Stephen D. Brown[1,2]
Elizabeth A. Rider[2,3]
Katherine Jamieson[4]
Elaine C. Meyer[2,5]
Michael J. Callahan[1]
Carolynn M. DeBenedectis[6]
Sarah D. Bixby[1]
Michele Walters[1]
Sara F. Forman[7]
Pamela H. Varrin[2,8]
Peter Forbes[9]
Christopher J. Roussin[4,10]

Keywords: communication assessment, communication competency, communication education, radiology, simulation

DOI:10.2214/AJR.16.17439

Received September 19, 2016; accepted after revision November 26, 2016.

Supported in part by a 2011 RSNA Education Scholar Grant.

[1]Department of Radiology, Boston Children's Hospital and Harvard Medical School, 300 Longwood Ave, Boston, MA 02115. Address correspondence to S. D. Brown (Stephen.brown@childrens.harvard.edu).

[2]Institute for Professionalism and Ethical Practice, Boston Children's Hospital, Boston, MA.

[3]Department of Medicine, Division of General Pediatrics, Boston Children's Hospital and Harvard Medical School, Boston, MA.

[4]Simulator Program, Boston Children's Hospital, Boston, MA.

[5]Department of Psychiatry, Harvard Medical School, Boston, MA.

[6]Department of Radiology, University of Massachusetts Medical School, Worcester, MA.

[7]Department of Medicine, Division of Adolescent Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA.

[8]Cotting School, Lexington, MA.

[9]Clinical Research Program, Biostatistics Core, Boston Children's Hospital, Boston, MA.

[10]Department of Anesthesiology, Perioperative and Pain Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA.

Supplemental Data
Available online at www.ajronline.org.

This article is available for credit.

AJR 2017; 209:351–357

0361–803X/17/2092–351

Development of a Standardized Kalamazoo Communication Skills Assessment Tool for Radiologists: Validation, Multisource Reliability, and Lessons Learned

OBJECTIVE. The purpose of this study was to develop and test a standardized communication skills assessment instrument for radiology.

MATERIALS AND METHODS. The Delphi method was used to validate the Kalamazoo Communication Skills Assessment instrument for radiology by revising and achieving consensus on the 43 items of the preexisting instrument among an interdisciplinary team of experts consisting of five radiologists and four nonradiologists (two men, seven women). Reviewers assessed the applicability of the instrument to evaluation of conversations between radiology trainees and trained actors portraying concerned parents in enactments about bad news, radiation risks, and diagnostic errors that were video recorded during a communication workshop. Interrater reliability was assessed by use of the revised instrument to rate a series of enactments between trainees and actors video recorded in a hospital-based simulator center. Eight raters evaluated each of seven different video-recorded interactions between physicians and parent-actors.

RESULTS. The final instrument contained 43 items. After three review rounds, 42 of 43 (98%) items had an average rating of relevant or very relevant for bad news conversations. All items were rated as relevant or very relevant for conversations about error disclosure and radiation risk. Reliability and rater agreement measures were moderate. The intraclass correlation coefficient range was 0.07–0.58; mean, 0.30; SD, 0.13; and median, 0.30. The range of weighted kappa values was 0.03–0.47; mean, 0.23; SD, 0.12; and median, 0.22. Ratings varied significantly among conversations ($\chi^2_6 = 1186$; $p < 0.0001$) and varied significantly by viewing order, rater type, and rater sex.

CONCLUSION. The adapted communication skills assessment instrument is highly relevant for radiology, having moderate interrater reliability. These findings have important implications for assessing the relational competencies of radiology trainees.

The American Board of Radiology and Accreditation Council for Graduate Medical Education (ACGME) include interpersonal and communication skills among the radiology-specific core competencies that must be taught and assessed during radiologic training [1–3]. Graduating trainees are expected to show competence in communicating "complex and difficult information, such as errors, complications, adverse events, and bad news" [1].

To our knowledge, no standardized validated method exists to assess such competencies within radiology. A number of radiology programs have implemented innovative communication skills exercises for their trainees that use either direct engagement with actual patients or simulations with professional actors or trained patients [4–8]. However, each program has applied somewhat different metrics for rating trainees' communication skills. The ACGME, American Board of Radiology, and others have advised that numerous approaches exist for assessing communication skills, and it seems likely that no single method will capture the full array of relational qualities that characterize patient-related communication competency [1, 2, 9]. Nonetheless, now that the ACGME Next accreditation system requires programs to provide summary reports for their residents, the development of standardized, electronic, and generalizable core competency evaluation tools has been recommended to facilitate the broadly available ACGME Milestones assessment and analysis [10].

