

Measuring Faculty Skill and Effectiveness in Conducting Interactive Learning Sessions—a Proposed Instrument

Patrick John Gannon^{1,2,3} · David Elkowitz^{3,4,5} · Alice Fornari^{3,6} · Gina Sorrentino³ · Joel N. H. Stern^{3,7,8} · Nancy Kheck² · Terri Shapiro⁹ · Naomi L. Lacy¹⁰ · David L. Steele¹⁰ · Latha Chandran¹¹

Published online: 9 May 2016

© International Association of Medical Science Educators 2016

Abstract Medical educators have created approaches such as the “flipped classroom” to cultivate robust knowledge transfer with high-order thinking skills and retention. Hybrid approaches, with independent self-directed learning and large group interactive “application” activities, may also better foster learning. No standard instrument exists to rate the effectiveness of interactive teaching in large groups. This study developed an objective criterion-based instrument to measure

the effectiveness of large group teaching by educators who engage with an audience that is pre-prepared by self-directed learning. We call such an educator a “Conductor of Interactive Learning” (COIL) and the interactive skillset, COIL skills. Based on needs assessment surveys of local and national end users, we designed a measurable standard-based instrument to evaluate effectiveness of a COIL. The instrument was designed to delineate measurable objective standards that characterize COIL skills and serve as a peer/self-assessment tool to identify individual COIL skills. After beta testing across institutional medical educators ($n = 28$), improved versions were presented to national experts ($n = 12$) and potential medical educator end users (three conference workshops) for constructive feedback on the quality and usefulness of the standards and the rating instrument. We developed standards that captured both interactive skills ($n = 7$) and non-interactive educational fundamentals ($n = 9$) for large group teaching. Using an iterative consensus building national process, we finalized a hybrid rating instrument consisting of these standards with anchors adjacent to merged 4-point Likert and visual analog scales. The standards and rating instrument for COIL skills could serve as an individualized guide for educator development.

✉ Patrick John Gannon
Beekanzi@gmail.com; pgannon@iau.edu.lc

- ¹ International American University, Vieux Fort, Saint Lucia, West Indies
- ² International American University College of Medicine, Vieux Fort, Saint Lucia, West Indies
- ³ Department of Science Education, Hofstra Northwell School of Medicine, Hempstead, NY 11549, USA
- ⁴ Academy of Medical Educators, Hofstra Northwell School of Medicine, Hempstead, NY 11549, USA
- ⁵ Department of Pathology and Laboratory Medicine, Hofstra Northwell School of Medicine, Hempstead, NY 11549, USA
- ⁶ Educational Skills Development Division, Hofstra Northwell School of Medicine, Hempstead, NY 11549, USA
- ⁷ Department of Neurology, Hofstra Northwell School of Medicine, Hempstead, NY 11549, USA
- ⁸ Department of Neurology, Lenox Hill Hospital, Northwell Health, New York, NY 10075, USA
- ⁹ Provosts Office, Hofstra University, Hempstead, NY 11549, USA
- ¹⁰ Paul L Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso, TX 79905, USA
- ¹¹ Stony Brook University School of Medicine, Stony Brook, NY 11794, USA

Keywords Self-directed learning · Interactive learning · Higher-order learning

Introduction

Medical educators generally agree that students should be trained in settings that promote self-directed learning (SDL) in

order to encourage higher-order thinking. As the best practices in teaching and learning have evolved, the literature reflects the fact that talented and seasoned teachers cultivate this process through a variety of pedagogical and assessment methods, even in the challenging arena of the lecture hall [1–6]. This poses fundamental questions of whether *all* educators can acquire the highly effective skillset of master teachers through targeted professional development [7], and if so, how that would be measured. In this paper, we describe the development of criteria for teaching faculty in their role as effective facilitators of student learning in interactive large group settings.

The studies by Newman and colleagues [8, 9] formulated a series of 11 criteria for lecturing skills that allowed “Overall lecture quality” to be assessed by peer review. Their criteria of lecturing performance such as “Encourages appropriate audience interaction” and “Monitors audience’s understanding of material” address some key issues of interactive teaching and learning. Srinivasan and colleagues introduced a framework of teaching skills that encompass a broad range of capabilities including content knowledge, learner centeredness, and practice-based reflection [5]. The work of Prober and Heath [1, 2] deals with innovative new approaches to medical education that are captured in their provocative article titles *Lecture Halls without Lectures* and *Medical Education Reimagined*. Rather than reinventing the large group lecture forum, these authors propose using methods in the classroom to foster retention through relevance, e.g., such as integrating patient discussion to render lectures “stickier (more comprehensible and memorable).” However, it was noted, this alone is still a lecture and works best with other strategies like the “flipped classroom,” (where students view video of the sticky lecture beforehand, allowing the lecturer to focus on interactive discussion and application of concepts in the large group setting) so that “Teachers would be able to actually teach, rather than merely make speeches” [1]. McLaughlin and colleagues have successfully implemented the flipped classroom approach in which students had access to a recorded library of content-focused lectures [3]. In addition, several activities were applied in their iteration of the flipped classroom (i.e., audience response and open-ended questions; multi-factorial pair and share) and micro-lectures to reinforce and consolidate learning. However, the preview of recorded lectures may not confer sufficient depth and breadth of foundational knowledge to prepare students for higher-order learning in large group encounters, and there is precious little time in the lecture setting to differentiate that. The teacher/facilitator has to gauge readiness, to optimize participation, and to redirect the exchanges of learners to foster meaningful application of biomedical science to patient care. Master teachers intuitively establish this alchemy and balance the needs of learners with instructional goals and objectives. It was our intention to create an instrument to guide instructors on the best practices for interactive teaching in a large group setting, based upon proven approaches in the literature and at our institutions.

