Measurement of the General Competencies of the Accreditation Council for Graduate Medical Education: A Systematic Review

Stephen J. Lurie, MD, PhD, Christopher J. Mooney, MA, and Jeffrey M. Lyness, MD

Abstract

Purpose
To evaluate published evidence that the Accreditation Council for Graduate Medical Education's six general competencies can each be measured in a valid and reliable way.

Method
In March 2008, the authors conducted searches of Medline and ERIC using combinations of search terms "ACGME," "Accreditation Council for Graduate Medical Education," "core competencies," "general competencies," and the specific competencies "systems-based practice" (SBP) and "practice based learning and improvement (PBLI)." Included were all publications presenting new qualitative or quantitative data about specific assessment modalities related to the general competencies since 1999; opinion pieces, review articles, and reports of consensus conferences were excluded. The search yielded 127 articles, of which 56 met inclusion criteria. Articles were subdivided into four categories: (1) quantitative/psychometric evaluations, (2) preliminary studies, (3) studies of SBP and PBLI, and (4) surveys.

Results
Quantitative/psychometric studies of evaluation tools failed to develop measures reflecting the six competencies in a reliable or valid way. Few preliminary studies led to published quantitative data regarding reliability or validity. Only two published surveys met quality criteria. Studies of SBP and PBLI generally operationalized these competencies as properties of systems, not of individual trainees.

Conclusions
The peer-reviewed literature provides no evidence that current measurement tools can assess the competencies independently of one another. Because further efforts are unlikely to be successful, the authors recommend using the competencies to guide and coordinate specific evaluation efforts, rather than attempting to develop instruments to measure the competencies directly.

Acad Med. 2009; 84:301–309.

In February 1999, the Accreditation Council for Graduate Medical Education (ACGME), which is responsible for accrediting all U.S. clinical residency and fellowship programs, unveiled its Outcome Project.[1] This 10-year plan began with a consensus process that defined six general competencies (patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice) thought to be common to physicians training in all specialties . The long-term goal of the Outcome Project is to develop a new model of accreditation based on defining outcomes linked to the six general competencies. Furthermore, because the Outcome Project was created in conjunction with the American Board of Medical Specialties, there is the potential for this model of certification to be extended to ongoing accreditation of U.S. physicians throughout their careers.

This new model was, at least in part, a reaction to a widespread feeling that "medical education seemed to be mired in legions of new requirements," resulting in "a geometric increase in the number of 'musts' and 'shoulds' facing the director of a GME [graduate medical education] program."[2] By contrast, an accreditation model based on general competencies was predicted to "invite creative responses to a challenge rather than prescribing a narrow set of particular responses."[2]

Having defined the six general competencies in a series of discussions with representatives of its constituent organizations, the ACGME then invited program directors to define specific behaviors that would reflect the general competencies in their own specialties. One goal of this project was that appropriate measures of the general competencies would be derived from the needs and insights of those most directly involved in GME, rather than imposed from above by centralized ACGME leadership. Ultimately, it was hoped that such appropriate specification of the general competencies would lead to more rigorous assessment methods: "Program and Institutional Requirements (will) . . . require programs to use increasingly more useful, reliable, and valid methods of assessing residents' attainment of these competency-based objectives."[1]

As a part of this process, the ACGME expected that the Outcome Project would provide a "new challenge for program directors" to "encourag(e) the use of evidence and measurement in the redesign of GME [graduate medical education]." Furthermore, this would help program directors in that

Dr. Lurie is director of assessment, Office of Curriculum and Assessment, University of Rochester School of Medicine and Dentistry, Rochester, New York.

Mr. Mooney is information analyst, Office of Curriculum and Assessment, University of Rochester School of Medicine and Dentistry, Rochester, New York.

Dr. Lyness is director of curriculum, Office of Curriculum and Assessment, University of Rochester School of Medicine and Dentistry, Rochester, New York.

Correspondence should be addressed to Dr. Lurie, University of Rochester School of Medicine and Dentistry, 601 Elmwood Ave, Box 601, Rochester NY, 14642; telephone: (585) 273-4323; e-mail: (Stephen_Lurie@urmc.rochester.edu).

"heretofore their work was viewed as administrative rather than academic, and they were often unsuccessful when they appeared before promotion and tenure committees." The authors concluded that this "legitimate knowledge-building agenda" would "ultimately result . . . in peer-reviewed publications."[2]

According to the Outcome Project timeline,[3] the goal of Phase Two of the project (which was to have occurred between July 2002 and June 2006) was to have involved "sharpening the focus and definition of the [core] competencies and assessment tools." This would then set the stage for Phase Three (July 2006 through June 2011), the goal of which is to achieve "full integration of the [core] competencies and their assessment with learning and clinical care."[3]

The Outcome Project has led to vast changes in evaluative strategy affecting every U.S. postgraduate medical training program (and potentially every practicing U.S. physician). Yet, it remains unclear to what degree the Outcome Project has achieved its stated Phase Two goals of measuring the six general competencies. This is a timely issue, not only because sufficient time has elapsed since the end of Phase Two for resulting literature to appear in print but also because the success of Phase Three seems to be depend, at least in part, on the project having reached its Phase Two goals.

