry rotation as PGY-I's. All residents were informed that the testing would not affect their individual grade. After taking the first test, all three groups received related study materials.

**Results:** A statistically significant increase in re-test scores occurred in the combined groups ( $p < .001$ ). The largest changes in scores were among Group 2 (the group taking the test on first and last day of their addiction psychiatry rotation).

**Conclusion:** The greatest learning seemed to occur when residents were tested at beginning and end of the rotation. However, all residents' test scores improved to some degree, regardless of their level of training or the timing of the test.

**Scientific Significance:** This study offers support for testing as a learning guide and as a means of stimulating residents' learning. (*Am J Addict* 2014;23:576–581)

## INTRODUCTION

Residency programs expect their trainees to develop competence in a broad range of substance use disorders, but often programs have not developed systematic ways to teach or test for the acquisition of knowledge on this topic. Perhaps it is not surprising that graduating residents report a lack of preparedness to treat substance use disorders. In 2002, the American Council on Graduate Medical Education (ACGME) and the Residency Review Committee (RRC) mandated that all residency programs evaluate residents on six defined competencies.<sup>1</sup> Although, optimal methods of teaching and evaluating residents need to be developed, medical education leaders have encouraged greater use of more frequent and explicit assessments during training to motivate learning and inform self-assessment.<sup>2</sup> In addition, in 2001 the RRC mandated that substance use disorders rotations be at least 4 weeks in psychiatry residencies. Sattar et al.<sup>3</sup> developed a 16-item scale called the Addiction Training Scale (ATS) to help evaluate psychiatry residents' self-assessment of knowledge about addiction. However, the ATS measures the test taker's subjective level of confidence in various areas of addiction training without objectively measuring the resident's knowledge.

It is important to develop methods that can help residents learn the basic concepts of that specialty<sup>4</sup> with the expectation that they will acquire a consistent and more standardized training from month to month and a solid foundation for later continuing medical education. We designed an educational tool in a multiple-choice test format that would "prime" residents on key concepts in addiction psychiatry. Trainees were informed that the test would not be used to grade them; rather, the test was intended to guide learning during their clinical

Received July 23, 2013; revised February 20, 2014; accepted April 12, 2014.

Address correspondence to Dr. Karam-Hage, Department of Behavioral Science, University of Texas MD Anderson Cancer Center, 1155 Pressler St, C.P.B. Unit 1330, Houston, TX 77030. E-mail: maherkaram@mdanderson.org

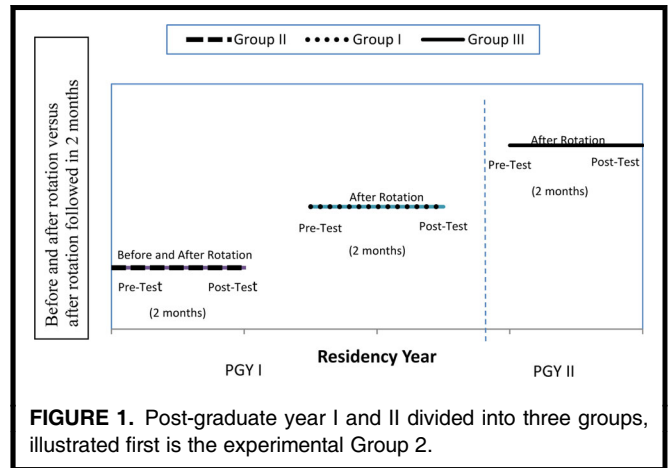
experiences and formal didactic instruction. The test was then repeated 8 weeks later to assess residents' acquisition of knowledge during that period of time. The 8 weeks corresponded to the duration of rotation in substance use/addiction treatment (SU-AT) at that time. This paper summarizes preliminary evidence of the feasibility, success, and advantages of this approach, which integrates clinical training and knowledge-based assessment.