To optimize interactive teaching/learning in different sized groups, our educators have engaged in a variety of hybrid approaches to large group teaching, which rely upon a foundational form of SDL framed by goals and learning objectives for each session. In this report, we describe the process of development of a skill rating instrument with specific standards that can be used to enhance educator effectiveness. While the criteria for the “Conductor of Interactive Learning” (COIL) tool were originally developed to promote mastery of interactive teaching skills, as it was beta-tested at multiple schools, we have determined broader applications for using the tool to *assess* faculty teaching skills. The COIL tool itemizes core teaching competencies in a visually structured way to form a platform for formative feedback: it describes key aspects of teaching style and learner engagement, and the criteria are anchored with descriptors across a continuum (from novice to master level). For faculty training and assessment purposes, this tool facilitates the discussion of these milestones after observed teaching. As such, we propose that the COIL tool is poised to become useful for *self and* peer assessment of interactive teaching and to cultivate a reflective practice to enhance interactive teaching skills.

Methods

At Hofstra Northwell School of Medicine, we use a hybrid SDL design for delivery of the medical science curriculum. In a hybrid SDL design, learning activities outside of class are followed by a large group interactive session conducted by a faculty member, who is both a content expert and skilled in the process of hybrid SDL-based interactive teaching with the Socratic Questioning pedagogy [10]. We refer to such a faculty member as a “Conductor of Interactive Learning” (COIL).

We defined a set of criteria necessary for COIL faculty in their role as effective facilitators of student learning. The goals and learning objectives for the session define the depth and breadth of the session content. A key component of higher-order interactive learning is to have COIL faculty trust learners to contribute to the learning partnership and feel comfortable with a collaborative classroom. If learners do not complete prerequisite learning activities (pre-work) that focus on foundational stages of knowledge and comprehension, the interactive session is more likely to devolve into a traditional, unidirectional delivery of a lecture by the content expert. Similarly, learners should trust the COIL to engage them in higher-order interactive learning and to avoid reiteration of their H-SDL foundational content in the classroom. To ensure success, it is vital to adhere to clearly defined COIL standards (skills) to guide educators in their role as a COIL educator. In the development phase, our standards were developed for two domains in the assessment of COIL skills: (1) specific interactive skills (seven items) and (2) general educational fundamentals (nine items). We refer to these criteria for interactive teaching collectively as the 16 “COIL standards.”

Development of the COIL Tool

The COIL tool consists of the 16 standards and a rating scale for each of the standards. Next, we describe the four phases for the process of creating and finalizing the COIL standards and the rating scale.

Phase I: Creating the First Draft of COIL Standards

Standards that characterized interactive educators were formulated by the following: (1) review of instruments that have been validated for use in the assessment of peer/self-rating of lecturing skills and educator competencies; (2) direct observation (by authors PJG, DEE) of in-house educators ($n = 16$), master educators selected as role models, who were in the process of refining their own interactive teaching skills as they conducted large group learning sessions; and (3) consultation with a national group of *medical education research experts* (MEREs, $n = 12$) who were recruited by the lead author (PJM) through their affiliations as alumnae of the Harvard Macy Institute. As a result, a first working draft of the COIL standards was generated. This was an initial step in construct validation and testing the reliability of the COIL instrument [11].