One potential tool for achieving standardization for communication skills is the Gap-Kalamazoo Communication Skills Assessment Form (GKCSAF), a Kalamazoo Consensus Statement assessment instrument [11–

13]. The Kalamazoo Consensus Statement was developed by 21 medical education leaders and communication experts from the United States and Canada [14]. The group delineated a set of essential elements in physician-patient communication to facilitate teaching and assessment of communication skills at all levels of medical education and subsequently created the Kalamazoo Essential Elements Communication Checklist [13]. Two additional iterations of the Kalamazoo checklist followed: the Kalamazoo Essential Elements Communication Checklist—Adapted and the GKCSAF [11–13, 15–17]. The original Essential Elements Communication Checklist and the Kalamazoo Essential Elements Communication Checklist—Adapted (modification with a Likert scale) identified seven essential communication competency domains, each of which incorporated multiple specific subcompetencies [13, 14, 17]. The GKCSAF subsequently recognized nine essential domains with 34 subdomains, and it entailed a multirater method with gap analysis, which is used for individual, interdisciplinary, and team assessments [11–13, 15].

The Kalamazoo instruments have been broadly used in medical education and have been validated for various clinical circumstances [11–15, 17]. Some radiology programs have used partial items from the instrument [6] or used a Kalamazoo instrument to establish its proof of concept as a potential tool [8]. However, the instrument was developed to assess communication skills outside of radiologic practice, and its validity has not been determined specifically for radiology. This would be a key next step if the instrument is to be adapted broadly for radiology, insofar as the conversations that radiologists have with patients and the settings in which they are conducted are distinct from other clinical realms.

The purposes of this study were to adapt a communication skills assessment tool well established elsewhere in medicine—the GKCSAF—for relevance specifically to radiology and to test its reliability in a standardized setting. We first validated the tool by applying the Delphi method to revise and achieve consensus on the various domains of the instrument among an interdisciplinary team of carefully chosen individuals. These experts assessed the applicability of the instrument to the evaluation of conversations between radiology trainees and trained actors portraying concerned parents using enactments video recorded during a communication workshop. We then assessed interrater reliability by using the revised instrument to rate a series of enactments between trainees and actors video recorded in a hospital-based simulator center.

Materials and Methods

The Boston Children's Hospital institutional review board determined that this project represented a departmental quality assurance and performance improvement initiative and was therefore exempt from review. We validated the adapted tool for radiology in two phases: three rounds of relevance testing using the Delphi method and one round of testing of interrater reliability.

Phase 1: Adaptation and Validation of Relevance for Radiology

In phase 1 of our study, we started with the GKCSAF, using the Delphi method to determine the relevance of the GKCSAF specifically as an evaluation instrument for radiologist-to-parent communication about bad news and to revise it accordingly. The Delphi method is a process of organizing consensus on a focused domain and has been used widely for generation of assessment tools [18–20].

For each round in phase 1, nine independent reviewers were asked to watch a video recording of a simulated bad news conversation between a radiologist and actors portraying the parents of an infant. The reviewers were two pediatric radiologists and a breast imaging specialist, all with substantial experience with simulation-based communication skills training; two pediatric radiology training program directors; an adolescent medicine specialist with fellowship training in medical humanism and professionalism; a clinical psychologist at a school for children with special needs who had substantial experience with simulation-based communication skills training for health care providers; a critical care nurse-psychologist who founded and directs an interdisciplinary health care communication skills training institute; and a pediatrician expert on medical education, physician competency assessment, health care communication, and medical humanism who was a primary architect and adapter of the GKCSAF.

For each separate round of phase 1, a different video enactment of the same bad-news scenario was shown, each featuring a different radiology trainee. In the scenario, the radiologist conveys to the actor-parents that their infant's abdominal ultrasound shows a probable liver cancer. The videos were previously recorded at workshops of the Boston Children's Hospital Program to Enhance Relational and Communication Skills for radiologists [4]. All video-recorded radiologists provided consent to having the videos used for educational program development.

For each round of phase 1, after viewing the video together, the reviewers used a 5-point Likert scale (1, poor; 2, fair; 3, good; 4, very good; 5, excellent) to independently score the relevance of each GKCSAF domain and subdomain as they pertained to the video-recorded conversation. Reviewers also provided written comments on the items and suggestions for new items. Once they completed their comments independently, the reviewers discussed the instrument, domains, subdomains, and the scenario as a group, and the discussion group conversation minutes were recorded.