## METHODS

This pilot study was conducted during the 2005–2006 academic year when a 2-month rotation in SU-AT was required. The clinical sites for the rotation were the Veterans Administration Healthcare System of Ann Arbor and an outpatient addiction treatment program of the University of Michigan. To guide and standardize the residents' learning of basic concepts in addiction and to objectively assess their acquired knowledge, we developed a 50-item multiple-choice test. Before using the test as part of the formal rotation, we evaluated its usefulness via a prospective case-control study using a within-subjects design (same test repeated in same subject). We preserved residents' confidentiality by using a pre-set code specific to each resident's pair of tests. During the testing period, we did not score, tabulate, or match any of the tests until the study was completed, and all participating residents were evaluated through the usual procedures. In other words, residents did not receive feedback on their individual test performance. We chose this approach to ensure residents' confidentiality and to avoid influencing the faculty's judgment when evaluating the residents.

Our original sample consisted of 26 residents; however, one PGY-I resident did not finish the post-test, another PGY-I did not take the post-test, and two PGY-II residents dropped out of the study, citing lack of time, and did not take the post-test. As a result, we had a study sample ( $N = 22$ ) comprising 14 PGY-I and 8 PGY-II residents. Eighty-one percent (18/22) of residents were Caucasian; 68% (15/22) were women; and their mean age was 30.3 years (range 27–40).

The residents were divided into three groups (see Fig. 1). Groups 1 and 2 were all PGY-I residents, whereas Group 3 comprised PGY-II residents. Group 1 (PGY-I) took their first test on last day of the rotation and their re-test (follow up) 2 months after. Group 1 served as controls for the training usually acquired during the rotation among PGY-Is. Group 2 took their first test the first day of the rotation and their re-test the last day of the 2-month SU-AT rotation. Group 2 comprised the "experimental group." Group 3 had finished their addiction rotation 2–12 months earlier so their first test administration was 2–12 months after finishing the rotation and their second test administration was 2 months after. Group 3 served as an overall control for the effect of repeated exposure to the test (priming effect) and for the possibility of cumulative learning taking place over time in subsequent clinical rotations (maturation effect). Residents were informed of the study



**FIGURE 1.** Post-graduate year I and II divided into three groups, illustrated first is the experimental Group 2.

and were invited to participate voluntarily, with the option to opt out. We obtained an exemption from the requirement of informed consent from the Institutional Review Board (IRB) because the study represents "education research."

The residents were informed that the tests taken would not be scored until the study was completed and were assured that the test scores would not be used to evaluate them throughout their rotation. We emphasized that the tests were part of a pilot study that we were conducting. After taking the test the first time, all participating residents were given eight pages of "general concepts in addictions," which covered the content in our test. Those concepts were discussing each question without providing a specific letter-answer on the multiple-choice items. As an incentive to participate and after completing the first test, we provided residents with a copy of the *Textbook of Substance Abuse Treatment, 3rd edition* (Galanter and Kleber, American Psychiatric Publishing, 2004). The textbooks were purchased through funds from the Medical Education Scholars Program at the University of Michigan Medical School, a program in which the first author (M.K.-H.) was participating as a faculty scholar. The first author (a board certified addiction psychiatrist) developed the different test items and concept answers (appendix-I), those were reviewed and edited by one of the senior authors (K.B. a board certified addiction psychiatrist and director of the addiction psychiatry fellowship program), and then they were tested among rotating residents before beginning the experiment. Those residents were asked to provide feedback on accuracy, ease of reading, usefulness and completeness of the questions and concepts. Several edits were incorporated as a result of the residents' input and two items were replaced as they had more than one answer. The results were analyzed across groups (combined on first test and re-test scores) and then separately by comparing the three groups. Each resident served as his/her own control (a repeated-measures design).

## Statistical Analysis

Stata version 12 (STATA Corp., TX) was used for data analysis. We summarized the demographic characteristics of

participants and the amount of time they reported that they spent reading in total and using the supplementary material associated with our test. The normality of data were tested using the Shapiro–Wilk test. Non-parametric tests (Kruskal–Wallis test and Wilcoxon signed-rank test) were used due to small sample size. The three groups were compared for significant differences in pre-test scores using the Kruskal–Wallis test. For a paired comparison of the first and second test scores within each group, we used the Wilcoxon signed-rank test. Then we applied the Student’s *t*-test; two individual *t*-tests were used for two comparisons: change scores before and after rotation versus those tested only after rotation; change scores in PGY-I versus PGY-II. Because only two groups were compared, *t*-tests were used instead of analysis of variance, with a *p* value of <.05 considered significant.