Phase II: Beta-Testing Standards and Finalization

We selected our COIL faculty from both senior and junior in-house educators who expressed enthusiasm and interest in learning new instructional skills for large group teaching. These early adopters assessed draft versions of our teaching criteria (COIL standards). They participated in the primary validation process using audience response clickers on anonymous setting to objectively rate each of the newly drafted standards. For example, they rated the statement “This Standard is Important for Both Self and Peer Rating of COIL- Skills” using a scale of (a) very important; (b) important; (c) usable with modification (= *major modifications required*); and (d) not usable (or not important enough to include). Results were subjected to multiple rounds of in-house overview by core COIL faculty. Outside expertise was solicited for subsequent round of reviews. MEREs were given the complete drafts and charged with independently determining which standards merited inclusion. In later stages, all critical comments and baseline psychometric data (audience response data from earlier review rounds) were shared to negotiate consensus about the final iteration of COIL standards. The approach was designed to be inclusive and recognize the diverse expertise of stakeholders who invested time/effort to support this project. After extensive consideration of all feedback from in-house educators and experts, we finalized the 16 COIL standards by consensus of the core COIL faculty and implemented the majority of recommendations (>90 %) after three rounds of external review (Table 1).

Phase III: Expanding the Definition of Standards for Clarity

As the standard design progressed, the core COIL faculty determined that more detail was needed, to limit variation in how the criteria were interpreted and applied, particularly as the standards were to be used for peer/self-assessment and individualized skill development. The developers then created elaborate descriptors to anchor each standard with specific examples (referred to as foundational components C1, C2, C3). We expanded on these descriptors with a series of statements that would capture the meaning of each standard with examples of specific behaviors. The expanded definitions of COIL standards were presented to the core-teaching faculty over the course of three workshops. The final versions of all 16 standards were placed on a website for easy access by collaborators http://medicine.hofstra.edu/pdf/department/scienceedu/scienceedu_coil_expanded_all.pdf.

Phase IV: Development of a COIL Skill Rating Scale

After developing the concise and elaborated version of standards that define the COIL skills, we focused on developing a rating scale to be used for each standard. Various rating scale formats were considered for an instinctive way to assess the behaviors defined as COIL skills. We deliberated upon eight individual scales and combinations, including 4-, 5-, and 7-point Likert scales [12] and the visual analog scale (VAS), commonly used for rating pain [13].

Variants of the VAS were considered, including gray and color continua, and as well as combinations of VAS/Likert scales. A dual approach was chosen to represent both quantitative (ordinal) and qualitative (interval) data. In-house medical educators ($n = 16$) assessed each rating scale option for the following perceived qualities: “ease of use, familiarity, usefulness, inter-rater reliability, internal consistency, and generalizability.” Prior to open-forum discussion, the group rated their preference using a scale from “Strongly Agree” to “Strongly Disagree.” Statements such as “Continuous Color Scales Would Work Best for Me” (personal best fit) were used for rating throughout.

For potential use as descriptors, two different word-series choices for the anchored rating scales were assessed by the 16 medical educators. We considered the 4-point approach “Strongly Disagree, Disagree, Agree, Strongly Agree” and then decided on the more representative “Novice, Advanced Beginner, Proficient, Master” descriptors based on the Dreyfus scale [14]. The statement used for rating “This Rating Scale is Both a Valuable and Easy to Use Instrument to Rate COIL skills” established this preference. Next, the MEREs independently evaluated all of their preferences for the COIL rating instrument, including numerical scale, VAS, and choice of descriptors. The core COIL faculty consolidated all feedback, written, and oral comments to finalize the COIL tool that included the 16 standards and the specific rating scale.

Table 1 The 16 COIL standards: concise version

COIL standard number	Concise version of standard
S1 ^a	Formulated and applied in context, question(s) that were directly linked to goal/s and gave rise to effective, higher order discussions
S2 ^a	Engaged multiple individuals to inspire interactive learning across the large group
S3 ^a	Observed and responded appropriately to the comprehension of material across the group
S4 ^a	Recapped the main components at the end of each subsection of the session with students engaged to contribute to discussion
S5 ^a	Integrated material to enable comprehension that could not have been derived from primary or secondary sources
S6 ^a	Summarized and explained key concepts in an interactive concluding wrap up
S7 ^a	As moderator of an interdisciplinary teaching faculty TEAM presentation, effectively managed COIL contributions and student involvement
S8	Assigned pre-readings that were relevant to the topic, at the appropriate size, depth and breath, and related to the Goal/s and Learning Objectives
S9	Presented and briefly elaborated the Goal/s
S10	Framed session to communicate importance and relevance of topic to the current theme
S11	Explained relevance of topic in context with previous and/or future exposures
S12	Demonstrated good content knowledge that was expressed with enthusiasm
S13	Articulated contextually appropriate interdisciplinary knowledge
S14	Used information technology creatively and appropriately to augment and reinforce Content
S15	Appropriate timing of key components across session allowed for a good understanding of depth and breadth
S16	Started and finished on time

^a Indicates specific interactive skills; others (S8–S16) are educational fundamentals

Results

Results of the Beta Testing of COIL Standards

Once the COIL tool was finalized, as described above, preliminary beta testing of the COIL tool was conducted by the in-house faculty ($n = 16$). They used the COIL standards to peer assess one another's teaching skills, as well as that of volunteer faculty and guest presenters (who knew of the mandate to be interactive, but did not have standards to guide them). Each session was followed by open-forum discussion to talk about their results in implementing the COIL tool. The results reported are percentiles from the midline of both sides of the 4-point rating series, being 0 to -100 % versus 0 to +100 %. As seen in Fig. 1, 15 of the 16 standards were rated "Very important" or "Important" by +75 to +94 % of observers. This gave us valuable insights for refining the language of COIL standards by consensus of the core COIL program faculty (authors), and that of the expert consultant group ($n = 12$).