In total, there were three rounds of review. After each round, the results of the reviewer ratings were combined and analyzed with the SPSS statistical program (version 23.0 for Apple Macintosh, IBM). Domains and subdomains were assessed as relevant if they were rated, on average, as 4 or higher on the 5-point scale. Most items with consistent ratings of relevance were left unchanged, and small edits were made to a few items according to discussion group deliberations. All items rated less than 4 were edited primarily with language that seemed more relevant for radiology, and some changes were made to the order of domains or the placement of subdomains. No further revisions were made after the second round.

In addition, during the third round, raters were asked to assess the relevance of the instrument for assessing communication competencies for two additional and different video-recorded conversations. In one conversation a radiology trainee discussed with an anxious parent the risks of radiation from a CT examination to be performed on a 7-year-old child with suspected appendicitis. In the other, a radiologist discussed a missed ultrasound finding that resulted in a 3-month delay in a cancer diagnosis. As with the original bad news scenario, these videos were recorded during Program to Enhance Relational and Communication Skills for radiologists workshops.

Phase 2: Assessing Interrater Reliability of the Adapted Instrument

We named the adapted instrument the Kalamazoo Communication Skills Assessment Tool—Radiology (KalRad). Once the relevance of KalRad was established for the bad-news radiology scenario, we assessed variation among raters (interrater reliability) in using the tool to assess communication competencies among a cohort of radiology trainees. Using the Boston Children's Hospital simulator program, seven pediatric radiology fellows using a different bad news scenario from that in phase 1 were independently recorded in simulated enactments with professional actors portraying the child's parents. In this scenario, the physicians explained to the parents of an 8-year-old boy that a posterior fossa mass and hydrocephalus were found on an outpatient MRI examination performed because the boy had headaches. The

seven different interactions between physicians and actor-parents were video recorded. Eight of the nine original team members viewed the video-recorded enactments and used KalRad to rate the fellows' communication skills. Raters completed the surveys independently of one another. Surveys were administered to raters in both paper-based and electronic form, according to rater preference. Data were entered and analyzed with SPSS version 23.0 statistical software.

Statistical Methods

Each of the eight raters watched all seven video-recorded conversations and rated the quality of each radiologist's communication in 43 domains and subdomains. Rater agreement for each item was assessed with the Fleiss weighted kappa statistic for multiple raters and intraclass correlation coefficient. Generalized estimating equations (Proc GENMOD, SAS version 9.3, SAS Institute), a regression method appropriate for correlated observations, was used to assess the effect of each of the following on mean rating: video conversation (n = 7), viewing order at two levels (conversations 1 to 7 [n = 5 raters], conversations 7 to 1 [n = 3 raters]), rater type at three levels (radiologist [n = 5], nonradiologist physician [n = 1], nonphysician [n = 2]), and rater sex (female [n = 6], male [n = 2]). In post hoc tests, Tukey adjustment was used to control for chance differences due to multiple comparisons.

Results

Phase 1: Delphi Results

Figure 1 shows the results of each round of rating and revision in the Delphi process among nine reviewers. After two rounds of revision, analysis of round 3 responses revealed that 42 of the 43 domains and subdomains (98%) had an average rating of very good or excellent relevance (i.e., 4 or greater on the 5-point scale) with moderate variation in responses across raters (Table 1).

Table 1 shows reviewers' mean ratings of the relevance of the revised instrument for radiologist-parent conversations about error disclosure and radiation risk. For both of these conversations, 43 of 43 items were rated as very good or excellent. The one item that was rated below 4 for the bad news conversation (item D.1, Asks about/addresses life events, circumstances, other people that might affect how they receive or process the information being conveyed) was retained, because it was rated as relevant for the other two conversations. The resulting adaptation, representing the final tool, contained nine essential domains and 34 total subdomains, for a total of 43 items.

Phase 2: Interrater Reliability Results

To assess the reliability of KalRad, the ratings from the eight raters about the seven conversations (each conversation representing a different trainee) were used to compute intraclass correlation coefficients and weighted kappa statistics for each of the finalized 43 items.

Rater agreement measures—Reliability and rater agreement measures for most items were only moderate with intraclass correlation coefficient values ranging from 0.07 to 0.58 (mean, 0.30; SD, 0.13; median, 0.30). Weighted kappa values ranged from 0.03 to 0.47 (mean, 0.23; SD, 0.12; median, 0.22). As an example, Table 2 shows ratings for a single item (item H.1 in Table 1) and is representative of results for items with higher reliability. Raters were consistently able to identify a better conversation (conversation 2) and a poorer conversation (conversation 1), but ratings for the other five conversations exhibited considerable variation. For item H.1, the intraclass correlation coefficient was 0.54 and the weighted kappa value 0.45, making it one of the more reliably rated items.