## RESULTS

During the study period, 26 residents were recruited, and 22 (84.6%) returned both their first and second tests. Table 1 shows the demographic characteristics of the remaining 22 residents, including 4 (18%) in Group 1, 10 (45%) in Group 2, and 8 (37%) in Group 3. The three groups had similar demographic characteristics. Table 2 indicates the self-reported time (per week or total hours) spent reading about SU–AT during the 2-month interval between tests. Residents in Group 2 spent significantly more hours per week and in total hours reading about SU–AT in the last 2 months (*p* = .00). However, when divided by method of reading, there were no significant differences in the amount of time spent in reading in the textbook or answer concepts that we provided, between the three groups.

Most residents in Group 2 (80%) reported spending more than 10 total hours reading about SU–AT, during the entire study period; compared to only 12.5% of the residents in Group 3, and no residents in Group 1.

The Kruskal–Wallis test showed no significant differences in performance on the pre-test (*p* = .197) among the three groups. Figure 2 depicts the pre- and post-test scores in Groups 1, 2, and 3. The Wilcoxon signed-rank test was conducted to compare the performance between pre- and post-test within each group. In Group 2 (experimental group), a significant difference in performance was seen between the two tests, with the participants doing better in the post-test (median = 41.5 points) than in the pre-test (median = 34 points) (*p* = .005); no significant difference was found in both control Group 1 and Group 3 (*p* = .068 and *p* = .123, separately).

Because the test performance improved apparently more in the experimental group (Group 2), we expected that the greatest test score change would occur for that group (when testing is done before and after the rotation). To make a direct comparison on that basis, we collapsed Groups 1 and 3 and compared their score change with Group 2 score change. We used an unpaired two sample Student’s *t*-test to compare the mean change score (from the first to second test) of the residents in Group 2 (mean = 7.7, SD = 3.56, *p* = .000) versus the residents in Groups 1 and 3 (mean = 3.8, SD = 4.44, *p* = .006). This was a between groups comparison by unpaired two sample student’s *t*-test. The two samples are two groups of score changes. Group 2 had a significantly higher mean change score (between pre- and post-test) than did a combined Groups 1 and 3 (*p* = .019). Using the same methods, to check for any year of training differences, we collapsed Groups 1 and 2 (all PGY-I residents) resulting in a mean change score of 7.1 (SD = 3.65, *p* = .000) that was also compared with PGY-II

**TABLE 1.** Demographic characteristics of 22 residents

|                | Group 1 (N=4) |      | Group 2 (N=10) |      | Group 3 (N=8) |      | Overall sample |      |
|----------------|---------------|------|----------------|------|---------------|------|----------------|------|
|                | <i>n</i>      | %    | <i>n</i>       | %    | <i>n</i>      | %    | <i>n</i>       | %    |
| Sex            |               |      |                |      |               |      |                |      |
| Male           | 3             | 75.0 | 7              | 70.0 | 5             | 62.5 | 15             | 68.0 |
| Female         | 1             | 25.0 | 3              | 30.0 | 3             | 37.5 | 7              | 32.0 |
| Age (range)    |               |      |                |      |               |      |                |      |
| 26–30          | 3             | 75.0 | 6              | 60.0 | 6             | 75.0 | 15             | 68.0 |
| 31–35          |               |      | 2              | 20.0 | 1             | 12.5 | 3              | 14.0 |
| 36–40          | 1             | 25.0 | 2              | 20.0 | 1             | 12.5 | 4              | 18.0 |
| Race           |               |      |                |      |               |      |                |      |
| Caucasian      | 2             | 50.0 | 7              | 70.0 | 4             | 50.0 | 13             | 59.0 |
| Asian-American | 1             | 25.0 | 1              | 10.0 | 3             | 37.5 | 5              | 23.0 |
| Other          | 1             | 25.0 | 2              | 20.0 | 1             | 12.5 | 4              | 18.0 |
| Income (range) |               |      |                |      |               |      |                |      |
| <50 K          | 1             | 25.0 | 3              | 30.0 | 2             | 25.0 | 6              | 27.0 |
| 50–100 K       | 2             | 50.0 | 2              | 20.0 | 2             | 25.0 | 6              | 27.0 |
| 101–200 K      | 0             | 0.0  | 1              | 10.0 | 2             | 25.0 | 3              | 14.0 |
| >200 K         | 1             | 25.0 | 4              | 40.0 | 2             | 25.0 | 7              | 32.0 |