Notably, only Standard 7 (S7) received the low negative rating of -69 % across the "Usable with modification" (UWM) and "Not Usable" (NU) categories. This standard differed from all others as it was formulated to distinguish the situation where a clinical core leader brought in multiple colleagues to present an interdisciplinary case. Standard 7 was modified to reflect the leader's role within an inter-professional team of teachers. After rewriting ("As moderator of an interdisciplinary group teaching presentation, effectively managed

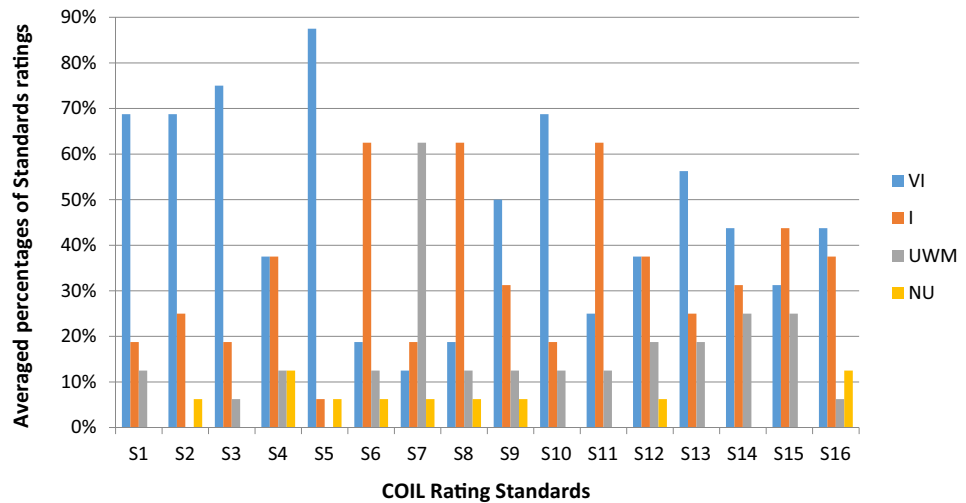
contributions and student involvement"), it was independently rated by all reviewers as "important." In this case, the other standards applied to all members of the inter-professional team. Similarly, other standards (with slightly negative ratings, 20 % range), S4, S12, S14, and S15, were reworked and optimized by the core COIL program faculty and sent out for review.

Results of COIL Standards Grouping and Elaboration of Definitions

During open-forum discussion and workshops with the core COIL faculty, it was noted that not all 16 of the first group of standards referred precisely to "interactive" COIL skills. By consensus, it was decided to sort the standards into two distinct domains: standards (S1–S7) apply directly to interactive skills (asterisked standards in Table 1), whereas standards (S8–S16) represent fundamental educational skills that can be applied to a variety of presentations, regardless of group size and pedagogical approach (Table 1, standards without an asterisk).

As previously described for phase III of development, we created elaborated definitions of the standards to support the use of this instrument for (peer/self-directed) professional development. For brevity, examples of the concise and elaborated COIL standards for teaching skills are shown in Table 2 while the full version including fundamental non-interactive standards may be accessed on the website: http://medicine.hofstra.edu/pdf/department/scienceedu/scienceedu_coil_expanded_all.pdf.

Fig. 1 Results of beta testing COIL standards. Raw data (table below) and rounded percentages in a bar chart of anonymous clicker rating (1–4) for COIL standards as rated by in-house faculty ($n = 16$ responders) to “Value for self and peer assessment of COIL-Skills”. *VI* very important, *I* important, *UWM* usable with modification, *NU* not usable



	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
VI	11	11	12	6	14	3	2	3	8	11	4	6	9	7	5	7
I	3	4	3	6	1	10	3	10	5	3	10	6	4	5	7	6
UWM	2	0	1	2	0	2	10	2	2	2	2	3	3	4	4	1
NU	0	1	0	2	1	1	1	1	1	0	0	1	0	0	0	2

Results of Beta Testing of the COIL Skill Rating Scale

To inspire educators to achieve the goal of becoming an adaptive expert in conducting interactive large group sessions, we provided them with the COIL rating scale, with the option of using this instrument for professional skill development via self and/or peer assessment of COIL skills (Fig. 2). We developed the tool to show both quantitative and qualitative rating of the COIL teaching standards being assessed. The example depicted in Fig. 2a allows raters to determine categorically which anchor statement and Likert box most closely corresponds to the standards being assessed during direct observation of teaching. It also allows a continuous rating of skills or fundamentals by use of a color spectrum-based visual analog scale across the top of the anchor statement and Likert box (Fig. 2b). When using the instrument, raters were instructed to first decide which anchored Likert statement best characterizes the observed skill and then to use the VAS to quantify the level of proficiency the teacher demonstrates in that skill. Thus, two types of data collection, with different resolution (see Fig. 2b), are supported by the COIL instrument (non-parametric from the Likert scale and parametric from the VAS).