Mean item ratings for combined conversations—For the seven conversations as a whole, the mean ratings for 41 of the 43 items were between 3 (good) and 4 (very good). The two exceptions were the item asking about life events and circumstances that may affect how the parent receives or processes information (item D.1; mean, 2.98; SD, 1.2), and the item explaining other participating clinicians' input (item F.2; mean, 4.03; SD, 0.77).

Comparison of ratings across conversations—Ratings varied significantly among conversations (χ^2_6 = 1186; p < 0.0001) (Fig. 2). The lowest mean ratings were found for conversation 1 (mean, 2.48; SD, 0.98) and the highest ratings for conversation 2 (mean, 4.59; SD, 0.65). Conversation 2 also had the lowest variation in ratings, indicating a high level of rater agreement about the conversation. The SDs of ratings of the other six conversations ranged from 0.94 to 1.12.

In regression models, the video conversation was the most significant predictor of rating. In pairwise post hoc comparisons of mean ratings, ratings for conversation 1 were significantly lower (i.e., worse) than the ratings for all conversations except conversation 6 (mean, 3.06; SD, 0.94). Ratings for conversation 2 were significantly higher than the ratings of all other conversations except conversation 4. Ratings also varied significantly by rater type; the one nonradiologist physician gave lower ratings (β = –0.30; p = 0.004) than the average of the four radiologists. Rater sex approached significance; women raters gave lower ratings than men did (β = –0.22; p = 0.058).

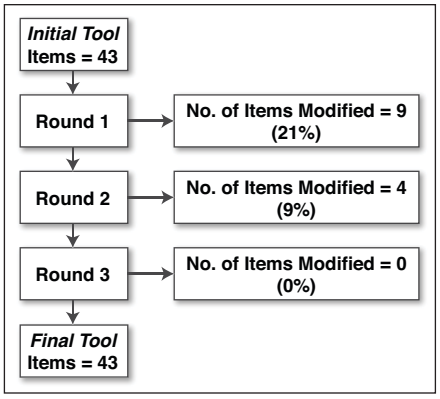

Fig. 1—Chart shows results with Delphi method in three iterative rounds. Revisions were made after rounds 1 and 2.

Comparison of ratings by viewing order—The effect of viewing order on mean rating was assessed by use of a viewing order–by–conversation interaction, which tested whether ratings for a conversation depended on viewing order. This interaction was highly significant (χ^2_6 = 765; p < 0.0001). Pairwise comparisons showed that conversation 1 was rated significantly lower when viewed first than when it was viewed last (mean rating viewed first, 2.12; viewed last, 3.06; p < 0.006). For the highest-rated conversation (conversation 2), ratings were not significantly different whether the conversation was viewed second or second-to-last. Conversation 7 was rated marginally lower when viewed first than when viewed last (mean rating viewed first, 3.22; SD, 1.04; viewed last, 4.02; SD, 1.06; p = 0.07).

Discussion

In this study, we used a simulation-based method to develop and validate a well-established standardized communication skills assessment tool (GKCSAF) for use in radiology. The revised instrument, KalRad, exhibited strong relevance for radiologist-to-patient communication and moderate interobserver reliability. Our group of reviewers found it straightforward to use with an approximately 10- to 15-minute average completion time per encounter. We anticipate similar times for those familiar with assessing communication competencies. KalRad is readily available (Supplement S1,

TABLE 1: Kalamazoo Communication Skills Assessment Tool—Radiology (KalRad): Final Validation Among Nine Reviewers for Bad News, Error Disclosure, and Radiation Risk