**TABLE 2.** Self-reported time spent reading about substance use and addiction psychiatry (SU-AP) in the prior 2 months

|  | Group 1 (N=4) |     | Group 2 (N=10) |    | Group 3 (N=8) |      | p-Value    |
|--|---------------|-----|----------------|----|---------------|------|------------|
|  | n             | %   | n              | %  | n             | %    |            |
| Hours per week reading about SU-AT in the last 2 months  |               |     |                |    |               |      |            |
| 0-2  | 4             | 100 | 2              | 20 | 8             | 100  | <b>.00</b> |
| 2 or more  | 0             | 0   | 8              | 80 | 0             | 0    |            |
| Total hours reading about SU-AT in the last 2 months     |               |     |                |    |               |      |            |
| 0-10   | 4             | 100 | 2              | 20 | 7             | 87.5 | <b>.00</b> |
| 10 or more   | 0             | 0   | 8              | 80 | 1             | 12.5 |            |
| Total hours reading in the textbook in the last 2 months |               |     |                |    |               |      |            |
| 0-2  | 4             | 100 | 6              | 60 | 7             | 87.5 | .27        |
| 2 or more  | 0             | 0   | 4              | 40 | 1             | 12.5 |            |
| Total hours reading from the Q & A in the last 2 months  |               |     |                |    |               |      |            |
| 0-1  | 2             | 50  | 4              | 40 | 7             | 87.5 | .10        |
| 1 or more  | 2             | 50  | 6              | 60 | 1             | 12.5 |            |

residents mean = 3 (SD = 4.72,  $p = .058$ ). Interestingly, with that collapsing, PGY-I residents had statistically significant higher test change scores than did PGY-II residents ( $p = .017$ ). We calculated the effect size (Cohen's  $d$ ) comparing the change score in Group 2 to the change scores in Groups 1 and 3, this resulted in Cohen's  $d$  value of 0.975, a large effect size (Fig. 2).