Although the ACGME has published an online toolbox of assessment methodologies,[4] including general psychometric properties of these tools, the document does not comment on how the tools relate to the core competencies. Thus, we sought to evaluate the evidence about whether the six general competencies can currently be measured independently of one another in a valid and reliable way. Indeed, if the six core competencies cannot be measured independently of one another, there would be little practical utility in specifying them as independent criteria of competence. In their description of assessment of the core competencies, the ACGME requires "use of dependable measures to assess residents' competence in patient care, medical knowledge, practice-based learning and improvement, interpersonal and communication skills, professionalism, and systems-based practice." In addition to assessing individual residents, programs are also expected to "use resident performance and outcome assessment results in their evaluation of the educational effectiveness of the residency program."[5] This language implies that the competencies can be measured, at least to some degree, independently of one another for purposes of evaluation.

Assessment of the core competencies has become an immediately pressing issue for residency directors, who must demonstrate attainment of these competencies by their trainees. The concept of competency-based assessment has also been gaining ground both in undergraduate medical education and as a central aspect of ongoing board certification of practicing physicians.[6] Thus, we felt that it was timely to address the question of the reliability and validity with which these competencies can be directly assessed by current measurement tools.

As a secondary question, we sought to evaluate the literature on the two newly defined competencies—systems-based practice (SBP) and problem-based learning and improvement (PBLI). Because these latter two competencies were a particularly innovative aspect of the Outcome Project and did not exist in a formally stated way before 1999, we felt that a complete review of these specific competencies would be achievable within the scope of our study and would further shed light on the new achievements of the Outcome Project. By contrast, the other four competencies have been discussed by medical educators for decades, and each has its own respective and vast literature.

Finally, we sought to assess the nature of the peer-reviewed literature that has addressed the general competencies. This question addresses the ACGME's goal that the Outcome Project would lead to new intellectual activity and publications.

Method

We searched Medline and ERIC using combinations of the search terms "ACGME," "Accreditation Council for Graduate Medical Education," "general competencies," and "core competencies." We also searched on the terms "systems based practice" and "practice based learning and improvement" for publications appearing from 1999 until March 2008.

We then reviewed reference lists of initially identified studies for any studies that were missed by our search. We included publications that presented descriptions of assessment modalities that had been used in specified samples. Because of the very diverse nature of this literature, we felt it would have been inappropriate to have been overly restrictive in this criterion. In addition to including studies that presented psychometric data, we included studies that simply provided narrative accounts. Studies were included if the authors explicitly stated that their aim was to develop and test an assessment tool as it related to the ACGME core competencies and if the article presented any kind of result based on previously unpublished experience in a specific sample. Thus, our inclusion criterion allowed us to exclude opinion pieces, review articles, and reports of consensus conferences because none of these are based on new data relating to the performance of specific tests. We also excluded studies that were not published in peer-reviewed journals.

Results

Our search yielded 127 articles, of which 56 met our inclusion criteria. Because of the exploratory nature of this study, we did not have preconceived ideas of how to organize these studies. Based on our review of the content of these articles, the following four categories emerged as most reflective of the articles that we found: (1) quantitative/psychometric evaluation of the six general competencies, (2) preliminary studies of the general competencies, (3) studies specifically about SBP and PBLI, and (4) surveys or about the general competencies.

Because we had no a priori sense of the kinds of studies we would encounter in the review, we were unable to develop a prespecified quality index. In the cases of survey studies, we judged the quality of work according to the three following traditional criteria: (1) Was there a clear description of the sampling strategy? (2) Did the sample represent a nationally representative sample of the population of interest rather than a local or

convenience sample? and (3) Was the response rate at least 60%?

Finally, we discovered that a number of publications have provided grids describing, in various specialties, which assessment methods would in principle be expected to reflect which of the six general competencies. We examined these grids to assess whether their conclusions were similar to the findings of the studies we reviewed.

Quantitative/psychometric studies of evaluation tools of the six general competencies

Global rating forms. Summary rating forms, which allow faculty to assess trainees' abilities over multiple occasions, are probably the most ubiquitous assessment tools in residency programs. We identified five studies that specifically evaluated the ability of global rating forms to assess the six general competencies (Table 1). These studies have relatively large numbers of participants, which perhaps reflects the widespread nature of these assessment tools. In the largest of these, Silber et al[7] derived items on a global rating scale directly from the language of the general competencies. These authors then determined the scale's structure based on a sample of nearly 1,300 residents. They found that the 23 items on the scale clustered into the two dimensions of medical knowledge and interpersonal skill rather than the six general competencies on which the items were based.

In general, the other four studies also support the conclusion that evaluators cannot distinguish trainees' levels of attainment of the six general competencies in a global rating scale. When individual core competency scores were computed from a global rating form, all six of these scores were significantly related to a written exam,[8] suggesting that the scores were also significantly correlated with one another. Another study[9] found that derived scores (which were related to some but not all of the general competencies) were all significantly correlated with one another. Finally, Reisdorff et al[10] found that all six core-competency scores improved with level of training. In a follow-up analysis,[11] these authors reported that each of the six subscales seemed to be unidimensional in their factor structures, although there was considerable variability across the six competencies. The authors did not analyze the factor structure of all the items considered together, and thus they did not address the extent to which the six scales shared common variance.

360-degree evaluations. In principle, evaluation by colleagues and coworkers provides feedback from persons who may directly observe one another's actual daily behaviors. The method may be further refined by framing the questions in terms of the six core competencies. We identified six studies that provided statistical analysis of such assessment tools (Table 1). These relatively small studies do not provide support for the idea that 360-degree evaluations can be used to distinguish individuals' levels of attainment of the six general competencies. Two of the studies found that all the items clustered on a single factor,[12,13] whereas another[14] found that the items separated into three factors that were not related in a simple way to the six general competencies. One study[15] found that residents and attending physicians had little agreement on ratings of residents' competencies. The other two studies[16,17] did not explicitly look at the degree of concordance between the items and the general competencies.

