IRis

Northeastern University

Bouvé Faculty Publications

Bouvé College of Health Sciences

December 15, 2008

Development of a peer teaching-assessment program and a peer observation and evaluation tool

Jennifer M. Trujillo
Northeastern University

Margarita V. DiVall
Northeastern University

Judith T. Barr
Northeastern University

Michael J. Gonyeau
Northeastern University

Jenny A. Van Amburgh
Northeastern University

See next page for additional authors

Recommended Citation

Trujillo, Jennifer M.; DiVall, Margarita V.; Barr, Judith T.; Gonyeau, Michael J.; Van Amburgh, Jenny A.; Matthews, S. James; and Qualters, Donna M., "Development of a peer teaching-assessment program and a peer observation and evaluation tool" (2008). *Bouvé Faculty Publications.* Paper 33. http://hdl.handle.net/2047/d20002373

This work is available open access, hosted by Northeastern University.

Author(s)
Jennifer M. Trujillo, Margarita V. DiVall, Judith T. Barr, Michael J. Gonyeau, Jenny A. Van Amburgh, S. James Matthews, and Donna M. Qualters

This article is available at IRis: http://iris.lib.neu.edu/bouve_fac_pubs/33

American Journal of Pharmaceutical Education **2009; 72 (6) Article 147.**

RESEARCH ARTICLES

Development of a Peer Teaching-Assessment Program and a Peer Observation and Evaluation Tool

Jennifer M. Trujillo, PharmD,[a] Margarita V. DiVall, PharmD,[a] Judith Barr, ScD, MEd,[a] Michael Gonyeau, PharmD,[a] Jenny A. Van Amburgh, PharmD,[a] S. James Matthews, PharmD,[a] and Donna Qualters, PhD[b]

[a]Northeastern University School of Pharmacy
[b]Northeastern University, Center for Effective University Teaching

Submitted December 18, 2007; accepted March 9, 2008; published December 15, 2008.

Objectives. To develop a formalized, comprehensive, peer-driven teaching assessment program and a valid and reliable assessment tool.
Methods. A volunteer taskforce was formed and a peer-assessment program was developed using a multistep, sequential approach and the Peer Observation and Evaluation Tool (POET). A pilot study was conducted to evaluate the efficiency and practicality of the process and to establish interrater reliability of the tool. Intra-class correlation coefficients (ICC) were calculated.
Results. ICCs for 8 separate lectures evaluated by 2-3 observers ranged from 0.66 to 0.97, indicating good interrater reliability of the tool.
Conclusion. Our peer assessment program for large classroom teaching, which includes a valid and reliable evaluation tool, is comprehensive, feasible, and can be adopted by other schools of pharmacy.
Keywords: peer assessment, large classroom teaching, peer observation and evaluation tool

INTRODUCTION

A majority of American colleges and schools of pharmacy use standardized student rating of large classroom teaching to assess the effectiveness of teaching.[1] These assessment tools are used to evaluate teaching effectiveness for such purposes as merit, promotion, and tenure decisions. Although a large volume of literature endorses the reliability and validity of standardized student ratings of teaching effectiveness, concerns have been raised that these rating systems do not capture the total teaching experience, nor should teaching effectiveness be judged solely from the student perspective.[2-9] Additional assessment questions that may not be answered by using only student evaluations include: Do learning objectives align with the overall course goals? Is the lecture material relevant to the course objectives? Is the content current and up-to-date? Are the assessment strategies consistent with learning objectives? Are connections to prior learning within the curriculum made effectively? How has student learning been documented? How does this course fit into the overall curriculum goals? While student evaluations are the traditional method of measuring teaching effectiveness, pharmacy faculty members have indicated an interest in utilizing alternative methods of evaluating teaching.[10]

The 2007 American Council of Pharmaceutical Education (ACPE) Guidelines for the Accreditation of Doctor of Pharmacy Programs state that teaching faculty members should be evaluated annually. The Guidelines further indicate that assessment procedures should include self-assessment and "…appropriate input from peers, supervisors, and students. The use of self-assessment and improvement tools, such as portfolios, by faculty and staff is encouraged."[11]

As part of the university-wide teaching review process, Northeastern University requires that at least 2 methods of teaching assessment be used. Standard student rating evaluations must be included for yearly merit as well as tenure and promotion decisions. However, what constitutes "best practices" for assessment of large classroom teaching is highly contextual given the differences in institutional goals, resources, priorities, and discipline paradigms. Therefore, other recommended means of documenting teaching effectiveness include peer classroom visits, peer evaluations of class materials, teaching portfolios (including self-assessment), and evaluations by alumni of the program.[12]

While peer assessment has been recommended, it has not been required at our institution. Within the Department

Corresponding Author: Margarita V. DiVall, PharmD, BCPS, Northeastern University School of Pharmacy, 360 Huntington Ave, 206 Mugar, Boston, MA 02115. Tel: 617-373-5370. Fax: 617-373-7655. Email: m.divall@neu.edu

of Pharmacy Practice, faculty members have been encouraged to have a colleague of their choice evaluate at least 1 lecture per year. The evaluation typically consisted of a colleague observing the lecture and evaluating the instructor using a checklist-style evaluation tool. Many factors limited the utility of this process. Because the process was not mandatory and individual faculty members were responsible for initiating the peer evaluation, not all faculty members consistently participated. Also, faculty members were not trained in peer observation and the tool's reliability and validity had never been tested. Thus, the Department did not view the peer assessment process as entirely useful and expressed interest in revising the program.

