Summative assessment in a doctor of pharmacy program: a critical insight

Kerry Wilbur
College of Pharmacy, Qatar University, Doha, Qatar

Background: The Canadian-accredited post-baccalaureate Doctor of Pharmacy program at Qatar University trains pharmacists to deliver advanced patient care. Emphasis on acquisition and development of the necessary knowledge, skills, and attitudes lies in the curriculum’s extensive experiential component. A campus-based oral comprehensive examination (OCE) was devised to emulate a clinical viva voce and complement the extensive formative assessments conducted at experiential practice sites throughout the curriculum. We describe an evaluation of the final exit summative assessment for this graduate program.

Methods: OCE results since the inception of the graduate program (3 years ago) were retrieved and recorded into a blinded database. Examination scores among each paired faculty examiner team were analyzed for inter-rater reliability and linearity of agreement using intraclass correlation and Spearman’s correlation coefficient measurements, respectively. Graduate student ranking from individual examiner OCE scores was compared with that of other relative ranked student performance.

Results: Sixty-one OCEs were administered to 30 graduate students over 3 years by a composite of eleven different pairs of faculty examiners. Intraclass correlation measures demonstrated that examiner team reliability was low and linearity of agreements was inconsistent. Only one examiner team in each respective academic year was found to have statistically significant inter-rater reliability, and linearity of agreements was inconsistent in all years. No association was found between examination performance rankings and other academic parameters.

Conclusion: Critical review of our final summative assessment implies it is lacking robustness and defensibility. Measures are in place to continue the quality improvement process and develop and implement an alternative means of evaluation within a more authentic context.

Keywords: pharmacy education, educational measurement, quality improvement

Introduction
Demand for higher education is increasing worldwide and precipitating international partnerships that reflect standardization of health sciences education across borders and the desire to emulate perceived global leaders in this regard.1-3 In the last decade, a number of Gulf Coast Corporation countries have experienced marked economic growth and in turn are devoting significant resources to augmenting provision of health care services to their populations with parallel investments within the health education sector. The Canadian-accredited College of Pharmacy at Qatar University began offering an undergraduate curriculum conferring a Bachelor of Pharmacy degree, the first degree to practice in Qatar, as well as a post-baccalaureate graduate degree in advanced clinical pharmacy practice in the falls of 2007 and 2011, respectively.
The Canadian-accredited full-time Doctor of Pharmacy (PharmD) degree is a 36-credit program open to Qatar University pharmacy graduates that includes 32 weeks of experiential training (eight internships, each 4 weeks in duration, representing 32 credits) with pharmacist mentors in Qatar. Graduate students are also enrolled in a research evaluation and presentation course each semester (four credits). A part-time PharmD program study plan is open to eligible pharmacists practicing in the country whereby they complete the aforementioned internships and courses and up to 25 additional preparatory credit hours.

PharmD training supports an advanced pharmacy practice model whereby pharmacists are integrated members of the multidisciplinary health care team who collaborate with other clinicians in the management of patients. Such pharmaceutical care may include patient chart review, patient interview and education, ordering and interpretation of laboratory tests, physical assessment, formulation of clinical assessments identifying potential or actual drug-related problems, development and implementation of therapeutic plans according to best available evidence, and patient follow-up to evaluate the safety and effectiveness of drug therapy.5,6 For post-baccalaureate PharmD students to acquire and practically develop these competencies in the graduate program curriculum, emphasis needs to be placed on the experiential components supervised by pharmacist mentors who demonstrate pharmaceutical care in an advanced clinical practice (known as “preceptors”).6,7

Structured formative assessment of these clinical graduate students occurs continuously throughout the professional program. Mid-point and final internship evaluations are completed by the advanced practice-based preceptors according to 25 predetermined criteria mapped to the 168 relevant student learning objectives across the curriculum.7,8 Similarly, the PharmD program campus-based faculty coordinating the research evaluation and presentation course uses several methods to assess graduate student comprehension and associated learning needs, including projects, skill checks, and exercises. Such multiple means help program faculty understand students’ academic progress and respond with supportive adjustments; however, the internship experience is not uniform in that no two students are enrolled in the same specific advanced clinical rotations or with common preceptors. A common summative integrated assessment of student achievement at the end of our professional program was deemed necessary.

We implemented a campus-based oral comprehensive examination (OCE) to replicate a clinical viva experience. In PhD programs, the viva voce is an oral examination characterized by interaction between a graduate student and multiple examiners, and serves as the final defense of submitted thesis work.9,10 Likewise, the purpose of our assessment is to evaluate the examinee’s critical reasoning through expressed synthesis and interpretation of patient data and associated judgments pertaining to case management. However, as our program matures and the literature evolves, we are obligated to examine if this final summative assessment approach is robust and defensible. We present data from our existing assessment and identify the need for a new evaluation strategy.