Skill	Bad News		Error Disclosure		Radiation Risk	
	Mean	SD	Mean	SD	Mean	SD
A. Overall ability to establish a rapport	4.88	0.35	4.71	0.49	4.75	0.46
1. Greets and shows interest in the patient and patient's family	4.78	0.44	4.67	0.5	4.67	0.50
2. Uses words that show care and concern throughout the interview	4.89	0.33	4.78	0.44	4.67	0.50
3. Uses tone, pace, eye contact, and posture that show care and concern	4.89	0.33	4.78	0.44	4.78	0.44
4. Responds explicitly to patient and family statements about ideas and feelings	4.89	0.33	4.78	0.44	4.78	0.44
B. Overall ability to open the discussion	4.67	0.52	4.67	0.52	4.83	0.41
1. Allows patient and family to state their understanding of why they are here	4.78	0.44	4.44	0.53	4.78	0.44
2. Asks "is there anything else?" or another open-ended question to allow space for patient to express concerns	4.56	0.53	4.67	0.500	4.44	0.73
3. Explains and/or negotiates an agenda or reason for the visit	4.00	1.00	4.22	0.97	4.11	1.05
C. Overall ability to seek/elicit further information	4.57	0.53	4.29	0.76	4.57	0.53
1. Addresses patient and family statements using open-ended questions	4.56	0.53	4.33	0.71	4.44	0.53
2. Clarifies details as necessary with more specific or "yes/no" questions	4.44	0.73	4.22	0.67	4.44	0.53
3. Summarizes and gives family opportunity to correct or add information	4.67	0.50	4.44	0.73	4.56	0.53
4. Transitions effectively to additional questions	4.33	0.71	4.25	0.71	4.56	0.73
D. Overall ability to understand the patient's and family's perspectives	4.40	0.89	4.50	0.55	4.29	0.76
1. Asks about/addresses life events, circumstances, other people that might affect how they receive or process the information being conveyed	3.67	1.22	4.44	0.53	4.11	0.78
2. Elicits/allows space for patient's and family's beliefs, concerns, expectations about their immediate situation	4.11	0.60	4.56	0.53	4.22	0.67
E. Overall ability to share information	4.83	0.41	4.75	0.46	4.75	0.46
1. Assesses patient's and family's understanding of problems and desire for more information	4.78	0.44	4.78	0.44	4.67	0.50
2. Explains using words that family can understand	4.78	0.44	4.78	0.44	4.78	0.44
3. Clearly conveys immediate next steps	4.56	0.53	4.67	0.5	4.78	0.44
4. Asks if family has any questions	4.71	0.49	4.67	0.5	4.78	0.44
F. Overall ability to reach agreement	4.29	0.76	4.88	0.35	4.71	0.49
1. Includes/assists family in any immediate decisions that must be made	4.33	0.71	4.67	0.50	4.78	0.44
2. Checks for mutual understanding of immediate further steps to be taken	4.44	0.73	4.89	0.33	4.78	0.44
3. Asks about acceptability of immediate further steps to be taken	4.22	0.67	4.89	0.33	4.56	0.53
4. Identifies additional resources as appropriate	4.22	0.44	4.89	0.33	4.33	0.50
G. Overall ability to communicate accurate information	4.57	0.53	4.71	0.49	4.71	0.49
1. Accurately conveys the seriousness of the patient's condition	4.67	0.50	4.67	0.50	4.56	0.53
2. Explains other participating clinicians' input	4.67	0.50	4.78	0.44	4.67	0.50
3. Clearly presents and explains possible immediate next steps	4.44	0.53	4.78	0.44	4.56	0.53
4. Gives enough information to assist with informed decision-making	4.44	0.53	4.67	0.5	4.56	0.53
H. Overall ability to demonstrate empathy	4.71	0.49	4.86	0.38	4.71	0.49
1. Clinician's demeanor is appropriate to the nature of the conversation	4.78	0.44	4.88	0.35	4.78	0.44
2. Shows compassion and concern	4.78	0.44	4.88	0.35	4.56	0.53
3. Identifies/labels/validates patient's and family's emotional responses	4.56	0.53	4.75	0.46	4.67	0.50
4. Responds appropriately to patient's and family's emotional cues	4.56	0.53	4.75	0.46	4.56	0.53
I. Overall ability to provide closure	4.67	0.52	4.83	0.41	4.67	0.52
1. Asks if the patient and family have questions, concerns, or other issues	4.56	0.53	4.75	0.46	4.67	0.50
2. Summarizes	4.44	0.53	4.88	0.35	4.67	0.50
3. Clarifies/reiterates the immediate next steps—where they are going and/or who they will see next	4.44	0.53	4.88	0.35	4.78	0.44
4. Provides appropriate contact information—for radiologist or clinical team—if interim questions arise	4.44	0.53	4.75	0.46	4.56	0.53
5. Acknowledges patient and family, and closes interview	4.67	0.50	4.75	0.46	4.78	0.44

Note—Values are ratings on a 5-point Likert scale (1, poor; 2, fair; 3, good; 4, very good; 5, excellent).

which can be viewed in the *AJR* electronic supplement to this article, available at www.ajronline.org).

The moderate interobserver reliability in our study is striking given that our raters met on numerous occasions to achieve consensus on what constituted effective and strong communication skills for radiologists. This variation likely reflected the diverse backgrounds of our reviewers, who may have held contrasting perspectives concerning particular competencies of specific communicators. Ratings may also have been influenced by how well reviewers knew those being assessed, sex differences, and differences in radiologists' and nonradiologists' personal or professional expectations of radiologists. Other studies conducted with multisource feedback in medical education, including within radiology, have shown significant differences in competency evaluations between assessor groups [7, 21, 22]. Nonetheless, our use of disparate representative stakeholders as reviewers and raters was consonant with beliefs that interdisciplinary collaboration is optimal for assessing competency in communication and interpersonal skills and that those with expertise in humanistic and psychosocial aspects of health care bring uniquely valuable insights into the evaluation process [11, 15, 22, 23]. Such an interdisciplinary approach may be particularly important for assessing communication and relational skills in radiology, in which no established standards for excellence exist, few validated communication skills programs have been developed, and faculty development remains largely ad hoc. The use of multisource feedback for assessing clinical competencies has been further endorsed for its recognition that health care delivery is becoming increasingly interdisciplinary and team based and that this trend requires assessors from a variety of perspectives [11, 15, 22, 24]. This is of substantial relevance to radiology, the value of which is becoming increasingly tied to its embeddedness within disparate health care system domains [25–28].

Previous studies of interrater reliability specifically for Kalamazoo instruments have had mixed results [12, 17]. Among cohorts of faculty, standardized patients, and students in one study, Joyce et al. [17] found high consistency within each group for ratings of the students' communication skills but lower correlation between groups, in particular students' self-assessments compared with their assessments by faculty and standardized patients. Using the instrument from which ours was specifically adapted, Peterson et al. [12] reported high interobserver reliability in ratings from faculty and peer observers while scoring simulated conversations between standardized patients and the participants being assessed. However, that study was designed to help create a communication skills curriculum rather than primarily to test interrater reliability. Before completion of the assessment tool, all raters participated in postsimulation debriefings with other observers, participants, and actor-patients, which, the authors noted, could have substantially influenced the raters' assessments. In our study, to maintain purity of perspective, no cross communication occurred among raters with other raters or participants before the completion of ratings.

One further potential limitation to our study was that the instrument was validated specifically for bad news conversations, which are only one sort of difficult conversation in radiology. However, our final phase of validity testing showed the relevance of the tool for assessing communication with patients about radiation risks and diagnostic errors. It would be impractical to develop different instruments for every imaginable scenario, although we would welcome further validation in other arenas of radiologic communication. Certainly many, if not most, of the relational skills accounted for within the instrument are relevant to other types of difficult conversations in radiology, although given how extensive the Kalamazoo instrument is, it may be difficult to provide deeply informed ratings for each domain, especially for short physician-patient and physician-parent interactions.

Overall, our experience points to some key lessons to be considered for leaders within radiology programs that are developing initiatives to assess trainees' communication skills. First, if multiple individuals are being evaluated serially, the order in which they are observed may influence observers' ratings. This is important to keep in mind even for programs in which order randomization may be impracticable. Most important, to be fair to both trainees and the process, our results suggest that assessment of radiology trainees' communication skills should be performed by interdisciplinary teams whose members represent the disparate perspectives of those who interact with radiologists

TABLE 2: Number of Ratings for Instrument Item H.1: Clinician's Demeanor Is Appropriate to the Nature of the Conversation

Video Conversation	Poor	Fair	Good	Very Good	Excellent	Total No. of Ratings
1	2	5	1	0	0	8
2	0	0	0	1	7	8
3	0	2	2	2	2	8
4	0	1	0	2	5	8
5	0	0	4	3	1	8
6	0	1	5	1	1	8
7	0	0	2	2	4	8

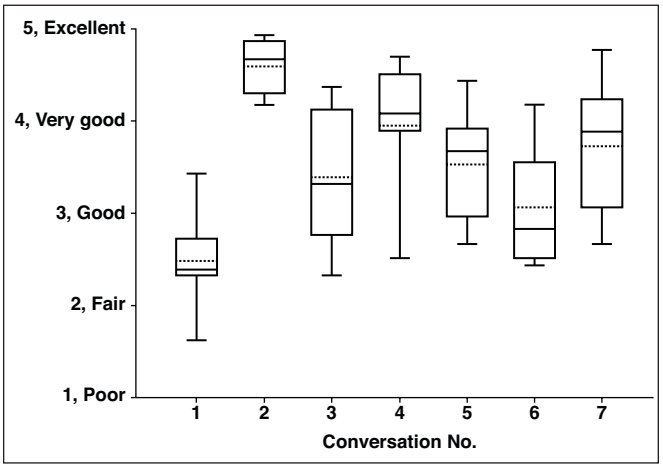

Fig. 2—Plot shows distribution of mean ratings for eight raters of enactments between actors and seven different trainees for same bad news conversation. Statistically significant variation was found in ratings among conversations. Reliability was highest for highest-rated conversation. Extremes of whiskers denote 5th and 95th percentiles.

as peers, mentors, and patients. Although our data suggest that most observers will mutually recognize the best examples of communication, to avoid various potential biases in assessment, such a team of observers might optimally include those who know the trainees well and those who do not. Training programs should be aware of the potential for bias in assessing communication skills that is related to both the rater characteristics and the assessment process itself.

Communication skills assessment instruments such as ours are not intended for use in grading trainees any more than we would provide such metrics for diagnostic interpretive skills. For a competency and field in which standards of excellence are not yet clearly established, the benefits of the assessment process may therefore be threefold: to document standardized core competency, to provide formative feedback to trainees, and to allow trainees to hear (and appreciate) the disparate subjective perspectives of various stakeholders. With its intrinsic subjectivity recognized, a standardized process for assessment of communication skills would ideally allow learners to better understand their strengths and weaknesses, identify the actions necessary to improve their clinical effectiveness, and track their progress over time.

Programs that want to provide standardized assessment of communication competencies may want to establish teams that, like ours, work together to gain mutual understanding of the instrument and expected competencies. The videos created during simulation exercises proved highly valuable for this purpose. The ability to assemble a constant mix of raters will likely be challenging for many programs. Nonetheless, many institutions may have reasonably accessible resources, including other clinical programs that have developed or are interested in developing communication skills training curricula, and various programs for chaplaincy, social work, ethics, family services, medical humanism, and interpreter services. We have tapped into many such resources at our institution to assemble our communication skills training and assessment program.

Our study results underscore the value of using simulation in assessing the communication skills of radiologists and, in doing so, build on a burgeoning experience within radiology. Early experiences assessing radiology trainees' competencies entailed direct observation of residents in actual clinical encounters with real patients [7] and in simulated encounters with patients specifically trained to enact difficult conversations, make assessments, and provide feedback [6]. More recently, DeBenedectis et al. [8] reported on the use of video-recorded simulated enactments between radiology residents and trained professional actors. Those investigators found potential for evaluating and teaching skills for a variety of difficult conversation scenarios. Similarly, our program uses video-recorded simulated enactments between trainees and trained professional improvisational actors, which allowed our raters to view and assess the videos remotely and at their convenience. It also allows the training program directors to assess performances independently and to review the videos with trainees, each of whom receives feedback both from the program directors and from the actors. As with the other programs that have been described, performance of the assessments provides an important opportunity for learning.

Although simulation is not the only means of assessing the array of relational and communication skills that radiologists must possess over the range of conversations they must hold with patients, the method has many powerful attributes for the development of robust communication training curricula. We and others have found the value of assessing those skills. With the development of this adapted Kalamazoo Communication Skills Assessment Tool, radiology now has a standardized method not only for assessing and providing feedback on communication and relational skills but also for testing any educational innovations that may be implemented.