The objectives of this study were to develop, implement, and evaluate a peer assessment program for large classroom teaching (more than 75 students). The presentation of this effort is in 2 parts. First, the multi-step development process and the rationale used to proceed through each step are described. During this process, we determined that it was necessary to *de novo* develop a Peer Observation Evaluation Tool (POET) to standardize the observation. Therefore, in the second part of the report, the methods and results of the process to develop a valid and reliable assessment tool. Overall, the peer assessment program should allow faculty members to document teaching effectiveness over time; assist in the assessment of teaching at merit, promotion, and tenure reviews; and contribute to the improvement of teaching and learning within our program.

METHODS

The Northeastern University School of Pharmacy (NUSOP) consists of 2 departments: the Department of Pharmaceutical Sciences and the Department of Pharmacy Practice that includes both practice and social and administrative sciences faculty members. The 210 semester-hour doctor of pharmacy program is organized in a 0-6 year format. Thirty-one percent of the credit hours are taught by the College of Arts and Sciences, 22% by the Department of Pharmaceutical Sciences, and 47% by the Department of Pharmacy Practice. The latter credits are further divided into 36 semester hours of advanced pharmacy practice experience and 62 semester hours of didactic courses. Seminars and laboratories are included in the department's offering; however, the courses consist primarily of single, multiple, and team-taught courses using a large lecture format.

Development of the peer assessment program was initiated jointly by the Dean's Office and the Chair of the Department of Pharmacy Practice. The Assistant Dean for Academic Affairs assembled a departmental taskforce in the summer of 2005. All faculty members of the Department were invited to participate as a member of the taskforce and/or as a trained peer observer.

The taskforce was charged to develop and implement a formalized, comprehensive, peer-driven, teaching assessment program that would promote the improvement of large classroom teaching and learning. The taskforce collaborated with the University's Center for Effective University Teaching (CEUT). The CEUT director served as the educational expert advisor to the group.

The taskforce met with the CEUT director to determine key components and qualities of a peer assessment program and previewed Webb and McEnerney's stepwise approach to constructing a successful peer observation system, summarized in Table 1.[13] Our sequential, developmental approach outlined below was based primarily on this 10-step process.

Step 1: Establish a clear vision. The taskforce collaboratively developed and achieved consensus on our statement of purpose for the peer assessment program: "to promote the improvement of teaching and learning." Departmental teaching philosophy statement was also revised by the faculty and adopted by the department in May 2006. The elements of this philosophy served as the blueprint for the peer assessment program.

Step 2: Differentiate between formative and summative review. The Taskforce met as a group to make a fundamental decision regarding whether the system should be formative, summative, or a combination of the 2. The group agreed that a *formative* system that provides feedback for professional growth and development was the most appropriate option.

Step 3: Identify program leader and coordinator. Strong leadership and consistent long-term coordination were deemed critical components of a successful program. The Taskforce elected 2 chairpersons to provide leadership during the developmental and implementation phase of this process. The co-chairpersons paid particular attention to achieving consensus within the taskforce, attaining department-wide faculty buy-in, and incorporating peer assessment into the departmental culture. A program coordinator would likely be necessary to oversee the implementation and monitoring of the peer assessment program once approved by the Department.

Step 4: Identify participants and peer observers. The Peer Assessment Taskforce considered several issues including which faculty members should be assessed (eg, tenured, non-tenured, junior, and faculty members who had received poor student evaluations) and whether participation should be mandatory or voluntary. The decision was made that participation would be *mandatory* for *all* faculty members in the Department.

Peer observer training is essential to increase the acceptance of the process and to have participants feel comfortable that the observer would understand what to view and how to view it. The taskforce felt that, with appropriate training, all faculty members interested in being peer observers would be qualified, especially with a well-established process and a valid and reliable peer-evaluation instrument. The decision was made that participation as a peer observer would be voluntary.

The group agreed that only those who completed formal training could serve as peer observers and that all faculty members should be encouraged to receive this training. The peer assessment program coordinator will match each faculty member with a trained peer observer based on scheduling and availability. Faculty members could decline one peer evaluator pairing.

Step 5: Establish a process. A pre-observation meeting was used to allow the observer to review lecture materials, learn about the class climate, assess the single lecture within the context of the entire course, and target the observation to particular requests of the instructor (eg, provide feedback on active learning strategies).