Materials and methods
The OCE for PharmD students at Qatar University’s Canadian-accredited College of Pharmacy was modeled after the format of another graduate PharmD program in North America.11 Two course coordinators in the PharmD program collaborate to create a single patient case scenario featuring the presentation of a high priority acute comorbidity (eg, life-threatening) as well as three additional active medical and associated drug-related problems. The examinations undergo internal review between at least four other clinical faculty members who participate in direct patient care in Qatar. The three-stage examination emphasizes the following skill areas in drug-related problem-solving processes in advanced patient care: data gathering; drug-related problem identification and prioritization; determination of viable treatment alternatives; development of a pharmacy care plan that includes evidence-based recommendations for management of drug-related problems, as well as specific patient and drug monitoring parameters; and, finally, overall verbal and written communication.

Data gathering and drug-related problem identification and prioritization
In the first 30-minute closed-book phase of the assessment, the PharmD student joins two examining faculty in an examination room and is given a written patient case summary, which includes a brief history of the present illness, some laboratory values and physical findings, and other medical history. The graduate student reviews the content and may ask examiners for additional relevant missing information (key points have been intentionally omitted in the summary provided to the student). Once the student perceives this data gathering process to be exhausted and has no additional questions, he or she is then provided with a sheet containing all available patient case information, including any data they may not have
Drug-related problem work-up
During the next phase of open-book individual self-study, the graduate student is supervised in a private room to work up the patient case. This work-up involves identifying information provided in the patient case to justify identified and prioritized drug-related problems and arrive at patient-specific goals of therapy. Graduate students are expected to search primary literature to arrive at the best alternative to resolve each drug-related problem. These recommendations must be specific, including the dose, route, and duration of therapy. Graduate students must prepare to discuss all alternatives and provide a rationale for the selected treatment plan. Patient and medication monitoring plans for both effectiveness and safety must be devised, including the timing, frequency, and duration for the given parameter.

Before leaving this phase, graduate students must also prepare a written pharmacy note addressing one of the three drug-related problems that would be placed in the patient’s medical chart. The pharmacy note is written according to a predetermined and practiced structured format and will be graded with a rubric familiar to the students and contributes to the overall examination score.

Pharmaceutical care plan presentation
In the final 30-minute phase of the assessment, the student rejoins the two examining faculty to present recommendations to solve the identified drug-related problems, including patient-specific goals of therapy, the therapeutic alternatives considered, the ultimate recommended regimen, primary literature to support the evidence-based recommendations, and the patient and medication monitoring plan. During this final encounter, the students are evaluated on the appropriateness of expressed patient-specific goals of therapy, thoroughness of alternatives considered, quality of the primary literature evaluated, accuracy of the final recommendation, and organization and completeness of the monitoring plan.

This examination is administered twice during the final post-baccalaureate PharmD academic year (AY); once as a low-stakes assessment in the fall semester and again as a high-stakes exit assessment in the spring semester at completion of the program. Each examination delivery is a distinct patient case iteration (different comorbidities and drug therapy and associated drug-related problems).

For this quality assessment study, the results of all OCEs administered since the PharmD program inception were evaluated. Archived examiner scoring sheets were retrieved and recorded into a blinded database. The six individual skill components over the three stages (described above and in Figure S1) and overall scores for each graduate student were entered.

Scores were compared between each paired examiner team for each discrete examination period. The reliability of raters was determined by calculating intraclass correlations (ICCs) and associated 95% confidence intervals using a two-way mixed model (absolute agreement type). Statistically significant ICC values of greater than 0.7 are considered optimal, with greater than 0.9 indicating excellent agreement. Linearity of agreement for examiners was also determined using Spearman’s correlation coefficient measurements. Finally, graduate student ranking from individual examiner scores in the OCE was compared by Kendall’s tau with ranked student performance in the PharmD program according to two other assessments, i.e., the PharmD admission undergraduate grade point average and scores obtained in the program research evaluation and presentation course. All statistical tests were two-tailed and based on a significance threshold of \( \alpha < 0.05 \). The analysis was conducted using Statistical Package for the Social Sciences version 21.0 software (SPSS Inc, Chicago, IL, USA).

Results
Thirty graduate students enrolled in the PharmD program have participated in an OCE since its inception in 2011 (AY11). During these 3 years, 61 OCEs were administered. Four faculty examiners were paired in AY11, five in AY12, and seven in AY14, resulting in a total of eleven combinations of examiner teams (Table 1). Four examiner pairs were reteamed for different graduate student cohorts, sometimes in different AYs.

Examiner team reliability is reported in Table 2. When overall OCE scores were considered, we have evidence to support the reliability of the measurements between members of both assessor groups in AY11. Examiner team 1 demonstrated statistically significant, but decreasing magnitude of agreement, in the two examination sittings (ICC 0.811, \( P = 0.014 \) in the first spring semester examination [SS1] and ICC 0.570, \( P = 0.024 \) in the second spring semester examination [SS2]). Examiner team 2 had statistically
significant and almost perfect agreement according to the ICC of their two separate examination sittings (ICC 0.95, \(P=0.006\) in SS1; ICC 0.926, \(P=0.015\) in SS2). In AY12, only one of the three assessor groups demonstrated statistically significant reliability. Examiner team 4 demonstrated almost perfect agreement (ICC 0.99, \(P=0.001\)). Finally, of the seven distinct assessor pairs in AY13, just two demonstrated levels of statistically significant reliability, one of which was low: ET-8 (ICC 0.571, \(P=0.031\)) and ET-11 (ICC 0.99, \(P=0.001\)). However, when all scores from separate examination sittings for reteamed assessor groups were combined, examiner teams 3 and 5 demonstrated high overall reliability (ICC 0.98, \(P=0.011\), and ICC 0.94, \(P=0.001\), respectively).

Linearity of agreement was found to be inconsistent across examiner teams in each year (Table 2). When OCE performance rankings were compared with the graduate school admission grade point average and with the research evaluation and presentation course, statistically significant relationships were found in only four instances (Table 3).

### Discussion
Like established US programs challenged to find reliable means of determining achievement of student learning outcomes, our nascent program is also seeking to purposively devise valid assessments of student abilities.\(^{13,14}\) Multiple means for evaluation are available and include use of written examinations,
Evaluating the final exit assessment in a PharmD program

Table 3 Comparative graduate student ranking between oral comprehensive examination scores and other program assessments

<table>
<thead>
<tr>
<th>Students</th>
<th>Comparative assessments</th>
<th>Kendall’s τ</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AY11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE1 and GPA</td>
<td>0.50</td>
<td>0.083</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.143</td>
<td>0.652</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE1 and course</td>
<td>0.429</td>
<td>0.138</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.643</td>
<td>0.138</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE2 and GPA</td>
<td>0.659</td>
<td>0.090</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.609</td>
<td>0.026</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE2 and course</td>
<td>0.270</td>
<td>0.281</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.171</td>
<td>0.527</td>
</tr>
<tr>
<td>AY12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE1 and GPA</td>
<td>0.524</td>
<td>0.999</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE1 and course</td>
<td>0.143</td>
<td>0.652</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE2 and GPA</td>
<td>0.333</td>
<td>0.293</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.143</td>
<td>0.652</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE2 and course</td>
<td>0.333</td>
<td>0.293</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.143</td>
<td>0.652</td>
</tr>
<tr>
<td>AY13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE1 and GPA</td>
<td>−0.013</td>
<td>0.951</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>−0.245</td>
<td>0.246</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE1 and course</td>
<td>0.529</td>
<td>0.012</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.400</td>
<td>0.058</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE2 and GPA</td>
<td>0.013</td>
<td>0.951</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.194</td>
<td>0.359</td>
</tr>
<tr>
<td>Examiner 1</td>
<td>OCE2 and course</td>
<td>0.364</td>
<td>0.086</td>
</tr>
<tr>
<td>Examiner 2</td>
<td></td>
<td>0.545</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Notes: The GPA is the undergraduate record at admission; course is the final score of the campus-based research evaluation and presentation course at the conclusion of each semester.

Abbreviations: AY, academic year; GPA, grade point average; OCE, oral comprehensive examination.

assessing by supervising preceptors, direct observation, clinical simulations (such as oral structured clinical examination stations with standardized patients), multi-source assessments, and portfolios. Our curriculum is consistent with other advanced pharmacy practice training programs in its emphasis on situational learning through internships. However, reliance on internship preceptor reports as reliable summative assessments may be limited by incomplete documentation, conflicts of interest, or bias. The OCE was intended to complement these internships and the course-based multi-method assessment strategies in one final exit examination.

The small number of graduate students in our sample population disadvantages our evaluation and consequently the statistical analysis lacks robustness. Despite this, the retrospective evaluation of our OCE raises concerns regarding its utility as a valid and defensible high-stakes summative assessment. We found essentially no correlation between OCE grades and other indicators of academic performance. Having said this, admission grade point average might not be a sufficient predictor of OCE performance. In AY13, for example, practicing pharmacists enrolled in the part-time PharmD program reached the internship phase and participated in the OCE. For some, there has been as much as a decade since they completed their undergraduate degree from non-North American-oriented programs and so their high admission grade point average may be inconsistent with their abilities to excel in a patient-centered program. Lack of any correlation with the on-campus program course is more unexpected, but may be attributed to the majority of the graduate pharmacy training being internship-based with a diverse group of advanced clinical practice preceptors. These internship evaluations could be a more useful predictor of OCE performance, but offer less discretion among student outcomes, given their categorical outcome of pass/fail, and therefore present difficulties for comparison.

From our small dataset, it is clear that reliability is lacking. Despite clear grading guidelines, influence of different judgments of faculty on inter-rater reliability seems to be exacerbated by the number and various compositions of examiner teams. Other potential sources for this observed disparity includes the examiner team compositions of both practice-based and campus-based clinical faculty who may have different perspectives on the scoring of patient-related decision-making, as well as PharmD faculty who have been exposed to this type of assessment in their own past training compared with clinical PhD faculty who have not. While the final phase may represent the examination element of highest fidelity to a patient care setting (the efficient consideration and execution of patient care decisions) it is possible that the allotted time (30 minutes) allows insufficient opportunity for students to adequately articulate thorough assessment, justified management, and monitoring plans for three distinct issues. In its current summative form, we also forego the opportunity to offer student-specific formative feedback on correctable inconsistencies or to clarify the grading outcome.

Undergraduate health professional curriculums’ integration of simulated experiences in their assessment strategy is important and can be effective learning opportunities. However, by virtue they are not contextual and can lack genuineness. For example, while among the merits of the OSCE format is that multiple students are assessed according to identical and predetermined criteria on the same scenario, so too are these features its detriment. The OSCE may become predictable, resulting in rehearsed performances and offering...
little indication of how students may perform in an actual patient care environment.\textsuperscript{21,22}

An evolving body of research supports the use of more authentic assessment strategies in health professional training programs. Nursing curricula throughout Europe are especially active in adoption of models evaluating students in actual patient care environments.\textsuperscript{20–24} There is a profusion of similar literature from medical school curricula regarding the development and testing of in-training evaluation strategies.\textsuperscript{25,26} While these health professions have devised assessment tools to document observed skills within the students’ clinical practice, these instruments have fixed predetermined criteria and associated checklists. Examiners judge and record against observed student tasks and actions or make numeric conversions of student performance into scores and grades. Checklist criticism may be rooted in perceived flaws regarding its underlying epistemology, ie, that full assessment of student ability may be deconstructed and assigned a numeric value.\textsuperscript{27,28} Other arguments include that when examiners focus on checking such isolated items (as opposed to evaluations of the full context of care) the assessment may overlook or obscure other necessary domains of patient management, such as empathy or caring.\textsuperscript{29,30} As such, attention has shifted to the use of narrative descriptions to replace grades and ratings as a framework for assessment of clinical performance in medical education.\textsuperscript{31,32} Narrative descriptions stem from observations relating to all aspects of patient care and are shared with students by faculty as a means of ongoing formative assessment of their learning. This timely and specific feedback offers credible judgments of student abilities that are not consistently captured through traditional experiential supervision. Emphasis on qualitative evaluation assimilates social learning theory, consistent with authentic constructivist/interpretist approaches to in-training assessments that have been recently argued.\textsuperscript{28} Subsequent to these findings, our program is now embarking on efforts to repurpose our current OCE and develop a contemporary final formative and summative assessment for clinical pharmacy graduates exiting our PharmD program. For our post-baccalaureate graduate students who have already met the required competencies for pharmacist licensure, but must demonstrate achievement of the clinically oriented student learning outcomes of our program, these may be considered superior to other means of assessment.

**Conclusion**

Critical review of the final summative assessment for graduate clinical pharmacy students in our PharmD program identified deficiencies in its validity and reliability. We continue the quality improvement process of exploring alternative means of evaluation within a more authentic context.

**Disclosure**

The author reports no conflict of interest in this work.

**References**


Supplementary material

Figure S1 Oral comprehensive examination scoring scheme.

Phase 1
Data gathering, problem identification, and initial recommendations (35 marks)
  - Ability to identify missing case information (15 marks)
  - Systematic approach to information gathering (10 marks)
  - Ability to appropriately prioritize most important problems (5 marks)
  - Appropriateness of initial recommendations (5 marks)

Phase 2
Independent case work-up

Phase 3
Presentation of action plan and justification (55 marks)
  - Principal drug-related problem (25 marks)
    - Goals of therapy (5 marks)
    - Therapeutic alternatives (5 marks)
    - Recommendation and rationale (7 marks)
    - Safety and effectiveness monitoring plan (7 marks)
  - Second drug-related problem (15 marks)
    - Goals of therapy (3 marks)
    - Therapeutic alternatives (2 marks)
    - Recommendation and rationale (5 marks)
    - Safety and effectiveness monitoring plan (5 marks)
  - Third drug-related problem (15 marks)
    - Goals of therapy (3 marks)
    - Therapeutic alternatives (2 marks)
    - Recommendation and rationale (5 marks)
    - Safety and effectiveness monitoring plan (5 marks)

Chart documentation (10 marks)