Primary assessment of COIL rating scales by the in-house faculty ($n = 16$) gave rise to divided opinions about the 4-point Likert scale. The VAS was most highly rated (+54 % on the “Agree” side of the 4-point scale) while the “Continuous Gray Scale” was rated lowest (–91 % it on the “Disagree” side). Initially, broad resistance to 4-point [forced choice] Likert scale was overridden by the use of the VAS combination. This offered the ability to go meaningfully across a Likert box as a

continuum. The final rating scale option was a dual scale of a 4-point anchored Likert scale, adjacent to the continuous VAS. This approach was considered easy to use and allowed for co-collection of ordinal and interval data, with 100 % of the in-house faculty rating its efficacy as strongly agree. This version was adopted as the COIL rating scale of choice for implementation. Feedback from medical science educators after we presented COIL at national/international workshops suggested that using a visual scale in combination with the anchored Likert scale was an effective and complementary way to qualify the rater’s perception of observed teaching skills. This remains to be explored in subsequent validation research.

This last iteration of the COIL skill standards with the dual rating instrument was presented to the MEREs ($n = 12$) for final consideration with the same questions used for institutional faculty. They were accepted unanimously as being “acceptable and innovative.” The MEREs all preferred the “Novice to Master” option (Fig. 2), which was adopted. Comments from MEREs included “This is timely and should be pursued quickly; it is well justified since problems and needs are clearly there; Appropriate approach to evaluate faculty development in an innovative pedagogy.”

Discussion

Our initial goal for establishing COIL standards was to help educators adopt a new teaching approach, aligned with our school’s educational directives (to find an antidote to traditional

Table 2 Elaborated versions of COIL skills (interactive skills only)*Standard 1**Formulated and applied in context, question(s) that were directly linked to goal/s and gave rise to effective, higher order discussions*

Effectively used higher order questions or prompts that directly related to the Goal/s and Learning Objectives of the session (C1). Continued to use these effectively in order to generate discussions that were clearly continuous and related in sequence (C2). Avoided lapses into inappropriately placed, unidirectional lecture style (C3).

*Standard 2**Engaged multiple individuals to inspire interactive learning across the large group*

Encouraged and facilitated large group uniformity of participation (e.g., by moving from section to section, engaging left to right, front to back, using names, looking for and focusing on participation gaps (C1), modeling but not overusing highly active participants (C2), Regularly pausing for questions while looking around (C3).

*Standard 3**Observed and responded appropriately to the comprehension of material across the group*

In addition to promoting responses and engagement, was able to read the group as to level of comprehension of material. For example, encouraged questions to be asked and answered (C1); asked for best guesses by the students through facilitation of a specific discussion (C2); simplified a topic that wasn't working well (even the most active students were quiet) to provide a meaningful staged approach to arrive at the goal (C3).

*Standard 4**Recapped the main components at the end of each subsection of the session with students engaged to contribute to discussion*

Consistently used an approach that followed from framing the session upfront, where different components were summarized and placed into context (C1). Also at this time: a) interactive learning techniques were used to assess broad understanding before moving on (C2); and b) the relevance of knowledge gained in this component to those that follow were elaborated to make the sequential shifts transparent (C3).

*Standard 5**Integrated material to enable comprehension that could not have been derived from primary or secondary sources*

Was able to reliably and consistently promote creative scenarios (C1), where examples unfolded to arrive at novel elements of comprehension (C2) and students synthesized distinct but related topics to formulate the big picture (C3).

Primary sources include peer reviewed published research articles. Secondary sources include textbooks, websites and published reviews/meta-analyses, etc.

*Standard 6**Summarized and explained key concepts in an interactive concluding wrap up*

Placed all of the components of the session into a manageable package that illustrated their cohesion and relevance (C1). What started the session upfront with a framed overview of relevance (Fundamental Standard 10) was followed up here with a concluding synopsis that elaborated the progressive unfolding of the session's components (C2), along with their relatedness and multifactorial relevance (C3).

*Standard 7**As moderator of an interdisciplinary teaching faculty TEAM presentation, effectively managed COIL contributions and student involvement*

Similar to individual presenters, but instead, there is a group of educators (two or more) that are contributing, within their area of knowledge expertise, towards a common goal for the session. A moderator (or group leader) oversees the process by: a) introducing the members of the team and their content expertise, b) explaining the relevance of their contributions (C1), c) managing the timing, relevance, and interdisciplinary dynamics of the session (C2), and d) promoting faculty and student interactions (C3). Contributors, other than the moderator, may also be evaluated independently using standards that are applicable to their sub-session roles.

lectures). These criteria for interactive teaching were intended to guide faculty to the “how and what to do” to effectively engage learners in large group setting. The COIL tool development process originated by comparing previously published tools with the real-life, independently derived practices of our best teachers. From the systematic observation of our master teachers and in-house teaching faculty during interactive learning sessions (by PJG, DEE), two distinct domains emerged (interactive skills and educational fundamentals) to ground the organization of our COIL standards. That was followed by the challenge to *qualify* the degree of proficiency teachers demonstrated for each of these criteria, and it was this step that transformed our COIL standards from guidelines to model best practices in large group teaching into a versatile instrument for the assessment of faculty teaching.

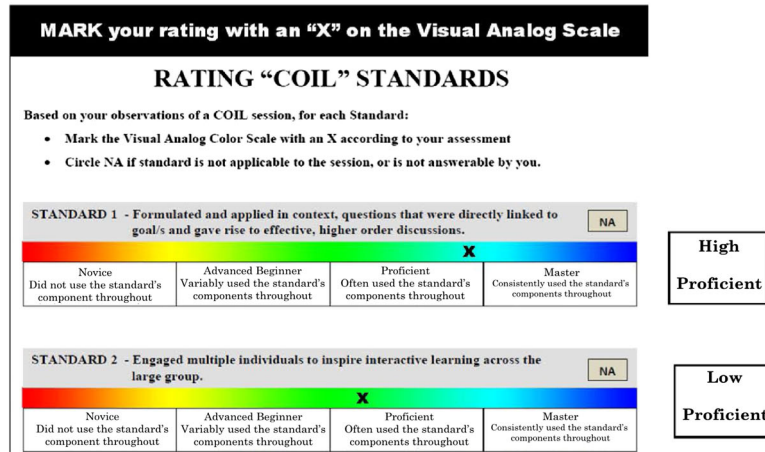
While developing the elaborated definitions for COIL standards (to ensure reliability in application), we tested the

practicality of using the COIL instrument for faculty development in interactive teaching and to provide structured formative feedback to teachers as they practiced these new strategies. The COIL standards can apply to learner groups of variable size and have the potential to complement interactive engagement of learners in multiple educational venues. Having combined standards for both interactive skills and educational fundamentals with unambiguous rating scales to measure competency, all in a single instrument, offered flexibility and ease of adaptation for the purposes of training *and* assessment of teaching skills.

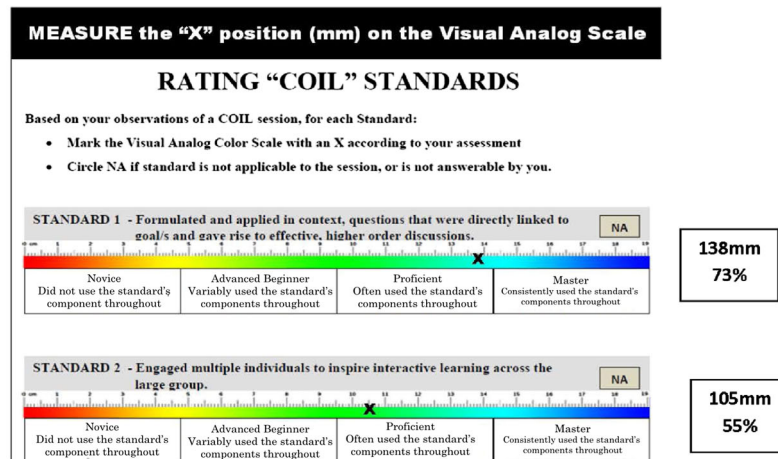
Medical educators look for ways to make their educational sessions pertinent and clinically meaningful to the budding physicians of tomorrow. The first component (C1) of the expanded definition of Standard 5 states “Was able to reliably and consistently promote creative scenarios where examples unfolded to arrive at novel elements of comprehension.” This

Fig. 2 The Hybrid Rating Instrument for COIL standards allowing alignment of continuous and categorical ratings. **a** The Rating Instrument is a hybrid of standards placed above a “Color Spectrum Visual Analog Scale” (VAS) and below that a 4-point Likert scale with anchor statements across standards that assess progressive competencies from Novice to Master. To rate the educational fundamental skills (S8–16), the VAS (only) is marked with a *cross* along its progression. (NA not able to rate). **b** The “X” mark position on the parametric scale is measured by mm ruler and derived as a whole scale percentage as well as its alignment with the 4-point non-parametric Likert scale being noted. Different parametric values can be associated with a single Likert box. The full series can be downloaded from the COIL website as a pdf. http://medicine.hofstra.edu/pdf/department/scienceedu/department/scienceedu_coil_rating_all.pdf

RATE the standard by considering Anchors and place a cross in the Visual Analog Scale (VAS).



MEASURE THE “x” from the left of VAS with a mm rule and transform VALUE to % of full VAS



prompts clinician educators to introduce working examples of key clinical experiences into the interactive learning session or integrate findings from case studies. Our observations of sessions conducted by practicing physicians who often utilize standard 5 support the perception that this fosters applied learning. Similarly, basic scientist educators could engage scientific foundations with clinical correlates and evidence-based medicine to achieve a similar goal.

The elaborated definitions of the standards have become an integral tool for professional development and added value to our program. Further, when students acclimated to the roll out of COIL and consistently completed SDL activities, our COIL faculty were capable and ready to move learners up to higher levels of critical thinking in the large group sessions. It was essential that COIL faculty were consistent in facilitating the applied and interactive learning in a compelling and intellectually challenging way. Large group attendance at most institutions is at record low; this interactive milieu brings learners to the classroom to gain insights they simply cannot accomplish via SDL.

When we developed the elaborated definitions of COIL standards, it served to (1) provide individualized guidance to educators along varying developmental stages, (2) standardize how our teachers interpreted the standards (and behaviors which demonstrate their application), and (3) support novice raters while observing interactive sessions. We anticipated that the COIL tool would provide a systematic path to help end-users at any level of proficiency to observe, assess, and practice exemplary teaching skills in large group settings.

The design of this instrument happened organically, through a process of iterative refinements that specifically addressed the needs of our teaching faculty. Utilitarian features like the dual rating scales (Likert and visual analog scales) evolved because rating the demonstration and proficiency of skills is inherently challenging. This requires the rater to make nuanced observations of behaviors and then to translate the decision of “how well” into checking off a Likert box (forced-choice set of 4; Fig. 2a). Data from a rating practice workshop (in Fig. 2b) shows how the visual analog scale

permits the rater to calibrate the relative degree of proficiency once they have selected the “Proficient” box on the Likert scale. During beta testing of the rating instrument, participants universally appreciated the freedom this feature offered and claimed to have frequently “struggled to justify a choice between adjacent anchors in a Likert scale because the best fit was somewhere in between.” Note that in Fig. 2b, the range of values (from 55 to 77 %) on the visual analog scale under the “Proficient” Likert anchor is broad and reflects a *low* through *high* degree of proficiency. This information would be potentially relevant and useful to monitor progression of individual skills (for ongoing professional development).

The main strength of this study is that it focuses on the pedagogy of large group interactive teaching and presents a “how to do it well” model with measurable standards for assessment. No standard instrument exists to rate the effectiveness of interactive teaching in large groups. Testing the COIL tool in multiple contexts and settings, we have learned that it positively supports teachers unfamiliar with this pedagogical approach and guides them through progressive stages of faculty development and practice to attain effective interactive teaching skills.

COIL’s dual-rating scale for teaching skills (with descriptive anchors and dual qualitative/quantitative scales) stimulated animated discussion at national and international workshops among beta-testers; the oft repeated subjective feedback was that COIL is “an innovative way to overcome the common *restrictions* of other assessment tools.” Most notably, teachers appreciated how the visual scale permitted a nuanced way to characterize the proficiency of teaching skills in contrast to the fixed Likert rating. A potential challenge of this approach is that it requires substantial professional development to foster consistency (inter-rater reliability); particularly important if the COIL tool is used as a peer assessment instrument to support continuous quality improvement.

Future Directions

The development of COIL standards and teaching skill rating scales marked an important transition for the professional development of our teaching faculty and how we prepared them for interactive teaching. With refinements, as this approach and the use of the COIL tool become systematic, it would enable longitudinal study of developmental milestones for effective teaching (interactive skills and educational fundamentals) in the large group setting. Combining the use of the expanded COIL standards within frame of reference training [9, 15] will also help newly recruited faculty to become proficient with COIL teaching skills.

It is critical that learners understand the importance of hybrid SDL so that at the start of any session, students attain at least level 2 of Bloom’s taxonomy “Remember – Understand” [16, 17]. This introduces the issue of whether

the COIL tool should be used by students to rate faculty. It has been debated extensively in multiple arenas, including faculty development sessions at participating medical schools ($n = 5$) and in workshops ($n = 3$) at national and international meetings. The consensus opinion was that giving students access to the COIL tool could be constructive. Moreover, it was noted that students often co-opt faculty resources to gain advantage and benefit their own learning. This may foster transparency and more insightful participation as a partner in the interactive classroom. We have concluded that the COIL tool has attained an appropriate level of content validation through our preliminary beta-testing efforts. It has also demonstrated potential applications for skill acquisition and self/peer assessment. In the next steps, we (1) aspire to optimize reliability across users, (2) evaluate how the COIL tool may be utilized to track the progressive development of teaching skills through reliable peer/self-rating, (3) plan to partner with evaluation experts to refine our validation studies, and (4) engage a larger pool of medical schools in testing the tool for potential use among learners and for faculty professional development.

Acknowledgments Thanks to the consultants at the Hofstra North Shore-LIJ School of Medicine for the support and advice in the development of the COIL program: David Battinelli MD; Samara Ginzburg MD; and Lawrence Smith MD.

The following medical education research experts acted as consultants in the formulation of the COIL program for which we thank them very much: Elizabeth G Armstrong PhD, Director of Harvard Macy Institute; Lisa Coplit MD, Associate Dean for Assessment & Faculty, Frank H. Netter MD School of Medicine at Quinnipiac University; Susan Farrell MD EdM, Faculty of Harvard Macy Institute, Professor Harvard Medical School; Robert Milner PhD, University of Massachusetts Medical School; Lori R Newman EdM, Harvard Medical School; Klara K Papp PhD, Director of Assessment and Evaluations, Office of Curricular Affairs, Case Western Reserve University School of Medicine; David Roberts MD, Harvard Medical School; Luanne E Thorndyke MD, Vice Provost of Faculty Affairs, University Massachusetts Med School. A particular thank you to Maryellen E Gusic MD who served as a major academic contributor to the COIL program since its conception at the 2012 Harvard Macy Institute.

References

1. Prober CG, Heath C. Lecture halls without lectures—a proposal for medical education. *N Engl J Med.* 2012;366:1657–9.
2. Prober CG, Khan S. Medical education reimaged: a call to action. *Acad Med.* 2013;88:1407–10.
3. McLaughlin JE, Roth MT, Glatt DM, Gharkholonarehe N, Davidson CA, Griffin LM, Esserman DA, Mumper RJ. The flipped classroom: a course redesign to foster learning and engagement in a health professions school. *Acad Med.* 2014;89:236–43.
4. Chandran L, Gusic M, Baldwin C, Tumer T, Zenni E, Lane JL, Balmer D, Bar-On M, Rauch DA, Indyk D, Gruppen LD. Evaluating the performance of medical educators: a novel analysis tool to demonstrate the quality and impact of educational activities. *Acad Med.* 2009;84:58–66.

5. Srinivasan M, Li ST, Meyers FJ, Pratt DD, Collins JB, Braddock C, Skeff KM, West DC, Henderson M, Hales RE, Hilty DM. "Teaching as a competency": competencies for medical educators. *Acad Med.* 2011;86:1211–20.
6. Milner RJ, Gusic ME, Thorndyke LE. Perspective: toward a competency framework for faculty. *Acad Med.* 2011;86:1204–10.
7. Berman AC. Good teaching is good teaching: a narrative review for effective medical educators. *Anat Sci Educ.* 2015;8:386–94.
8. Newman LR, Lown BA, Jones RN, Johansson A, Schwartzstein RM. Developing a peer assessment of lecturing instrument: lessons learned. *Acad Med.* 2009;84:1104–10.
9. Newman LR, Brodsky DD, Roberts DH, Pelletier SR, Johansson A, Vollmer Jr CM, Atkins KM, Schwartzstein RM. Developing expert-derived rating standards for the peer assessment of lectures. *Acad Med.* 2012;87:356–6.
10. Kost A, Chen F. Socrates was not a pimp: changing the paradigm of questioning in medical education. *Acad Med.* 2015;90:20–4.
11. Artino Jr AR, Durning SJ, Creel AH. AM last page: reliability and validity in educational measurement. *Acad Med.* 2010;85:1545.
12. Lee H, Paek I. In search of the optimal number of response categories in a rating scale. *J Psychoeduc Assess.* 2014;32:663–73.
13. Englbrecht M1, Tarner IH, van der Heijde DM, Manger B, Bombardier C, Müller-Ladner U. Measuring pain and efficacy of pain treatment in inflammatory arthritis: a systematic literature review. *J Rheumatol Suppl.* 2012; 90:3–10.
14. Dreyfus SE, Dreyfus HL. A five stage model of the mental activities involved in direct skill acquisition. Washington, DC: Storming Media; 1990.
15. Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med.* 2014;89:721–7.
16. Bloom BS [ed.], Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of educational objectives: the classification of educational goals. Handbook 1: cognitive Domain. New York, David McKay. 1956.
17. Anderson, L. W. and Krathwohl, D. R., et al (Eds.). A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. Allyn & Bacon. Boston, MA (Pearson Education Group). 2001.