A *teaching observation record* was used as a recording tool during the actual classroom observation. The teaching observation record consisted of 2 columns, a "plus" column to document positive aspects and a "delta" column to document areas of potential improvement, areas of best practices not observed, or questions the observer had about the classroom interactions. Observers were to record what they observed during the lecture in these columns and to review their observations after the lecture to identify the 2 or 3 most significant positive and negative observations or themes.[14] The teaching observation record was used as a resource when completing an assessment tool after the observation.

A post-assessment (eg, examination) meeting was held between the observer and instructor to allow the observer to review examination questions and item analysis to determine and discuss student mastery of course material covered during the observation.

Through literature review and expert advice, we identified 2 key concepts to weave into our process: self-reflection and personal interaction. *Reflection-on-action* prompts individuals to examine why they selected a particular approach over other options, what happened during the action (in this case, presentation of the lecture), what went well, what needs improvement or further development, and what new strategies or other options could lead to improvement.[15] The group chose to incorporate a self-reflective component into all 3 instructor/observer meetings, with the most intensive occurring at the post-observation meeting. While time constraints are always an issue, the importance of meeting face to face in the postobservations became a critical feature of the system.

The group recognized that the timeline and faculty workload needed to be reasonable and feasible to ensure that faculty members adhered to the process. The major components and suggested timeline of the process are outlined in Appendix 1.

Step 6: Identify a peer evaluation instrument. Several evaluation instruments were identified through literature and web searches; however, none were validated or reliable. The decision was made to develop a peer assessment instrument that was both valid and reliable for use in a large classroom setting as part of the comprehensive assessment program.

Step 7: Initiate training. Training is critical to the validity and value of the peer assessment process.[16] All taskforce members were formally trained. The 4-hour training session, conducted by the Department's education consultant, focused on effective peer observation techniques and provided faculty members with a model for observation and effective feedback. Faculty members were trained to effectively use the teaching observation record during the lecture and then use the record as a resource later when completing the assessment tool after the observation. After observing a portion of a prerecorded lecture, attendees practiced peer-review techniques using role play. Training for all members of the Department was conducted in fall 2007. The Taskforce obtained an intra-institutional instructional development grant to support these training sessions.

Step 8: Decide on logistics. As with any new program, the most difficult part is the implementation details. During this step, the Peer Assessment Taskforce identified and discussed many unanswered questions regarding the logistics of the program and achieved consensus on several policy statements.

How many observers should attend and evaluate each lecture? The group agreed that once the tool was proven valid and reliable one observer would be both sufficient and sustainable. Provision was made for instructors to request multiple observers if they would like the peer assessment to be summative to support promotion or tenure.

Should classroom visits be announced or unannounced? Announced visits offer advantages of allowing for a pre-observation discussion and for the instructor to be more comfortable with the observation, and for the instructor to be seen at his or her best. Unannounced visits can create a climate of distrust and fear, but can likely assess the level of instructor preparedness for any given lecture. Because of the value placed on the pre-observation meeting and the formative nature of our

Table 1. Steps to Creating a Peer Assessment Program

Step	Factors to Consider
1. Establish a clear vision	Create a statement of purpose that is collaboratively developed and consistent with the organization's mission.
2. Differentiate between formative and summative review	Formative reviews provide feedback for professional growth and development while summative reviews are used for personnel decisions/judgments.
3. Identify program leader and coordinator	Determine desirable characteristics of a leader, level of administrative support, available resources, individual versus team leadership.
4. Identify participants and peer observers	Determine whether participation is mandatory or voluntary and which instructors will be assessed. Identify a pool of peer observers (eg, teaching colleagues, external reviewers, education staff) and address issues of qualifications, credibility, experience, seniority, content expertise, and training requirements.
5. Establish a process	Determine how and when the observer should interact with the instructor with consideration to the feasibility and reasonability of the timeline and workload and the role of self-reflection.
6. Identify a peer evaluation instrument	Determine the desired format of the tool (eg, checklists, rating scales, written comments, combination) and seek input from faculty, administration, and "expert" colleagues. Align the instrument with your organization's mission and goals, the purpose of your PA program, and what is supported in the literature.
7. Initiate training	Assess available resources (within your institution or external) to identify experts to offer formal training and determine time commitment and frequency for training sessions.
8. Decide on logistics	In the context of your institution's resources and restraints, consider logistical questions such as who will participate, which lectures will be targeted, and when or how often will assessments occur.
9. Establish a time commitment and identify incentives and/or consequences	Determine a realistic estimate of time commitment and communicate this to all faculty. Identify incentives for those who participate and determine consequences for those who do not.
10. Establish communication modes and record keeping/reports	Determine necessary forms of communication and record keeping and discuss the importance of conveying the "rules of the game" to all faculty.

process, the group agreed that all observations will be announced and scheduled.

Does the observer need to stay for the entire lecture period or is part of the class sufficient? Observation of the entire lecture period is optimal; however, it can be more time consuming, and attending the first half or the second half of the longer lecture is a possibility. It is important to observe instructor student interaction either before or after class to assess classroom climate, instructor approachability, and instructor/student rapport.

Other selected questions that were debated by the Taskforce were: *How many times per year or semester should peer observations occur? Which lecture should be observed? Should faculty be allowed to choose which lecture is peer observed? Should lectures with poor student evaluations be selected for peer assessment? Should the same observer assess the same instructor over time?* The group agreed that each instructor should participate at least once per year. Each instructor would identify the lecture to be peer assessed and would be encouraged to select new lectures, lectures with perceived problems, or poor student evaluations for peer assessment.

Step 9: Establish a time commitment and identify incentives and/or consequences. The impact of peer assessment time commitment on already heavy faculty workloads is always a concern. The process needs to be comprehensive but feasible; faculty member buy-in is more likely if they have a realistic estimate of the time commitment. Such an estimate should be clearly communicated to all faculty members and estimates should include training, preparation for review, review, postreview feedback, paperwork, and meetings. Based on review of the literature and experience from field tests, the estimated total time commitment for 1 peer observation (including actual observation, three observer/instructor meetings, review of materials, and completion of final letter) was 4-8 hours.

Since the role of the peer observer requires the largest time commitment, incentives or rewards for participating

in this role should be carefully considered. The Department as a whole was strongly committed to improving teaching and learning. In addition, a brief, web-based, anonymous survey of the departmental faculty indicated that the majority agreed that peer assessment was a necessary step in teaching assessment and was willing to participate in the peer-assessment process. Most were also willing to attend a training session and serve as peer observers, which would limit the commitment to 1-2 peer observations per year. With that in mind, the group agreed that incentives and consequences could be limited. Annual participation will be mandatory and documented during the yearly merit evaluation. The process is formative, so peer assessment results are not included in merit reports, only the documentation that a faculty member participated in the process. Faculty who volunteer to serve as peer observers will receive credit toward service requirements for the year.

Step 10: Establish communication modes and record-keeping/reports. A written document communicating the "rules of the game" (process, polices, and expectations) was constructed and distributed to the department. Program coordinator will keep track of peer observation matches and faculty participation.

A written letter summarizing the outcomes of peer assessment is provided to the instructor by the observer. This letter is for formative purposes only to improve teaching and learning and is given only to the instructor. This instructor can, but is not required, to include this letter in merit, promotion, and tenure documents.

After completing the multi-step developmental process, the Taskforce proposed a comprehensive teaching assessment policy to the Department faculty which combined peer assessment with self-reflection and student evaluations. The proposal was unanimously approved and implemented in January 2008.

Development of the POET

A PubMed literature search (from January 1966 through September 2005) was conducted using the MeSH keywords *peer assessment; peer evaluation; peer observation; improving teaching and learning; evaluation tool or instrument; large classroom teaching*. Using the same keywords, we also searched Education Resources Information Center database and the archives of the *Journal*. A Web search was also conducted to identify existing peer assessment/evaluation systems in higher education. Several books (selected chapters) were reviewed. Identified articles and evaluation instruments were reviewed and all items were compiled into a master list.[1-10,16-27]

Prior to creating our own instrument, the Taskforce members deconstructed the elements of the Department's educational philosophy; these ultimately served as the blueprint for our evaluation instrument. The prototype for our peer-evaluation tool was constructed by linking each unique component of our educational philosophy with items from the master list that would best evaluate that component.

The first draft of our tool contained 123 items. Initially, each member of the Peer Assessment Taskforce and the education expert provided feedback on the organization, terminology, and clarity of the tool. Once similar items were combined and clarifications were made, each Taskforce member independently ranked each item on a scale of 1 (no importance) to 5 (high importance). All ranks were combined and averaged; items with a score of less than 3 were removed. Remaining items were modified or clarified based on feedback and reorganized to align with the peer-assessment process. Next, all Peer Assessment Taskforce Members piloted the tool on a prerecorded lecture. Based on consensus, some items were further streamlined and many item descriptors were added for clarification. The resultant document was called the Peer Observation Evaluation Tool or POET (a copy of the instrument is available by request from the author).

The POET consists of 39 ranked items, organized into 4 sections: pre-observation visit (8 items and pre-scripted interview questions); classroom observation – content (5 items), teaching strategies and presentation skills (17 items), and classroom climate (5 items); post-observation meeting (no ranked items, contains a rubric for reflection); and post-assessment meeting (4 items). Ranked items were evaluated using the following scale: accomplished well (AW), accomplished (A), needs development (ND), and needs significant development (NSD). To force a commitment to a specific response, options such as undecided, neutral, or not applicable were not used. Out of the 27 classroom observation items, the Taskforce identified 4 that might not routinely be observed; therefore, these items justified the addition of a fifth response item, did not observe (DNO). A series of open-ended dialogue questions was included in the pre-observation and postobservation meetings to guide the peer and self-assessment process.

A pilot study was conducted to establish interrater reliability of the POET. Eight taskforce members used the POET in a 2-semester pilot test to determine instrument reliability and to evaluate efficiency and practicality of the peer-assessment process. All 8 members served as the observer pool; 3 of these were also faculty members being observed. All pilot study faculty members had been trained to conduct peer assessment. Since observers and faculty members being observed in the pilot study

involved only volunteers from the taskforce, institutional review board approval was not necessary.

To determine interrater reliability, 8 separate lectures were evaluated by 2-3 peer observers. A different panel of observers was constituted for each lecture. All lectures were given to classes of 85-130 students.

For each lecture, the team of observers followed the previously described peer assessment process. For practical reasons, the team designated 1 peer observer to hold the pre-observation, post-observation, and post-assessment meetings. However, all peer observers were present during the classroom observation and independently completed the POET sections pertaining to classroom observation. Therefore, reliability testing was performed only on the POET's 27-item, classroom observation component. All members of the observation team reviewed the final written feedback provided to the instructor.

Interrater reliability of POET's classroom observation section was evaluated using intraclass correlation coefficient (ICC), SPSS, version 15.0 (Chicago, IL). Two-way random model analysis was used because both raters and instructors were selected from a pool and therefore were deemed "random." The consistency type ICC analysis was used to determine average measure reliability.[28,29] Intraclass correlation coefficients were computed for the overall classroom observation component of each lecture as well as for the 3 subsections: (1) content; (2) teaching strategies and presentation skills; and (3) classroom climate. For lectures with 3 observers, ICCs were computed for all raters (1, 2, 3), as well as for individual pairs of raters (1 and 2, 1 and 3, 2 and 3). Statistical significance of the proportion of variance that is systematic was set at $p < 0.01$ due to the large number of comparisons.

To further establish content validity, 3 nationally recognized leaders in pharmacy education and assessment were solicited to serve as external reviewers. These 3 experts were asked to rank each POET item on a scale of 1 (no importance) to 5 (high importance), and provide general comments about the POET and peer assessment process. Feedback was reviewed and incorporated into the POET based on consensus of peer assessment taskforce members.

RESULTS

There were 23 distinct classroom observations: 7 lecturers had 3 observers and 1 lecturer had 2 observers. Each observer completed the 27-item classroom component of the POET. The data set included 216 item-comparison clusters. In the majority of comparisons, observers agreed within 1 response level of each other ("low variation"). In only 7 instances (3%) did the observer team differ by more than 1 response level ("high variation"), all of which occurred in 3 lectures. Five lectures had no items with high variation.

When examining all DNO responses (n = 22), many items were in common among observers. Thirteen of the DNO responses (59%) were used for the item: "The instructor reacts to student professional behavior issues appropriately." In 4 lectures, all observers selected DNO for this item and in 1 lecture, 2 of 3 observers selected DNO for this item. In 2 lectures, 5 of 6 observers selected DNO for another item: "Divergent opinions or conflicting views presented when appropriate".

Intraclass correlation coefficients (ICCs) for the 8 observed lectures are presented in Table 2. ICCs for all available raters for each lecture ranged from 0.66 to 0.97. For 6 of 8 lectures (75%), the ICCs were above 0.89. All were significant with a $p < 0.001$. For the 7 lectures for which 3 observers were present, individual comparisons between observers produced ICCs ranging from 0.43 to 0.98, with 17 out of 21 (81%) individual comparison ICCs above 0.7. All but 3 ICCs were statistically significant with $p < 0.01$.

Table 2. Intraclass Correlation Coefficients for POET Items 9 through 35 (classroom observation)

	ICC (95% CI)			
Observers	**1, 2**	**1, 3**	**2, 3**	**All Observers**
Lecture 1	0.89* (0.76-0.95)	0.83* (0.58-0.93)	0.85* (0.64-0.94)	0.91* (0.81-0.96)
Lecture 2	0.76* (0.48-0.89)	0.96* (0.90-0.98)	0.73* (0.39-0.88)	0.89* (0.78-0.94)
Lecture 3	0.85* (0.66-0.93)	0.66† (0.24-0.85)	0.51 (-0.10-0.78)	0.78* (0.58-0.90)
Lecture 4	0.88* (0.74-0.95)	0.77* (0.49-0.90)	0.87* (0.70-0.94)	0.89* (0.79-0.95)
Lecture 5	0.93* (0.85-0.97)	0.93* (0.85-0.97)	0.98* (0.96-0.99)	0.97* (0.93-0.98)
Lecture 6	0.73* (0.40-0.88)	0.43 (-0.29-0.75)	0.43 (-0.28-0.74)	0.66* (0.34-0.84)
Lecture 7	0.87* (0.71-0.94)	0.88* (0.73-0.94)	0.84 *(0.64-0.93)	0.90* (0.81-0.96)
Lecture 8 (2 observers only)	0.93* (0.84-0.97)			0.93* (0.84-0.97)

Two-way random effects model; consistency type analysis; average measures
*p ≤ 0.001
†p = 0.005

ICCs were also calculated for each of the subsections of the classroom observation portion of the POET. For the content subsection (5 items), ICCs ranged from 0.53 to 0.99; for teaching strategies and presentation skills subsection (17 items), ICCs ranged from 0.43 to 0.90; for classroom climate subsection (5 items), ICCs ranged from 0.6-1.0. Interrater reliability remained strong for each individual section.

Responses from our external reviewers were favorable. All of the items in the POET were ranked at 3 or above by each reviewer. Additional feedback focused primarily on providing clarifications and adding more descriptors to some individual line items as well as the rubric. Based on this feedback, minor modifications were made to the POET. These changes were made after the pilot testing of the POET, but due to the minor nature of the changes, it did not warrant reevaluation of reliability.

DISCUSSION

We successfully developed a faculty-driven peer assessment program that is both feasible and sustainable. The program is based on a faculty-derived philosophic framework, which we consider to be crucial to the success of the program. We enlisted the aid of an educational expert in our initial stages, who became an integral part of the development process. We highly recommend that other institutions explore available resources and utilize the knowledge and skills of an educational expert.

One strength of our program is the utilization of a pre-observation meeting, which allows for customization of the process to meet the faculty member's specific needs and helps to alleviate any potential anxiety. The post-assessment meeting is also valuable because it allows the observed faculty member to formally examine student mastery of lecture material and to compare what students actually learned to the faculty member's expectations of their learning. This experience can serve as a spring board for continued revision and development, allowing for further refinement of both classroom teaching techniques and assessment procedures/examination writing skills. Another strength of our program is the self-reflection component, which permeates the process, including a formalized faculty reflection held before the post-observation meeting.

We established content validity and good interrater reliability of the POET, making it easily transferable to other institutions. The rubric and item descriptors are clear and straightforward which likely contributed to the good interrater reliability results we observed. Based on an analysis of missing data, a DNO option was added to two POET items. No other changes were deemed necessary.

During initial discussions, the Peer Assessment Taskforce questioned whether content expertise was necessary when pairing peer observers and instructors. The pilot study involved 4 courses and a diverse group of observers and instructors with a wide array of content expertise. The majority of participants agreed that a lack of content expertise did not impede peer assessment, which was further supported by the reliability results. Content is not typically the major focus of peer assessment, and regardless of content expertise, a well-trained observer can provide valuable feedback on lecture delivery, teaching strategies, and assessment of learning. Content expertise by a peer observer may be desirable in certain situations and should be discussed ahead of time.

Our training program focused on appropriate ways to critique and provide constructive feedback utilizing a teaching observation record in order to create an atmosphere which would encourage change. We were advised by our educational expert to avoid completing the POET during the actual classroom observation period, but to complete the teaching observation record with specific comments. This prevented a "checklist" mentality and decreased pressure on the peer observer to document every aspect of the POET during the lecture time. Instead, observers were able to focus more on the observed faculty member's strengths and areas for improvement, as well as on other intangibles such as faculty/student rapport and overall classroom climate. This process then promoted peer-observer reflection when completing the POET. Specific comments from the teaching observation record were integrated into the context of the POET after the classroom experience, providing a number of concrete examples to the faculty member being observed.

Success of a peer-assessment process depends on faculty "buy-in." Our department was open and supportive of this initiative and we did not focus on incentives for participation, although school-wide implementation may necessitate this. Incentivizing the process may make implementation easier at other institutions where faculty members may be hesitant to participate. Incentives may be intrinsic and extrinsic. The primary intrinsic example is the desire of the faculty member to improve upon their teaching skills by obtaining specific feedback about their strengths and areas for improvement as perceived by their peers. While this may be enough motivation for many faculty members, extrinsic incentives, such as fulfillment of a service requirement/committee work or incorporation of peer assessment results into promotion dossiers or yearly merit reviews, may aid in participation and should be explored if deemed necessary. We believe that this peer assessment program could be easily implemented at other colleges and schools of pharmacy.

The developed process is not without limitations. The development, implementation, and sustainability of a peer-teaching assessment program are time-consuming

processes. The pilot study participants identified the challenge of adhering to the entire process, especially the post-assessment meeting and final letter. This is a major concern of the Taskforce, as we believe the post-assessment meeting to be one of the most important and beneficial aspects of our process. There will also be a continuing need for peer observation and training, which necessitates a central coordinator to oversee the logistics of the process. And while we attempted to rigorously analyze POET to ensure interrater reliability and content validity, we do not have construct validity. Interrater reliability was established with members of the Peer Assessment Taskforce serving as peer observers. Because Taskforce members were responsible for developing the POET and no outside observers participated in the pilot study, this could have impacted the interrater reliability results or could limit the ability to generalize our results to other peer observers. Furthermore, interrater reliability was not established for the pre-meeting observation and post-assessment meeting sections because only 1 observer in each observation team completed such sections. POET does not directly assess or incorporate our school's educational ability-based outcomes, which may make it more difficult to objectively demonstrate any impact of our faculty's teaching development on overall student competence in the curriculum.

Many barriers exist for implementing and maintaining effective peer assessment. Faculty members may fear that such assessment will violate some perceived norms of the privacy of teaching.[30] Also, some faculty members prefer being nonjudgmental about others' teaching and refrain from offering themselves as a source of teaching expertise. Another concern is the practicality of the peer assessment process and the time it will add to an already heavy workload. Finally, faculty members may fear bias from personal and professional rivalries, as well as how the skills and personal teaching styles of reviewer may influence the evaluation.[31] Each unit can address these barriers by going through a stepwise approach of developing a practical system that will fit within their institution's culture. Many faculty concerns about bias can be addressed with a valid and reliable evaluation instrument and a formative process.

CONCLUSION

Our peer assessment program for large classroom teaching is comprehensive, feasible, and can be adopted by other schools of pharmacy. POET is a valid and reliable instrument that can be used to assess teaching and learning. The process also meets ACPE accreditation standards regarding multiple means to assess effectiveness and promote both self- and peer assessment of teaching. While the development, implementation and sustainability of a peer assessment program is a time-consuming process, fostering quality teaching through peer mentorship to elicit improved student learning is paramount to the success of every academy.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Mark Douglass, PharmD, Jennifer Kirwin, PharmD, and Tommy Pomfret, PharmD, for their service on the Peer Assessment Taskforce, and Emanuel J. Mason, EdD, MEd, for his assistance with statistical analysis. We would also like to thank our expert reviewers, Lynne M. Sylvia, PharmD, Cecilia M. Plaza, PharmD, PhD, and Nicholas G. Popovich, PhD, for their valuable feedback.

REFERENCES

1 Barnett CW, Matthews HW. Current procedures used to evaluate teaching in school of pharmacy. *Am J Pharm Educ.* 1998;62:388-91.

2 Cashin WE. Student ratings of teaching: The research revisited. IDEA paper No. 32; September 1995. The Individual Development and Educational Assessment Center, Kansas State University. Available at: http://www.theideacenter.org/sites/default/files/Idea_Paper_32.pdf. Accessed November 12, 2008.

3 Barnett CW, Matthews HW. Student evaluation of classroom teaching: a study of pharmacy faculty attitudes and effects on instructional practices. *Am J Pharm Educ.* 1997;61:345-50.

4. Barnett CW, Matthews HW, Jackson RA. A comparison between student ratings and faculty self-ratings of instructional effectiveness. *Am J Pharm Educ.* 2003;67(5):Article 117.

5. Arreola RA. Common questions concerning student ratings: what 80 years of research tells us. In: Arreola RA, ed. *Developing a Comprehensive Faculty Evaluation System.* Bolton, MA: Anker Publishing Inc; 2000: 79-92.

6. Leamon MH, Servis ME, Canning RD, Searles RC. A comparison of student evaluations and faculty peer evaluations of faculty lectures. *Acad Med.* 1999;74:S22-4.

7. Surratt CK Desselle SP. Pharmacy students' perceptions of a teaching evaluation process. *Am J Pharm Educ.* 2007;71:Article 06.

8. Hoyt DP, Pallett WH. Appraising teaching effectiveness: beyond student ratings. IDEA paper No. 36; November 1999. The Individual Development and Educational Assessment Center, Kansas State University. Available at: http://www.theideacenter.org/sites/default/files/Idea_Paper_36.pdf. Accessed November 12, 2008.

9. Gray M, Bergmann BR. Student teaching evaluations: inaccurate, demeaning, misused. Academe Online. Sept-Oct 2003. Available at: http://www.aaup.org/AAUP/pubsres/academe/2003/SO/Feat/gray.htm. Accessed November 12, 2008.

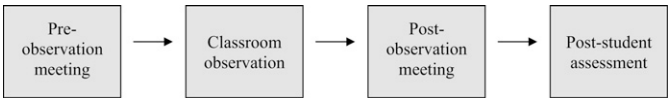
10. Wolfgang AP, Gupchup GV, Plake KS. Relative importance of performance criteria in promotion and tenure decisions: perceptions of pharmacy faculty members. *Am J Pharm Educ.* 1995;59:342-7.

11. Accreditation Council for Pharmacy Education. Accreditation standards and guidelines for the professional program in pharmacy leading to the Doctor of Pharmacy degree. Guideline 26.2. Available at: http://www.acpe-accredit.org/pdf/ACPE_Revised_PharmD_Standards_Adopted_Jan152006.pdf. Accessed November 12, 2008.

12. Northeastern University Faculty Handbook. Available at: http://www.facultysenate.neu.edu/documents/facultyhandbook.pdf. Accessed November 12, 2008.

13. Webb J, McEnerney K. Implementing peer review programs: a twelve-step model. In: *To Improve the Academy, vol. 16*. San-Francisco, CA: Jossey-Bass Publishers; 1997.
14. Qualters D. Observing students in a clinical setting. *Fam Med.* 1999;31:461-2.
15. Schon DA. *Education the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Profession.* 1st ed. San Francisco, CA: Jossey-Bass Publishers 305-44.
16. Chism NVN. Peer Review of Teaching: A Sourcebook. 1st ed. San Francisco, CA: Jossey-Bass Publishers 85-99.
17. Piburn M, Sawada D. Reformed teaching observation protocol (RTOP): reference manual. ACEPT Technical Report no. IN003. Arizona Collaborative for Excellence in the Preparation of Teachers. Available at: http://physicsed.buffalostate.edu/AZTEC/rtop/RTOP_full/PDF/RTOP_ref_man_IN003.pdf. Accessed November 12, 2008.
18. Schultz KK, Latif D. The planning and implementation of a faculty peer review teaching project. *Am J Pharm Educ.* 2006;70:Article 32.
19. Hansen LB, McCollum M, Paulsen SM et al. Evaluation of an evidence-based peer teaching assessment program. *Am J Pharm Educ.* 2007;71:Article 45.
20. Centra JA. Reflective faculty evaluation: enhancing teaching and determining faculty effectiveness. 1st ed. San Francisco, CA: Jossey-Bass Publishers; 205-13.
21. Braxton JM, Bayer AE, Finkelstein MJ. Teaching performance norms in academia. *Res Higher Educ.* 1992;33:533-69.
22. Nelson MS. Peer evaluation of teaching: an approach whose time has come (commentary). *Acad Med.* 1998;73:4-5.
23. Feldman P. Faculty performance reviews: accountability in teaching education. *Education.* 2005;125;349-52.
24. Rosenbaum ME, Ferguson KJ, Kreiter CD, Johnson CA. Using a peer evaluation system to assess faculty performance and competence. *Fam Med.* 2005;37:429-33.
25. Koops JB, Winsor KA. Creating a professional learning culture through faculty evaluation. *J Educ.* 2005;186:61-70.
26. Cohen PA, McKeachie WJ. The role of colleagues in evaluation of teaching. *Improving Coll University Teaching.* 1980;28:147-54.
27. DeZure, D. Evaluating teaching through peer classroom observation. In: Seldin P, ed. *Changing Practices in Evaluation Teaching.* San Francisco, CA: Jossey Bass Publisheres. 1999: 70-96.
28. Shrout, PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull,* 1979;86:420-8.
29. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1:30-46.
30. Shulman LS. Teaching as community property: putting an end to pedagogical solitude. *Change.* 1993;25:6-7.
31. Chism NVM. Why introducing or sustaining peer review of teaching is so hard and what you can do about it. *The Department Chair.* 2007;18 (2):6-8.

Appendix 1. Peer observation and assessment process.



1. **Pre-observation meeting** to review objectives, handouts, learning strategies, and teaching pedagogy (occurs 1 week before classroom observation).
    a. **Faculty member** should provide lecture materials to peer observer at least 1 week prior to this meeting for review
    b. **Peer observer** should utilize the scripted interview questions in the POET to complete Section 1

2. **Classroom observation:** the observation should last the entire length of the lecture
    a. **Peer observer** notes positive aspects of the lecture and identifies potential areas, if any, for improvement utilizing a "teaching observation record". Section 2 of the POET should be completed as soon as possible *after* attending the class. The peer observer should *not* complete the POET during the lecture. Three main focus areas for discussion with the lecturer should be identified.

3. **Post-observation meeting** (within 1-2 weeks)
    a. **Faculty member** should reflect on the lecture and complete Section 3 of the POET prior to the meeting with the peer observer
    b. **Peer observer** should give the opportunity to the faculty member to discuss his/ her reflection on the lecture and discuss strategies to improve lecture content, delivery, and assessment. Peer observer should provide positive feedback and constructive evaluation points, limited to 3 main focus areas.

4. **Post-assessment meeting** (within 2 weeks) of the first major examination following the classroom observation to review lecture assessment and student achievement of learning outcomes.
    a. **Faculty member** should submit original questions submitted for the exam, final version of exam questions (if applicable), and results of item analysis to the peer observer prior to this meeting
    b. **Peer observer** should complete Section 4 of the POET prior to this meeting and provide final comments and recommendations (limited to 2-3). A written letter is subsequently provided to the faculty member highlighting the major aspects of the peer assessment process, including the strengths of the lecturer and the 3 main areas of focus.