JOURNAL OF SURGICAL RESEARCH 193 (2015) 554–559

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Evaluation of the validity of medical students' self-assessments of proficiency in clinical simulations

Gerald A. Isenberg, MD,[a,*] *Vibin Roy, MD,*[b] *Jon Veloski, MS,*[d] *Katherine Berg, MD,*[c] *and Charles J. Yeo, MD*[a]

[a] *Department of Surgery, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania*
[b] *Department of Family and Community Medicine, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania*
[c] *Department of Medicine, and the University Clinical Skills and Simulation Center, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania*
[d] *Department of Psychiatry and Human Behavior, and the Center for Research in Medical Education and Health Care, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania*

ARTICLE INFO

Article history:
Received 12 August 2014
Received in revised form
12 September 2014
Accepted 25 September 2014
Available online 2 October 2014

Keywords:
Medical student education
Self-assessment
Simulation

ABSTRACT

Background: The accuracy of self-assessments has not been well supported in the literature. This study was undertaken to examine the validity of medical students' ratings of their proficiency during encounters with simulated patients and simulation devices.

Methods: Confidential self-assessments for 10 skills were collected from 195 students during a formal clinical skills assessment related to 3 cases at the end of a surgery clerkship. The cases required students to gather data from simulated patients and perform procedures such as rectal examinations, nasogastric tube insertions, and suturing on bench simulation models. The patients were trained to assess student performance.

Results: There were significant differences between student self-assessments and simulated patient scores for general clinical skills as opposed to procedural skills. Students' mean self-assessments in the data gathering and interpersonal skills were 2–6 % points higher than ratings of their proficiency by simulated patients. However, self-assessments on procedures were 5–8 points lower than patient ratings. The median correlation between self-assessments and patient ratings for general clinical skills such as data gathering and interpersonal skills was 0.08 (not significant), whereas the median correlation between student and patient ratings in procedures was 0.22 ($P < 0.01$).

Conclusions: Third-year medical students' self-assessments for specific procedures are more valid than self-assessments of general clinical skills. Students are less confident in their procedural skills compared with general clinical skills. Although self-assessments should not be used as the sole measure of performance in clinical simulations, self-assessments for specific procedures can provide supplemental information on proficiency.

* *Corresponding author*. Department of Surgery, Sidney Kimmel Medical College at Thomas Jefferson University, 1100 Walnut St. #500, Philadelphia, PA 19107. Tel.: +1 215 955 5877; fax: +1 215 955 2404.
E-mail address: Gerald.Isenberg@jefferson.edu (G.A. Isenberg).
0022-4804/$ – see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2014.09.036

1. Introduction

Self-assessment is a vital skill for all physicians. This is evidenced by the adoption of self-assessment and lifelong learning by the American Board of Medical Specialties as one of four elements in its Maintenance of Certification program [1]. Although self-assessment is the foundation of lifelong learning and self-improvement in medicine, the accuracy of these assessments has not been well supported by empirical research.

A seminal review in the early 1990s called attention to the low validity of knowledge and performance self-assessments. Self-ratings were closely associated with generalized perceptions of self-attributes and were not substantially influenced by test scores or faculty feedback [2]. In the following decade, a review on the state of the art of self-assessment in medical education, as well as higher education in general, also raised questions about the accuracy of self-assessment [3]. More recently, a systematic review of the congruence between physicians' self-assessments and external observations of their competence in the context of continuing education found little evidence to support the value of self-assessments. The authors concluded that external assessments might be more appropriate than self-assessments to support physicians' professional development activities [4].

These findings have been consistent with basic research in the behavioral sciences indicating that "People tend to hold overly favorable views of their abilities in many social and intellectual domains" [5]. Nevertheless, some studies of self-assessment in the context of clinical simulation have yielded encouraging findings. Three studies indicated similar findings of high concordance rates between identical checklists filled out by both standardized patients and students during simulated history and physical sessions [6–8]. A study of self-efficacy ratings for an 11-station, Objective Structured Clinical Examination (OSCE) found that students with high self-efficacy ratings scored higher on the OSCE than those with low self-efficacy ratings [9]. Although one study of first-year residents in clinical simulations did not support the validity of self-assessments [10], three studies of trainees using laparoscopic simulators and suturing with bench simulation models have revealed positive associations with elapsed time and number of errors [11–13].

A recent review of the validity of self-assessments of technical tasks in surgery reported mixed results and recommended that future studies assure that the external assessments used to evaluate self-assessments be reliable and valid, and that appropriate statistical methods be used [14]. In the present study, we examined the validity of third-year students' ratings of their own skills in encounters with simulated patients and simulation devices. Our hypothesis was that self-assessments related to specific procedures (e.g., suturing a laceration, inserting a Foley catheter) would be more accurate than self-assessments of general clinical skills (e.g., data gathering, communicating with patients).

2. Material and methods

2.1. Participants

Participants included 195 third-year students at a large private medical school in the United States in the 2011–2012 academic year. The University's Institutional Review Board approved the study by determining that this use of routine evaluation data collected as part of the educational program was exempt from human subjects review.

2.2. Study design

The study was conducted in the context of a routine OSCE at the end of the third-year surgery clerkship that had been shown to be reliable and valid in relation to faculty ratings of student performance during the clerkship [15]. The students participated in three clinical cases as follows: acute abdominal pain, gastrointestinal bleeding, and a forearm laceration. Each case required that students gather clinical data from simulated patients and perform procedures including rectal examination, intravenous (IV) insertion, nasogastric tube insertion, urinary catheter insertion, and suturing using anatomic and/or nasogastric tissue simulation models. The students performed rectal examinations using a Rectal Examination Model (Life/form Fort Atkinson, WI), inserted IVs using a Multi-venous IV Training Arm (Laerdal, Stavanger, Norway), inserted nasogastric tubes using an NG Tube and Tracheostomy Care Simulator (Laerdal), inserted a urinary catheter using the Advanced Catheterization Trainer (Limbs and Things, St Phillips, Bristol, United Kingdom), and performed suturing using a Suture Pad (Limbs and Things).

The simulated patients were trained to present the cases and to assess student performance at the end of each encounter using checklists and rating scales. The simulated patients were also trained to be able to perform the procedures themselves, so that they could understand how to evaluate the students' performance using criteria that had been described to the students at orientation and at a review session near the end of the clerkship. One of the authors (K.W.B.) evaluated each simulated patient's proficiency in performing each of the selected procedures to confirm that they would be able to accurately assess the students' proficiency.

2.3. Measures of proficiency and self-assessment

Students' proficiency in the general skills and on specific procedures was measured using checklists and rating scales that have been described elsewhere and are similar in format to those used by certifying and licensing boards [15,16]. Checklist scores represented the percentage of tasks completed in relation to each procedure and Likert rating scales were transformed to a 100-point interval grading scale.

Students also completed confidential summative self-assessment ratings at the end of the examination. One set of ratings involved global self-assessments of general clinical