Direct observation. We found only two studies that directly assessed how well faculty are able to rate learners' general competencies by observing them in specific situations (Table 1). Neither provides compelling evidence that this sort of instrument can be used to assess the general competencies in a valid way. The first[18] found that faculty were able directly to observe fewer than 7% of residents' behaviors in a naturalistic setting. In the second study,[19] raters observed a standardized video of a resident, whom they then rated on the general competencies. No data were presented on the degree to which the derived competency scores were related to one another.

Portfolios. The ACGME has recently launched a project to introduce portfolios into assessment of residents.[20] A portfolio comprises a series of documents that chronicle a learner's evolving competence. Thus, portfolios are appealing not only as summative evaluation tools but also because of the ways that they might guide learners to seek out experiences to help them to develop specific competencies. We were not able to identify any studies of portfolios that specifically sought to measure the ACGME general competencies. Nonetheless, the relatively small literature on portfolios suggests that portfolio scores will not be straightforward to interpret. In their systematic review of studies of portfolios, Carraccio and Englander[21] concluded that "Evidence to date, in studying unstructured portfolios, has demonstrated the difficulty in achieving what is typically considered acceptable standards of reliability and validity in educational measurement." In a subsequent psychometrically rigorous study of a structured portfolio, O'Sullivan et al[22] found that raters had good reliability for judging the overall quality of a portfolio but poor agreement on specific topics (which were derived specifically for psychiatry residents and were, thus, more specific than the ACGME general competencies).

Preliminary studies

We identified 18 peer-reviewed publications that described development or pilot studies of specific assessment tools but that did not provide any quantitative data relating to the tool's reliability or validity.[23–40] Although these articles do not address the reliability or validity of their respective measurement tools, we included them to fully characterize the current state of the literature. All are narrative studies with substantial methodological limitations, including very small sample sizes, lack of quantitative data, or atypical populations.

We discovered that three of these articles resulted in later quantitative follow-up studies.[27,35,36] We then contacted each corresponding author of the remaining studies and enquired whether he or she had any plans to further study the tool in terms of reliability or validity. We received responses from 11 of 15 authors, all of whom told us that they had no plans to further study the instrument they had described.

SBP and PBLI

We identified 14 studies that specifically addressed initiatives to assess the ACGME-defined competencies of SBP and PBLI (Table 2).[26,41–53] Because many of these studies stated that they aimed to

Table 1

Published, Peer-Reviewed Quantitative/Psychometric Evaluation Studies of the Accreditation Council for Graduate Medical Education General Competencies, 1999–2008*

Authors	Subjects	Institution	Measures and results
Global rating forms			
Silber et al (2004)[7]	1295 residents at a single institution	Thomas Jefferson/Albert Einstein	Items derived from the language of the six competencies clustered into two dimensions—medical knowledge and interpersonal skills
Reisdorff et al (2003)[10]	150 emergency medicine residents	Several programs in Michigan	Scores for all six competencies improved between year one and year three. No analysis of relationships among competencies
Reisdorff et al (2004)[11]	150 emergency medicine residents	Several programs in Michigan	Within each of the six competency areas, each group of items had a single major eigenvalue. Items were not combined for an overall analysis
Brasel et al (2004)[9]	36 surgical residents	University of Wisconsin	Factor analysis of preexisting evaluation tool yielded four factors that correspond to four of the six competencies. The four derived scores had correlations ranging from 0.64 to 0.75
Tabuenca (2007)[8]	332 general surgery residents	Multiple institutions	Scores for all six competencies increased with increasing year of training. All six scores correlated significantly with USMLE and in-training exams. There was no analysis of relationship among competency scores
360-degree evaluations			
Musick et al (2003)[17]	18 PM & R† residents	University of Pennsylvania	Descriptive statistics about items means; no analysis of how items relate to competencies
Higgins et al (2004)[16]	6 cardiothoracic surgery residents	Rush Presbyterian–St Luke's	No statistical comparison between the six competency scores. Residents improved on all competencies over time
Massagli et al (2007)[12]	56 PM & R† residents	University of Washington	All items were clustered on a single factor, rather than six
Weigelt et al (2004)[13]	10 residents on trauma service	Medical College of Wisconsin	Average scores across competencies were highly similar. Different raters were unable to distinguish competencies
Roark et al (2006)[15]	26 otolaryngology residents	Consortium of four New York City hospitals	Compared faculty versus peer ratings of six competencies. There were significant correlations for three of the six competencies, but none were significant if corrected for multiple comparisons
Rosenbaum et al (2005)[14]	21 family medicine faculty	University of Iowa	Items derived from the six ACGME competencies yielded acceptable subscale reliability. Factor analysis revealed only three subscales, which were unrelated to the six competencies
Direct observation			
Chisholm et al (2004)[18]	106 emergency medicine residents	Indiana University	In a natural setting, between 3.6% and 6% of resident behaviors were directly observed by faculty
Shayne et al (2006)[19]	82 emergency medicine faculty	16 academic emergency medicine programs	Faculty observed two simulated videos and rated them on five of the six competencies. Raters were internally consistent for all five scales. There was no analysis of relationships among competencies

* See the Method section for a description of inclusion criteria.
† PM & R, physical medicine and rehabilitation.

assess both competencies, we did not further subdivide the studies according to the two competencies. As shown in Table 2, eight of these[26,41–47] involved author-defined quality improvement projects. For each of these studies, the dependent measure was a relevant clinical outcome rather than assessment of participants.

The other six studies[48–53] presented a curriculum or elective opportunity and then measured participants' self-reported confidence or knowledge.