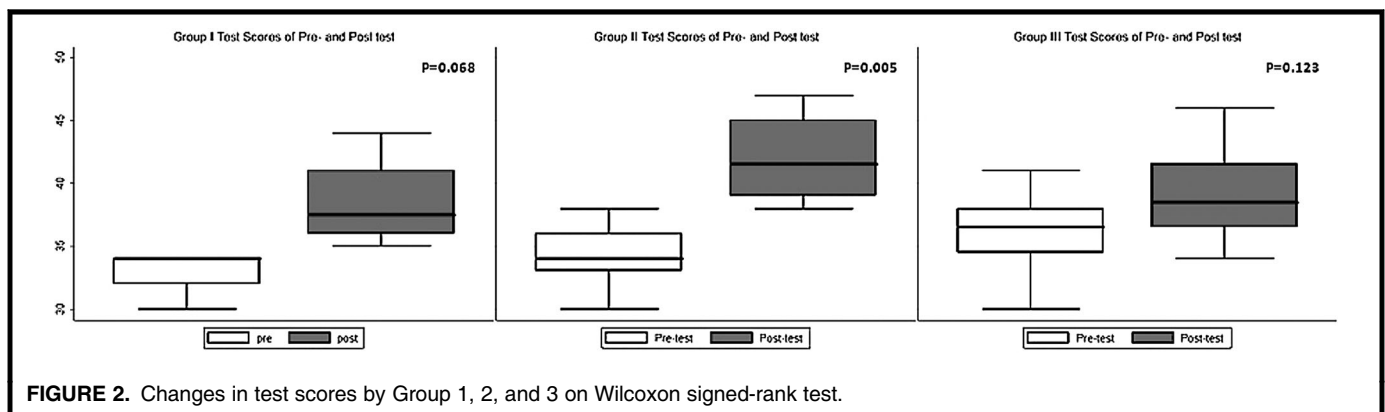
## DISCUSSION

The design of rotations and evaluation of residents' training is often based on subjective experiences and impressions.<sup>5</sup> When studies have utilized specific curricula and tests, they typically have not been tested in real-life settings or evaluated for validity and standardization.<sup>6</sup> Like many other subspecialties, addiction psychiatry is an expanding field in which a new learner can become overwhelmed with the vast array of concepts and competing theories as well as new discoveries and recent research findings. On the other hand, it is not realistic to expect a psychiatry resident to master a subspecialty's knowledge in the minimum of 4 weeks mandated by

the Accreditation Council for Graduate Medical Education (ACGME) or even in 2 months, as is the best-case scenario in a few residency programs.

In a busy clinical environment with limited time for reading, both theoretical (abstract) and experiential learning do become highly variable. This variability depends on the mix of cases at any particular time during the rotation and maybe mitigated if residency programs follow a substance use disorders curriculum covering all 4 years of residency.<sup>7</sup> Nevertheless, mastery of basic concepts is what ultimately constitutes clinical competency in a given specialty and is the premise of our proposed approach. We believe that our method offers an innovative tool to stimulate learning across the spectrum of residents and to provide a preliminary assessment of residents' abstract knowledge, which is a required competency that is objectively tested. More research is warranted to develop standardized tools designed to systematically and accurately assesses the wide array of knowledge to be covered by medical residents.

In summary, we provide a proof of concept regarding a novel method of guiding residents' study that doubles as an



**FIGURE 2.** Changes in test scores by Group 1, 2, and 3 on Wilcoxon signed-rank test.

objective evaluation tool. This method might be applicable in other subspecialties, especially those in which the learner has too little time to master extensive details and lacks the background necessary to prioritize and cover the basics. As we expected, our results suggest that repeated testing before and after rotation, a provision of content summary and concept answers are significantly associated with improvement in test scores. In other words, the ideal timing for the first test was before the rotation with a re-test after the rotation. We speculate that this particular order “sensitizes” residents by exposing them to what they need to learn, followed by providing a study guide that they can consult during their subspecialty rotation. We provided the residents with summaries of the concepts covered on the test, rather than just their score or an answer key. We believe that this order of exposure gives relevancy to the items on the test and helps consolidate the abstract learning by experiential learning through the rotation.

Medical educators advocate for the use of teaching interventions that promote learners’ self-regulated learning. In this study, this educational intervention took the form of a pre-instruction test, intended to serve as a catalyst for learners to anticipate key concepts and lessons. This use of testing was also intended to facilitate learners’ ability to reflect and connect with their prior related knowledge.

Of note, the experimental group (ie, the group who participated in the pre-rotation test) reported allocating more time to related reading than did their peers. This finding that trainees spent more time on related self-regulated learning behavior may simply be by chance this group was heavier readers; however, with the significant differences between groups as well as subsequently performing better on related assessments, this is likely to be a direct consequence, which is congruent with learning motivation theories in medical education. These theories call attention to the advantages of “priming” learners to reflect and activate previous related learning and to anticipate learning that will inform their growing competence.<sup>8</sup>