References

1. Accreditation Council for Graduate Medical Education, American Board of Radiology. The diagnostic radiology milestone project. Accreditation Council for Graduate Medical Education website. www.acgme.org/Portals/0/PDFs/Milestones/DiagnosticRadiologyMilestones.pdf. 2012. Accessed September 15, 2016
2. Accreditation Council for Graduate Medical Education website. ACGME program requirements for graduate medical education in diagnostic radiology. www.acgme.org/Portals/0/PFAssets/ProgramRequirements/420_diagnostic_radiology_2016.pdf. July 1, 2016. Accessed September 15, 2016
3. Vydareny KH, Amis ES Jr, Becker GJ, et al. Diagnostic radiology milestones. *J Grad Med Educ* 2013; 5:74–78
4. Brown SD, Callahan MJ, Browning DM, et al. Radiology trainees' comfort with difficult conversations and attitudes about error disclosure: effect of a communication skills workshop. *J Am Coll Radiol* 2014; 11:781–787
5. Chetlen AL, Mendiratta-Lala M, Probyn L, et al. Conventional medical education and the history of simulation in radiology. *Acad Radiol* 2015; 22:1252–1267
6. Lown BA, Sasson JP, Hinrichs P. Patients as partners in radiology education: an innovative approach to teaching and assessing patient-centered communication. *Acad Radiol* 2008; 15:425–432
7. Wood J, Collins J, Burnside ES, et al. Patient, faculty, and self-assessment of radiology resident performance: a 360-degree method of measuring professionalism and interpersonal/communication skills. *Acad Radiol* 2004; 11:931–939
8. DeBenedectis CM, Gauget JM, Makris J, Brown SD, Rosen MP. Coming out of the dark: a curriculum for teaching and evaluating radiology residents' communications skills through simulation. *J Am Coll Radiol* 2017; 14:87–91
9. Duffy FD, Gordon GH, Whelan G, et al. Assessing competence in communication and interpersonal skills: the Kalamazoo II report. *Acad Med* 2004; 79:495–507
10. Schmitt JE, Scanlon MH, Servaes S, Levin D, Cook TS. Milestones on a shoestring: a cost-effective, semi-automated implementation of the new ACGME requirements for radiology. *Acad Radiol* 2015; 22:1287–1293
11. Calhoun AW, Rider EA, Meyer EC, Lamiani G, Truog RD. Assessment of communication skills and self-appraisal in the simulated environment: feasibility of multirater feedback with gap analysis. *Simul Healthc* 2009; 4:22–29
12. Peterson EB, Calhoun AW, Rider EA. The reliability of a modified Kalamazoo consensus statement checklist for assessing the communication skills of multidisciplinary clinicians in the simulated environment. *Patient Educ Couns* 2014; 96:411–418
13. Rider EA. Interpersonal and communication skills. In: Rider EA, Nawotniak RH. *A practical guide to teaching and assessing the ACGME core competencies*, 2nd ed. Marblehead, MA: HCPro, 2010:1–84
14. Makoul G. Essential elements of communication in medical encounters: the Kalamazoo consensus statement. *Acad Med* 2001; 76:390–393
15. Calhoun AW, Rider EA, Peterson E, Meyer EC. Multi-rater feedback with gap analysis: an innovative means to assess communication skill and self-insight. *Patient Educ Couns* 2010; 80:321–326
16. Joyce BL, Scher E, Steenbergh T, Voutt-Goos MJ. Development of an institutional resident curriculum in communication skills. *J Grad Med Educ* 2011; 3:524–528
17. Joyce BL, Steenbergh T, Scher E. Use of the Kalamazoo essential elements communication checklist (adapted) in an institutional interpersonal and communication skills curriculum. *J*

Grad Med Educ 2010; 2:165–169

18. Cheung JJ, Chen EW, Darani R, McCartney CJ, Dubrowski A, Awad IT. The creation of an objective assessment tool for ultrasound-guided regional anesthesia using the Delphi method. *Reg Anesth Pain Med* 2012; 37:329–333
19. Cristancho S, Moussa F, Dubrowski A. Simulation-augmented training program for off-pump coronary artery bypass surgery: developing and validating performance assessments. *Surgery* 2012; 151:785–795
20. Dimitrow MS, Mykkanen SI, Leikola SN, Kivela SL, Lyles A, Airaksinen MS. Content validation of a tool for assessing risks for drug-related problems to be used by practical nurses caring for home-dwelling clients aged ≥ 65 years: a Delphi survey. *Eur J Clin Pharmacol* 2014; 70:991–1002
21. Moonen–van Loon JM, Overeem K, Govaerts MJ, Verhoeven BH, van der Vleuten CP, Driessen EW. The reliability of multisource feedback in competency-based assessment programs: the effects of multiple occasions and assessor groups. *Acad Med* 2015; 90:1093–1099
22. Hayward MF, Curran V, Curtis B, Schulz H, Murphy S. Reliability of the interprofessional collaborator assessment rubric (ICAR) in multi source feedback (MSF) with post-graduate medical residents. *BMC Med Educ* 2014; 14:1049
23. Meyer EC, Sellers DE, Browning DM, McGuffie K, Solomon MZ, Truog RD. Difficult conversations: improving communication skills and relational abilities in health care. *Pediatr Crit Care Med* 2009; 10:352–359
24. Wood L, Hassell A, Whitehouse A, Bullock A, Wall D. A literature review of multi-source feedback systems within and without health services, leading to 10 tips for their successful design. *Med Teach* 2006; 28:e185–e191
25. Boland GW, Thrall JH, Duszak R Jr. Business intelligence, data mining, and future trends. *J Am Coll Radiol* 2015; 12:9–11
26. Bruno MA. Advanced practice quality improvement: beyond the radiology department. *J Am Coll Radiol* 2014; 11:1150–1154
27. Kang SK, Fagerlin A, Braithwaite RS. A roadmap for personalized care in radiology. *Radiology* 2015; 277:638–643
28. Kaura D, Blickman JG, Boland GW. The haptic radiologist: being there, always. *J Am Coll Radiol* 2016; 13:104–105

FOR YOUR INFORMATION

A data supplement for this article can be viewed in the online version of the article at: www.ajronline.org.

This article is available for CME and Self-Assessment (SA-CME) credit that satisfies Part II requirements for maintenance of certification (MOC). To access the examination for this article, follow the prompts associated with the online version of the article.