Table 2

Published, Peer-Reviewed Studies of the Accreditation Council for Graduate Medical Education Competencies of Systems-Based Practice and Practice-Based Learning and Improvement, 1999–2008*

Authors	Subjects	Institution	Intervention	Outcome
Coleman et al (2003)[26]	3 teams of residents and attending physicians in family medicine	University of Louisville	Team-specific documentation projects: completion of medication lists, microalbumin screening, and completion of summary sheets	Outcome measure improved on all three projects
Canal et al (2007)[41]	15 surgical residents	Indiana University	Curriculum focusing on deriving a quality-improvement project	Self-reported knowledge, creation of four quality improvement projects
Englander et al (2006)[42]	Numbers of patients or residents not reported	Connecticut Children's Medical Center	Residents identified barrier to use of a standardized lab-ordering machine	Increased use of machine
Frey et al (2003)[43]	12 family medicine residents	Mayo Clinic Scottsdale	Individual projects in senior year	Self-reported knowledge
Miller et al (2006)[44]	110 patients with ventilator-associated pneumonia	Wake Forest University	Derivation of an institution-specific treatment algorithm to improve initial empiric treatment	Rate of appropriate prescribing increased
Mohr et al (2003)[45]	Improvement teams (including 8 residents) in pediatrics	University of Chicago	Identification of five changes in clinic process	Significantly increased immunization rates
Palonen et al (2006)[46]	70 residents in internal medicine and med-peds	University of Alabama	Comparison of chart review versus patient surveys to estimate rates of five clinical behaviors	Both methods yielded similar estimates
Paukert et al (2003)[47]	26 residents and 3 faculty in family practice	University of Texas	Residents and faculty audited 1005 charts	Documetation of preventive health services increased during the study
Rivo et al (2004)[48]	Third- and fourth-year medical students (total number not reported)	Consortium of eight medical schools†	Various curricula	Self-reported behaviors relating to systems-based practice
Siri et al (2007)[49]	4 groups of residents	University of Florida	Residents completed recommendations for four aspects of preoperative care	Self-reported satisfaction
Staton et al (2007)[50]	347 patients with diabetes in a general medicine internal medicine outpatient clinic	East Carolina University	Chart review by residents to improve adherence with foot examination	Improved adherence
Thomas et al (2005)[51]	46 internal medicine residents	May Clinic Rochester	Nonrandom assignment to conference versus small-group discussion of EBM versus no intervention	Small-group participants scored higher on a skills test and self-assessed knowledge
Tomolo et al (2005)[52]	45 internal medicine residents	Cleveland Veterans Affairs Hospital	Residents completed an outcomes card documenting medical errors	Acceptable interrater reliability for identifying types of errors
Weingart et al (2004)[53]	26 internal medicine residents	Beth Israel—Deaconess	3-week elective in quality improvement	Self-reported knowledge

* See the Method section for a description of inclusion criteria.
† Names of schools not explicitly described in article.

Other studies about the general competencies

Surveys. We identified 11 published studies that described surveys about the ACGME competencies with varying samples and response rates (Table 3).[54–64] These studies are difficult to summarize because of differences in methodology and populations studied. Only two of these studies met all three of our quality criteria for surveys.[56,60] In the first of these,[56] family medicine program directors consistently rated SBP and PBLI as their lowest educational priorities. Similar rankings of self-rated competency were found among physicians who had completed an allergy and immunology fellowship in the United States between 1995 and 2000.[60]

Grids. We identified seven publications in which authors developed grids that cross-referenced available assessment tools with the six competencies.[31,38,65–69] In general, the purpose of these publications is to develop a checklist of which general competencies can reasonably be assessed with which methodologies. In every case, multiple assessment methods mapped onto multiple general competencies. Thus, at a conceptual level, it did not seem that experts were able to define measurement tools that uniquely capture the general competencies, or general competencies that are unique to assessment methods.

Discussion

We find that the literature to date has not yielded any method that can assess the six

Table 3

Published, Peer-Reviewed Survey Studies of the Accreditation Council for Graduate Medical Education General Competencies, 1999–2008*

Authors	Sample	Response rate, no. (%)
Cogbill et al (2005)[54]	Psychiatry residents at University of Arkansas	16 of 23 (70)
Collins et al (2004)[55]	All U.S. radiology program directors	99 of 192 (52)
Delzell et al (2005)[56]	All U.S. family medicine program directors	287 of 444 (65)
Heard et al (2002)[57]	Program directors at the University of Arkansas	24 of 47 (51)
Johnson and Barratt (2005)[58]	Pediatric Continuity Clinic preceptors	336 of 2378 (14)
Joyner et al (2005)[59]	Urology program directors	105 of 119 (88)
Li et al (2003)[60]	Physicians completing U.S. allergy fellowship 1995–2001	253 of 373 (68)
Lynch et al (2003)[61]	National sample of family physicians	1,228 of 2,363 (54)
Michels et al (2007)[62]	Ophthalmologists in pacific northwest	147 of 676 (22)
Stiles et al (2006)[63]	Surgical residents at a single institution	25 of 25 (100)
Wald et al (2007)[64]	Undergraduate emergency medicine clerkship directors	92 of 132 (70)

ACGME general competencies as independent constructs. Rather, all currently available measurement tools generally yield a single dimension of overall measured competency or, sometimes, several measured dimensions that do not relate to the competencies in a simple manner. This lack of simple correspondence between the general competencies and measurement is mirrored in the several published attempts to conceptually map the general competencies onto observable behaviors—such attempts consistently yield grids in which all possible measurable behaviors consistently map onto three or more of the general competencies. Scores obtained by any of the currently available assessment tools represent various admixtures of the underlying hypothetical general competencies. That is, it currently does not seem possible to "measure the competencies" independently of one another in any psychometrically meaningful way.

In terms of our goal of characterizing the existing literature that has grown up around the ACGME competencies, we find that only 13 of 127 (10%) of published studies presented any psychometric data on assessment tools. Another 14 of these (11%) presented descriptions of interventions to assess PBLI or SBP, although not all of these were psychometrically rigorous. Of the 127 studies we identified, 18 of these presented preliminary data, although 15 of them (12%) did not have any subsequent follow-up publications. Finally, 11 of these studies (9%) presented survey data, although few of them met rigorous standards. The remaining 71 publications (56%) represent consensus conferences, editorials, thought pieces, etc.

The exception to this challenge of measuring competencies seems to be "medical knowledge." This competency is generally measured with written examinations in which the examinee answers a series of standardized questions that assess factual knowledge. Recently, this approach has been expanded with the use of script-concordance tests, which offer examinees a series of choices that attempt to mirror real-world decision making, and in which examinees' scores are determined by their degree of concordance with the responses of a reference panel of medical experts.[70–73] Because this technique assesses application of knowledge in typical clinical conditions of uncertainty, these tests seem to fulfill the ACGME's requirement that trainees demonstrate "application of knowledge to patient care." Furthermore, it has been shown that, in a large sample of physicians, paper-and-pencil tests of knowledge have significant relationships to later markers of quality of clinical care.[74,75] Thus, these measures, which reliably assess medical knowledge, also seem to be valid predictors of important later clinical behaviors. Much of this success seems to be a reflection of the way that "medical knowledge" is composed of a very large series of identifiable facts and relationships among facts, the veracity of which can be independently assessed.

By contrast, the other five competencies reflect, in varying degrees, personal attributes of trainees rather than knowledge of objectively derived information. Furthermore, the relative values of these attributes are more socially and culturally determined than are the abilities comprising "medical knowledge." Thus, to date, these competencies have proven considerably more challenging to quantify in a reliable and valid way. Although we did not systematically survey the literature on these additional competencies, each has been the subject of several prior review articles, which we believe are helpful for providing additional context. For instance, the construct of "professionalism," which predated the ACGME general competencies, has continued to defy a clear operational definition despite several decades of attempts to derive one. In addition to deep philosophical differences over the various possible meanings of the term "professionalism," the inherent challenges of measurement and psychometric analyses add additional layers of uncertainty. In her systematic review of measurement of professionalism, Arnold[76] concluded that "interrater agreement on humanistic terms can be particularly low." Even if raters could agree on how to judge particular items relating to such a high-order construct as "professionalism," relationships among items seem unstable; depending on the measurement tool chosen, a purely empirical definition of "professionalism" may contain as few as three subscales[77,78] or as many as seven[79]

or eight.[80] Thus, at a measurement level, the meaning of "professionalism" becomes mired in the technical minutiae of psychometric analysis, irrespective of any philosophical beliefs about the nature of the construct itself.

On the basis of our results, we suspect that such concerns will likely continue to thwart attempts at measurement of the other general competencies as well. This is not because the general competencies are, in any sense, "incorrect"; rather, it is a reflection of the Outcome Project's assumption that the general competencies, once defined, would reveal themselves in a straightforward fashion through measurement. It will remain a challenge to develop objective measures that correspond neatly to these generalized educational constructs. In addition to disagreements over theoretical issues, measurement of actual human behaviors is subject to a host of nontheoretical biases and technical challenges, including the well-known psychometric problems of method variance, observer biases, expectation and contextual effects, logistical constraints, and random error.

It would be unfortunate, however, if these failures of quantification were to lead to cynicism about the general competencies or to the conclusion that such principles are of no practical value. As initially conceived by the leadership of the ACGME, the general competencies were meant as a response to "overspecification" of training and assessment requirements. Although we agree with this concern in principle, we feel that the problem was not so much *over*specification (because measurement requirements must always be stated with some specificity) but, rather, a lack of *coherent* specification. Without an overarching set of principles, a list of detailed requirements runs the risk of seeming random and arbitrary. Thus, the general competencies could have an invaluable role in guiding assessment strategy as long as it is clear that the six general competencies themselves exist in a realm outside of measurement. What remains missing from the Outcome Project, in our view, is an explicitly stated set of expectations that would link the ideals of the general competencies to the realities of measurement. Thus, a next step in development of an overall theory of assessment would not be to abandon the general competencies but, rather, to explicitly develop a more fully elaborated model to rationalize and prioritize various assessment tools in light of the general competencies. Although it is possible that such a measurement model could arise from the kind of grassroots effort proposed in the Outcome Project, we suspect that this will need to come from further consensus and deliberation by the ACGME and its constituent organizations.

As one contribution to the development of such a model, we find that the two newer ACGME competencies—SBP and PBLI—are viewed by many authors as representing aspects of health systems and teams rather than those of particular individuals. Thus, it is possible that environmental variables may exert significant influence on trainees' behaviors surrounding these competences. It is possible, for instance, that a trainee with relatively good understanding of systems-based issues may nonetheless seem to perform poorly when placed in a practice environment that hinders good communication among caregivers. Further refinements of the operational definitions of these competencies should include measures of health systems in addition to any measures of individuals.

Our study has several limitations. First, we did not assess conference presentations, posters, or other unpublished material. We recognize that much communication among program directors, as well as between program directors and the ACGME, occurs on this informal, face-to-face level. It is possible that we may have missed important additional information that was communicated in this way. Nonetheless, we deliberately chose not to examine such material in light of the ACGME's stated intent that the competencies would result in enhanced scientific activity, which implies publication and peer review. Second, because of the ongoing nature of the Outcome Project, it is possible that our review failed to reflect studies that may be currently ongoing. We suspect, however, that we did not miss a significant number of these because we contacted all authors who had previously published preliminary or pilot descriptions of assessment projects. Finally, we did not consult officials of the ACGME in preparing our review.

Despite these difficulties, we recognize that attention to the six ACGME competencies has already led to some of their intended benefits. For example, many residency programs now have additional curricular time and effort devoted to areas such as interpersonal and communications skills, which were previously perceived to be lacking in many training programs. Future assessment methodologies should incorporate these beneficial attributes while striving to define assessments that can be measured reliably and, thus, to provide empirical benchmarks for further educational reform.

Acknowledgments

The authors thank Diane M. Hartmann, MD, and David R. Lambert, MD, for their many thoughtful comments on this work.

References

1 Accreditation Council for Graduate Medical Education. The ACGME Outcome Project: An Introduction. Available at: (http://www.acgme.org/outcome/project/OPintrorev1_7–05.ppt). Accessed November 13, 2008.
2 Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. Health Aff (Millwood). 2002;21:103–111.
3 Accreditation Council for Graduate Medical Education. ACGME Outcomes Project. Timeline—Working Guidelines. Available at: (http://www.acgme.org/outcome/project/timeline/TIMELINE_index_frame.htm). Accessed November 13, 2008.
4 Accreditation Council for Graduate Medical Education. Toolbox of Assessment Methods. Available at: (http://www.acgme.org/Outcome/assess/Toolbox.pdf). Accessed November 17, 2008.
5 Accreditation Council for Graduate Medical Education. ACGME Outcome Project. Available at: (http://www.acgme.org/outcome/comp/compMin.asp). Accessed November 13, 2008.
6 American Board of Medical Specialties. MOC competencies and criteria. Available at: (http://www.abms.org/Maintenance_of_Certification/MOC_competencies.aspx). Accessed November 13, 2008.
7 Silber CG, Nasca TJ, Paskin DL, Eiger G, Robeson M, Veloski JJ. Do global rating forms enable program directors to assess the ACGME competencies? Acad Med. 2004;79:549–556.
8 Tabuenca A, Welling R, Sachdeva AK, et al. Multi-institutional validation of a Web-based core competency assessment system. J Surg Educ. 2007;64:390–394.
9 Brasel KJ, Bragg D, Simpson DE, Weigelt JA. Meeting the accreditation council for graduate medical education competencies using established residency training program assessment tools. Am J Surg. 2004;188:9–12.
10 Reisdorff EJ, Hayes OW, Reynolds B, et al. General competencies are intrinsic to

emergency medicine training: A multicenter study. Acad Emerg Med. 2003;10:1049–1053.
11 Reisdorff EJ, Carlson DJ, Reeves M, Walker G, Hayes OW, Reynolds B. Quantitative validation of a general competency composite assessment evaluation. Acad Emerg Med. 2004;11:881–884.
12 Massagli TL, Carline JD. Reliability of a 360-degree evaluation to assess resident competence. Am J Phys Med Rehabil. 2007; 86:845–852.
13 Weigelt JA, Brasel KJ, Bragg D, Simpson D. The 360-degree evaluation: Increased work with little return? Curr Surg. 2004;61:616–626.
14 Rosenbaum ME, Ferguson KJ, Kreiter CD, Johnson CA. Using a peer evaluation system to assess faculty performance and competence. Fam Med. 2005;37:429–433.
15 Roark RM, Schaefer SD, Yu GP, Branovan DI, Peterson SJ, Lee WN. Assessing and documenting general competencies in otolaryngology resident training programs. Laryngoscope. 2006;116:682–695.
16 Higgins RS, Bridges J, Burke JM, O'Donnell MA, Cohen NM, Wilkes SB. Implementing the ACGME general competencies in a cardiothoracic surgery residency program using 360-degree feedback. Ann Thorac Surg. 2004;77:12–17.
17 Musick DW, McDowell SM, Clark N, Salcido R. Pilot study of a 360-degree assessment instrument for physical medicine & rehabilitation residency programs. Am J Phys Med Rehabil. 2003;82:394–402.
18 Chisholm CD, Whenmouth LF, Daly EA, Cordell WH, Giles BK, Brizendine EJ. An evaluation of emergency medicine resident interaction time with faculty in different teaching venues. Acad Emerg Med. 2004;11: 149–155.
19 Shayne P, Gallahue F, Rinnert S, et al. Reliability of a core competency checklist assessment in the emergency department: The standardized direct observation assessment tool. Acad Emerg Med. 2006;13:727–732.
20 Accreditation Council for Graduate Medical Education. ACGME Learning Portfolio: A Professional Development Tool. Available at: (http://www.acgme.org/acWebsite/portfolio/cbpac_faq.pdf). Accessed November 13, 2008.
21 Carraccio C, Englander R. Evaluating competence using a portfolio: A literature review and Web-based application to the ACGME competencies. Teach Learn Med. 2004;16:381–387.
22 O'Sullivan PS, Reckase MD, McClain T, Savidge MA, Clardy JA. Demonstration of portfolios to assess competency of residents. Adv Health Sci Educ Theory Pract. 2004;9: 309–323.
23 Alexander M, Pavlov A, Lenahan P. Lights, camera, action: Using film to teach the ACGME competencies. Fam Med. 2007;39: 20–23.
24 Carraccio C, Englander R, Wolfsthal S, Martin C, Ferentz K. Educating the pediatrician of the 21st century: Defining and implementing a competency-based system. Pediatrics. 2004;113:252–258.
25 Clay AS, Petrusa E, Harker M, Andolsek K. Development of a Web-based, specialty specific portfolio. Med Teach. 2007;29:311–316.
26 Coleman MT, Nasraty S, Ostapchuk M, Wheeler S, Looney S, Rhodes S. Introducing practice-based learning and improvement ACGME core competencies into a family medicine residency curriculum. Jt Comm J Qual Saf. 2003;29:238–247.
27 Dickey J, Girard DE, Geheb MA, Cassel CK. Using systems-based practice to integrate education and clinical services. Med Teach. 2004;26:428–434.
28 Dorotta I, Staszak J, Takla A, Tetzlaff JE. Teaching and evaluating professionalism for anesthesiology residents. J Clin Anesth. 2006; 18:148–160.
29 Greenberg JA, Irani JL, Greenberg CC, et al. The ACGME competencies in the operating room. Surgery. 2007;142:180–184.
30 Hayes OW, Reisdorff EJ, Walker GL, Carlson DJ, Reinoehl B. Using standardized oral examinations to evaluate general competencies. Acad Emerg Med. 2002;9:1334–1337.
31 Johnston KC. Responding to the ACGME's competency requirements: An innovative instrument from the University of Virginia's neurology residency. Acad Med. 2003;78: 1217–1220.
32 Lyman J, Schorling J, May N, et al. Customizing a clinical data warehouse for housestaff education in practice-based learning and improvement. AMIA Annu Symp Proc. 2006:1017.
33 Oetting TA, Lee AG, Beaver HA, et al. Teaching and assessing surgical competency in ophthalmology training programs. Ophthalmic Surg Lasers Imaging. 2006;37: 384–393.
34 O'Sullivan PS, Cogbill KK, McClain T, Reckase MD, Clardy JA. Portfolios as a novel approach for residency evaluation. Acad Psychiatry. 2002;26:173–179.
35 Reisdorff EJ, Hayes OW, Carlson DJ, Walker GL. Assessing the new general competencies for resident education: A model from an emergency medicine program. Acad Med. 2001;76:753–757.
36 Shayne P, Heilpern K, Ander D, Palmer-Smith V; Emory University Department of Emergency Medicine Education Committee. Protected clinical teaching time and a bedside clinical evaluation instrument in an emergency medicine training program. Acad Emerg Med. 2002;9:1342–1349.
37 Simpson D, Helm R, Drewniak T, et al. Objective structured video examinations (OSVEs) for geriatrics education. Gerontol Geriatr Educ. 2006;26:7–24.
38 Torbeck L, Wrightson AS. A method for defining competency-based promotion criteria for family medicine residents. Acad Med. 2005;80:832–839.
39 Triola MM, Feldman HJ, Pearlman EB, Kalet AL. Meeting requirements and changing culture. The development of a Web-based clinical skills evaluation system. J Gen Intern Med. 2004;19:492–495.
40 Webb TP, Aprahamian C, Weigelt JA, Brasel KJ. The surgical learning and instructional portfolio (SLIP) as a self-assessment educational tool demonstrating practice-based learning. Curr Surg. 2006;63:444–447.
41 Canal DF, Torbeck L, Djuricich AM. Practice-based learning and improvement: A curriculum in continuous quality improvement for surgery residents. Arch Surg. 2007;142:479–482.
42 Englander R, Agostinucci W, Zalneraiti E, Carraccio CL. Teaching residents systems-based practice through a hospital cost-reduction program: A "win-win" situation. Teach Learn Med. 2006;18:150–152.
43 Frey K, Edwards F, Altman K, Spahr N, Gorman RS. The "collaborative care" curriculum: An educational model addressing key ACGME core competencies in primary care residency training. Med Educ. 2003;37: 786–789.
44 Miller PR, Partrick MS, Hoth JJ, Meredith JW, Chang MC. A practical application of practice-based learning: Development of an algorithm for empiric antibiotic coverage in ventilator-associated pneumonia. J Trauma. 2006;60:725–729.
45 Mohr JJ, Randolph GD, Laughon MM, Schaff E. Integrating improvement competencies into residency education: A pilot project from a pediatric continuity clinic. Ambul Pediatr. 2003;3:131–136.
46 Palonen KP, Allison JJ, Heudebert GR, et al. Measuring resident physicians' performance of preventive care: Comparing chart review with patient survey. J Gen Intern Med. 2006; 21:226–230.
47 Paukert JL, Chumley-Jones HS, Littlefield JH. Do peer chart audits improve residents' performance in providing preventive care? Acad Med. 2003;78(10 suppl):S39–S41.
48 Rivo ML, Keller DR, Teherani A, O'Connell MT, Weiss BA, Rubenstein SA. Practicing effectively in today's health system: Teaching systems-based care. Fam Med. 2004;36 (suppl):S63–S67.
49 Siri J, Reed AI, Flynn TC, Silver M, Behrns KE. A multidisciplinary systems-based practice learning experience and its impact on surgical residency education. J Surg Educ. 2007;64:328–332.
50 Staton LJ, Kraemer SM, Patel S, Talente GM, Estrada CA. "Correction" peer chart audits: A tool to meet accreditation council on graduate medical education (ACGME) competency in practice-based learning and improvement. Implement Sci. 2007;2:24.
51 Thomas KG, Thomas MR, York EB, Dupras DM, Schultz HJ, Kolars JC. Teaching evidence-based medicine to internal medicine residents: The efficacy of conferences versus small-group discussion. Teach Learn Med. 2005;17:130–135.
52 Tomolo A, Caron A, Perz ML, Fultz T, Aron DC. The outcomes card. Development of a systems-based practice educational tool. J Gen Intern Med. 2005;20:769–771.
53 Weingart SN, Tess A, Driver J, Aronson MD, Sands K. Creating a quality improvement elective for medical house officers. J Gen Intern Med. 2004;19:861–867.
54 Cogbill KK, O'Sullivan PS, Clardy J. Residents' perception of effectiveness of twelve evaluation methods for measuring competency. Acad Psychiatry. 2005;29: 76–81.
55 Collins J, Herring W, Kwakwa F, et al. Current practices in evaluating radiology residents, faculty, and programs: Results of a survey of radiology residency program directors. Acad Radiol. 2004;11:787–794.
56 Delzell JE Jr, Ringdahl EN, Kruse RL. The ACGME core competencies: A national

survey of family medicine program directors. Fam Med. 2005;37:576–580.
57 Heard JK, Allen RM, Clardy J. Assessing the needs of residency program directors to meet the ACGME general competencies. Acad Med. 2002;77:750.
58 Johnson CE, Barratt MS. Continuity clinic preceptors and ACGME competencies. Med Teach. 2005;27:463–467.
59 Joyner BD, Siedel K, Stoll D, Mitchell M. Report of the national survey of urology program directors: Attitudes and actions regarding the accreditation council for graduate medical education regulations. J Urol. 2005;174:1961–1968.
60 Li JT, Stoll DA, Smith JE, Lin JJ, Swing SR. Graduates' perceptions of their clinical competencies in allergy and immunology: Results of a survey. Acad Med. 2003;78:933–938.
61 Lynch DC, Pugno P, Beebe DK, Cullison SW, Lin JJ. Family practice graduate preparedness in the six ACGME competency areas: Prequel. Fam Med. 2003;35:324–329.
62 Michels KS, Hansel TE, Choi D, Lauer AK. A survey of desired skills to acquire in ophthalmology training: A descriptive statistical analysis. Ophthalmic Surg Lasers Imaging. 2007;38:107–114.
63 Stiles BM, Reece TB, Hedrick TL, et al. General surgery morning report: A competency-based conference that enhances patient care and resident education. Curr Surg. 2006;63:385–390.
64 Wald DA, Manthey DE, Kruus L, Tripp M, Barrett J, Amoroso B. The state of the clerkship: A survey of emergency medicine clerkship directors. Acad Emerg Med. 2007;14:629–634.
65 Bingham JW, Quinn DC, Richardson MG, Miles PV, Gabbe SG. Using a healthcare matrix to assess patient care in terms of aims for improvement and core competencies. Jt Comm J Qual Patient Saf. 2005;31:98–105.
66 Chapman DM, Hayden S, Sanders AB, et al. Integrating the Accreditation Council for Graduate Medical Education core competencies into the model of the clinical practice of emergency medicine. Ann Emerg Med. 2004;43:756–769.
67 Jarvis RM, O'Sullivan PS, McClain T, Clardy JA. Can one portfolio measure the six ACGME general competencies? Acad Psychiatry. 2004;28:190–196.
68 Singh R, Naughton B, Taylor JS, et al. A comprehensive collaborative patient safety residency curriculum to address the ACGME core competencies. Med Educ. 2005;39:1195–1204.
69 Wang EE, Vozenilek JA. Addressing the systems-based practice core competency: A simulation-based curriculum. Acad Emerg Med. 2005;12:1191–1194.
70 Charlin B, Gagnon R, Pelletier J, et al. Assessment of clinical reasoning in the context of uncertainty: The effect of variability within the reference panel. Med Educ. 2006;40:848–854.
71 Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The script concordance test: A tool to assess the reflective clinician. Teach Learn Med. 2000;12:189–195.
72 Charlin B, van der Vleuten C. Standardized assessment of reasoning in contexts of uncertainty: The script concordance approach. Eval Health Prof. 2004;27:304–319.
73 Sibert L, Darmoni SJ, Dahamna B, Weber J, Charlin B. Online clinical reasoning assessment with the script concordance test: A feasibility study. BMC Med Inform Decis Mak. 2005;5:18.
74 Tamblyn R, Abrahamowicz M, Dauphinee WD, et al. Association between licensure examination scores and practice in primary care. JAMA. 2002;288:3019–3026.
75 Tamblyn R, Abrahamowicz M, Dauphinee D, et al. Physician scores on a national clinical skills examination as predictors of complaints to medical regulatory authorities. JAMA. 2007;298:993–1001.
76 Arnold L. Assessing professional behavior: Yesterday, today, and tomorrow. Acad Med. 2002;77:502–515.
77 Arnold EL, Blank LL, Race KE, Cipparrone N. Can professionalism be measured? The development of a scale for use in the medical environment. Acad Med. 1998;73:1119–1121.
78 DeLisa JA, Foye PM, Jain SS, Kirshblum S, Christodoulou C. Measuring professionalism in a physiatry residency training program. Am J Phys Med Rehabil. 2001;80:225–229.
79 Blackall GF, Melnick SA, Shoop GH, et al. Professionalism in medical education: The development and validation of a survey instrument to assess attitudes toward professionalism. Med Teach. 2007;29:e58–e62.
80 Tsai TC, Lin CH, Harasym PH, Violato C. Students' perception on medical professionalism: The psychometric perspective. Med Teach. 2007;29:128–134.