In addition to this study’s congruence with theoretical approaches to teaching and learning in medicine, this study’s design and reporting were also informed by emerging recommendations for medical education’s research methods and reporting of findings. In terms of research methods, Lynch et al. urge that medical education research designs address predictable limitations—particularly threats to internal validity that would leave unanswered sources of confounding—by “synthetic” research designs.<sup>9</sup> In our study, the research design augmented the traditional pre- and post-test group with both a comparison post-test only group and testing of a group of more advanced learners. In terms of reporting research findings, leaders and journals are increasingly advocating that studies report their findings in terms of their associated effect sizes. Effect sizes provide more information about the magnitude of a study’s impact than does statistical significance testing; a further advantage is that effect sizes can be directly compared across studies, guiding our interpretation of what approaches achieve the strongest impact.<sup>10</sup> In our study, effect sizes let us

appreciate that, while learning and improved test performance increased in all comparison groups, the magnitude of the change was substantially higher in the experimental (pre- and post-rotation testing) group.

Future in-depth research in this area is required before this concept can be implemented on a large scale. Some of the issues to be addressed include determining the psychometric characteristics of the test questions; evaluation of the concept’s applicability to other settings in psychiatry or different specialties; levels of training required; and whether it matters if the teaching institution has a research or clinical focus. Because of the positive feedback we received directly and anonymously through the comment section of this test, we commend using this or similar methods to help residents learn the basics of addiction psychiatry and to evaluate their increase in medical knowledge more systematically. Since reading time was significant predictor of learning, it would be important in future studies to build in some mechanisms to encourage reading, perhaps via focused scheduled discussions of different topics throughout the rotation. Accordingly, programs can better meet their training and evaluation goals, as well as satisfy ACGME requirement (to quantify abstract knowledge by having a formal assessment), by using this method instead of relying on subjective impressions. To our knowledge, no other alternatives to classic testing have been developed to meet the new evaluation requirements; therefore, we believe that this type of research needs to be expanded.

A potential drawback would be a scenario in which residents relied on the test concept answers as their only learning method. However, we believe that issue can be avoided or minimized with proper explanation and education and, when possible, by providing residents with additional materials such as classic papers or a textbook on addiction. The other possible disadvantage of this new approach is the potential for a resident’s pre- or post-rotation score to color the faculty’s opinion of the resident. This, in turn, could distort the faculty’s perception of other aspects of his or her training, including interviewing skills, engaging patients and families, diagnostic and analytic skills, and problem solving, which constitute the full spectrum of real clinical competency. Therefore the test score ought to be thought of as measure of competency only in the abstract knowledge acquired.

#### *Declaration of Interest*

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this paper.

#### **REFERENCES**

1. ACGME. *Accreditation Council for Graduate Medical Education*. ACGME; 2013: Available at: URL: <http://www.acgme.org/acgmeweb/About/ACGMEHistory.aspx>. Accessed June 25, 2013.
2. Van D V, Schuwirth LW, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34:205–214.

3. Sattar SP, Madison J, Markert RJ, et al. Addiction training scale: Pilot study of a self-report evaluation tool for psychiatry residents. *Acad Psychiatry*. 2004;28:204–208.
4. Srihari V. Evidence-based medicine in the education of psychiatrists. *Acad Psychiatry*. 2008;32:463–469.
5. Osser DN, Patterson RD, Levitt JJ. Guidelines, algorithms, and evidence-based psychopharmacology training for psychiatric residents. *Acad Psychiatry*. 2005;29:180–186.
6. Klein DF. Comments on psychiatric education. *Acad Psychiatry*. 2005;29:128–133.
7. Iannucci R, Sanders K, Greenfield SF. A 4-year curriculum on substance use disorders for psychiatry residents. *Acad Psychiatry*. 2009;33:60–66.
8. ten Cate O, Snell L, Mann K, et al. Orienting teaching towards the learning process. *Acad Med*. 2004;79:219–228.
9. Lynch DC, Whitely TW, Willis SE. A rationale for using synthetic designs for medical education research. *Adv Health Sci Educ*. 2000;5:93–103.
10. Hojat M, Xu G. A visitor's guide to effect sizes: Statistical significance versus practical (clinical) importance of research. *Adv Health Sci Educ*. 2004;9:241–249.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article: