Application of Objective Structured Clinical Examinations in an Assessment of Pharmacists’ Continuing Competence¹,²


INTRODUCTION
For over 20 years, the pharmacy licensing authority in British Columbia, the College of Pharmacists of B.C. (CPBC), has been developing procedures to assess and assure the continuing competence of its members. The proposed model investigated in this research began with a paper and pencil assessment of practice knowledge which served as a general screen of competence. Those not meeting the standard set for the general screen would be required to undergo a detailed assessment of performance. Those not meeting the standard established for the performance assessment would in turn be required to successfully complete specific remedial activities in order to continue licensure as a practicing pharmacist. The full model is outlined in Figure 1. Previous reports have detailed the identification and validation of the competency blueprint for the re-licensure process(1), the development and validation of the assessment of practice knowledge(2), and the establishment of a passing score for the knowledge assessment(3).

The present paper will provide details of the activities in the shaded area of Figure 2 related to the development and validation of the performance assessment instrument and administration procedure. The research questions specific to this part of the investigation were related to the number of simulated performance situations and the assessment time required to produce performance scores which could be reliably and validly interpreted for re-licensure decisions. As well, the acceptability and feasibility of the performance assessment as part of the proposed competency assessment program were assessed.

PERFORMANCE ASSESSMENTS
Currently, performance assessment is a popular topic in general education(4-7) and its application in “high-stakes” testing has been the focus of a number of symposia and special sessions at meetings of educators³. In part, this attention is due to claims that performance assessments are “direct” and “authentic” and permit the measurement of “cognitively complex” material(4,6). Consequently, there has been considerable discussion of the validation criteria to be applied in performance assessments(4,6,8) as well as critical appraisals of their suggested advantages over more traditional forms of assessment(5,9). To date, the advantages of performance assessments are largely unproved and the implication that other forms of assessment are less authentic and indirect has been branded as “nefarious”(10). Mehrens suggested that, with the current available evidence, some individuals have “… put on their advocacy hats before the data support it”(11). He did however add that performance assessments should be a part of the evaluation system in making licensure decisions because of “the high costs of false positives.” Similarly, after reviewing the application of four performance-based assessment methods in the health professions, Swanson, Norman and Linn(12) concluded a blend of knowledge and performance-based assessments would be both psychometrically and educationally desirable.

Objective Structured Clinical Examinations
In its earlier competency assessment programs(13), the CPBC used a peer review procedure to evaluate performance. Independent evaluations of that activity concluded...
that the peer review approach should be altered to assure adequate sampling of performance and to enhance its standardization. Therefore, as part of the research described here, an alternative approach was investigated.

The procedure adopted was the Objective Structured Clinical Examination (OSCE)(14,15). This assessment procedure has been used in medical training to assess specific content areas(16) and as a part of comprehensive licensure evaluations(17). In pharmacy, OSCEs have been used in the United Kingdom to assess the interviewing and counseling competence of undergraduate students(18) and to identify continuing education needs for practitioners(19).

Typically, an OSCE is composed of a series of “stations.” The candidates move from station to station solving problems derived from pharmacy practice. The problems could be presented in a written format or they can involve “standardized” patients. The problems could also require the execution of skills such as prescription compounding. For each station, the candidate’s performance is recorded by using a checklist of possible behaviors needed to solve the problem. The checklists are completed either by the standardized patients or by independent raters. Subsequently, the checklists are scored according to a predetermined scoring rubric to permit evaluation of an individual’s performance during the OSCE.

In an examination of the psychometric characteristics of OSCEs used with final-year Australian medical students, Newble and Swanson(20) observed that the OSCEs generally had lower reliabilities than other assessment formats of comparable length. They concluded that this finding was more likely attributed to the variability among stations due to case specificity rather than poor interrater reliability. They projected that an OSCE of six hours may be needed to obtain an “acceptable” level of reliability. As a consequence, Newble and Swanson suggested that short OSCEs (one to two hours) are not sufficiently reliable to be used alone in high stakes situations. They recommended that OSCEs should be combined with other tests to obtain acceptable reliability levels.

**Standardized Patients**

As indicated above, OSCEs often include the use of standardized patients (SPs) in some or all of the station problems. A SP is usually a member of the general public that has been trained to present a particular health-related problem in a consistent and accurate manner(21). These individuals can be trained to present rather complex medical problems in a convincing fashion. They are either selected because they have certain conditions or symptoms relevant to the situation of interest or they can be trained to simulate the situation of interest(21).

Standardized patients can also be trained to complete a checklist of behaviors that occur during each session in which they are the “patient” or client. As well as including a listing of the behaviors relevant to the solution of the problem, the checklist often permits a record of the communication and counseling performed during the session. Information recorded on the checklists can be used both for instructional and evaluation purposes.

Again, medicine seems to have the most experience with the use of SPs. Recent reviews have described their applications and summarized evidence supporting their use(21 -23). With proper training, SPs have been found to be accurate and consistent in the presentation of their simulations and in their recording of information.

**METHODOLOGY**

**Development of the OSCEs.**

Practicing pharmacists were recruited to construct the situations to be used as OSCE stations. These individuals were representative of pharmacy practice in B.C. and had experience in some aspect of the College’s previous performance assessments.

To provide an orientation to the OSCE format and development, a workshop was held under the direction of an experienced facilitator in May 1990(4). The workshop began with a structured discussion of OSCEs and their use. This was followed by a practice session in which the pharmacists worked in small groups to develop potential station scenarios. After the practice session, all scenarios were critically appraised by the workshop participants and the facilitator. Following this initial workshop, the participants formed a working group that met regularly to develop and review stations. In developing the station material, the pharmacists were asked to draw upon their practice experience to make the “problem” as realistic as possible, to identify required physical and human resources, to suggest a list of important pharmacist behaviors to be included in the accompanying performance checklist, and to develop any

---

4 This training session was coordinated by Dr. Arthur Rothman of the Faculty of Medicine, the University of Toronto, Toronto, Canada.
instructions necessary for SPs.

Initially, the plan was to develop each station stimulus or problem and the accompanying checklist of behaviors so that each station assessed one of the 12 competencies (Table I) previously identified and validated.(1) However, station authors found this was not possible if they were to maintain the realism of the station. Consequently, more than one competency was typically assessed at one station. In these cases, the station authors were asked to indicate the competency that related to each checklist behavior.

In total, 63 potential stations were created. An example of the development of one of these stations is presented in Appendix A. Part 1 contains a summary of the worksheet used to develop one station; Part 2 contains an early draft of an accompanying checklist for this station; and Part 3 presents the problem as the candidate encountered it during the assessment.

**Standardized Patients and Physicians.**

A number of the stations required individuals to role-play patients (SPs). In addition, some stations required “standardized” medical doctors (SMDs). Members of the general public were recruited to role play the SPs. Pharmacists were chosen to role play the SMDs. All SPs and SMDs were trained by experienced trainers to play their roles for specific stations. Both the SPs and the SMDs were observed and coached by the trainers until they were able to play their roles in a consistent and correct manner. For the pilot-test described in this study, a total of 38 individuals were used as SPs (n=19) or SMDs (n=19).

The SP and SMD were also responsible for completing the checklist after a candidate completed the OSCE station. Therefore, as part of their training, the SPs watched two video tapes of pharmacist-patient interactions and then completed a checklist for each one, recording what the pharmacist did and/or said during the interaction. Since the SMDs would be interacting with the pharmacists over the telephone, they listened to two audio tapes of pharmacist-physician interactions and then completed a checklist for each interaction.

To determine the accuracy of the SPs and SMDs in recording what they saw or heard, their checklists were compared with what actually was on the video and audio tapes. For the SPs, the mean percent accuracy was 82.0 percent with a standard deviation of 12 percent; for the SMDs, the mean percent accuracy was 79.5 percent with a standard deviation of 8.4 percent. These results met the suggested minimum of 80 percent and are similar to those reported in a review of the use of SPs in clinical assessments.(22)

**Pretests.**

In April and May 1991, all 63 stations were pretested with groups of volunteer pharmacy practitioners. The purpose of this pretest was to identify the need for revisions, if any, in the problems and checklists; to select the stations to be included in the final form of the OSCE; and, to assess the overall procedures to be used in the final assessment. The candidates were volunteer pharmacists who had expressed an interest in the OSCE process or who were involved in other College licensing activities.

The stations were grouped so that 20 stations with varying lengths (5 to 15 minutes) comprised an OSCE. The full OSCE required 3.5 hours to complete. This time included an initial orientation, the actual assessment, movement from station to station, breaks, and a final debriefing. The number of stations per OSCE (i.e., 20) was set to permit adequate coverage of the competency blueprint as well as to allow two complete OSCE’s to be carried out in one day. As well, Niewble and Swanson(20) had indicated that shorter OSCEs are unlikely to be sufficiently reliable to use in high stakes testing.

**Construction of Final OSCE Forms.**

Applying the results of the pre-tests, two “parallel” forms of the OSCE (A and B) were constructed. Each contained 20 stations, four of which were common to both forms. Given the desire that the same competencies be equally represented in both forms, pairs of stations which assessed essentially the same competencies were selected and randomly assigned to the two OSCE forms. Care was also taken to insure that the patient problems and therapeutic categories of drugs appearing in each form would be balanced.

Each form of the OSCE had 13 stations of five minutes each, three stations of eight minutes each, three stations of 10 minutes each and one station that took 15 minutes. Approximately three hours were required to complete each OSCE. This time included a period of time for orientation and debriefing. As structured, 19 candidates were assessed during a single OSCE session.

**Station Content Validation.**

After completion of the development of the two OSCE forms, a validation study was conducted to insure that the stations within these forms were assessing the areas intended by their authors. Panels of 5 or 10 practicing pharmacists were formed. Each panel was assigned different stations and panelists were asked to assign each behavior in the corresponding station checklists to one of the 12 competency areas. Agreement of 80 percent or more on the assignment given the individual behaviors was taken as consensus. Items with less than 80 percent agreement were discussed by the entire panel and then panelists repeated the assignment procedure. Consensus was reached on all items.

To further validate the OSCE’s and their stations, the validation panelists’ assignments were compared with the station authors’ assignments. There was agreement for 682 (85.7 percent) out of 796 checklist items. For the remaining 114 items, a number of behaviors initially assigned to competencies 2 and 9 by the station authors were reassigned to competency 5 by the panel members. This reassignment may be explained by the fact that verification of the prescription (competency 2) and fulfilling legal documentation requirements (part of competency 9) are also a part of the dispensing function (competency 5). Further, the weighting for competency 5 is higher than the initial target given the large number of behaviors required to satisfactorily complete the dispensing function (including compounding). See Table I.

**Weighting of Checklist Behaviors.**

Thirty of the 55 pharmacist participants in the OSCE pretests were asked to assign scoring weights to the behaviors contained in the station checklists. The following five point weighting system was used: +2 = must do - has to be done for the problem to be considered solved; +1 = should do - contributes to solving the problem; 0 = could do - does
The level of agreement was recomputed. A final weighting was assigned by each judge; and the weight given; pretest response frequency data were provided; a final weighting was assigned by each judge; and the level of agreement was recomputed.

Behaviors without consensus at this point were tabled until a subsequent meeting in which weightings were completed by two additional groups of five judges. These groups were composed of pharmacists who had participated in the weighting process previously as well as pharmacists new to the process. The procedure followed was as outlined above. For those few behaviors for which 80 percent consensus was not reached at this point, the weight assigned to the behavior was that of the majority. The final scoring rubric for each station was subsequently determined by additional panels at the time standards for passing each station were established.

Pilot Tests of Final OSCE Forms

Altogether, eight pilot test sessions spread over five days were conducted. The first two, on February 15, 1992, were held to assess final year pharmacy students. For these sessions, one form of the OSCE was administered in the morning and the other OSCE form was administered in the afternoon. The next four, morning and afternoon sessions on April 11 and May 2, 1992, were held for pharmacists. Two additional sessions were held for pharmacists, one each on April 12 and May 3, 1992.

Both student and pharmacist samples were selected as part of the validation study of the OSCEs. It was hypothesized that the students would perform less well than the pharmacists with the difference attributable to the difference in pharmacy practice experience between the two groups.

For the students, 38 individuals were selected according to the process outlined below. The number of students selected was limited to the maximum number of individuals that could be assessed in one day. For the pharmacists, a target of 100 individuals was set. To compensate for nonresponse, a 30 percent oversampling was included as outlined below.

OSCE Samples. As illustrated in Figure 2, 306 pharmacists completed one of two forms of the knowledge assessment instrument. Likewise, 81 final year pharmacy students completed one of the two forms. For both groups, since the mean scores for the forms of the knowledge assessment differed significantly(2), the raw scores were adjusted to a common mean by subtracting the mean score of the exam form (A or B) for the day of the week the assessment was written. Once adjusted, a tapered sampling design was used for each group to insure a broad representation of knowledge scores among those who completed the OSCE’s while at the same time ensuring a larger sample of individuals who scored below the standard set for the knowledge assessment.

For the student sample, the centered mean scores were sorted from lowest to highest. Beginning with the lowest and proceeding to the highest, the first 40 percent of the students were divided into four deciles (n=8 each) and the remainder were divided, in increasing order, into three groups containing respectively, 17, 16, and 16 students. Within each decile or 20 percent grouping, the individuals were numbered 01 to 08 or 01 to 16/17. A table of random numbers was then used to select 38 individuals using a design that resulted in more those who completed the OSCE’s while at the same time ensuring a larger sample of individuals who scored below the standard set for the knowledge assessment.

The selected students were assessed with the OSCE’s approximately one month after the knowledge test. Of the students who completed an OSCE, 21 were part of the initial sample and four were substitutes.

For the pharmacist sample, the adjusted scores were arranged from lowest to highest and separated, beginning with the lowest, into four deciles and three 20 percent groupings. All pharmacists in the first two, or the lowest, deciles...
were selected (N=62). Within the remaining two deciles and the three 20 percent groupings, the pharmacists were numbered consecutively. The tapered sampling procedure used selected individuals from the third decile (N=21), fourth decile (N=17), first 20 percent (N=13), second 20 percent (N=9) and third 20 percent (N=9) groupings. The selected pharmacists were contacted and asked to participate in the OSCE. They were permitted to choose which of the six pharmacist OSCE sessions they would like to attend. If individuals declined to participate, they were replaced, in the case of the upper two deciles and the 20 percent groupings, by randomly selecting a substitute from the same subgroup.

Altogether, 168 pharmacists were either initially selected or selected as a substitute. Of this number, 99 pharmacists agreed to participate and 96 of these pharmacists completed the assessment.

Administration. The OSCEs were administered at the Faculty of Pharmaceutical Sciences, the University of British Columbia. The stations were set up in faculty offices and the dispensing laboratory. The candidates moved from station to station when prompted by a signal. An honorarium was paid to all students ($60) and pharmacists ($150). As well, travel and accommodation costs incurred by out-of-town pharmacists were reimbursed.

Scoring. For each pilot-test participant, individual station scores were calculated applying the scoring rubric developed for each station. Station scores were then summed to produce a total OSCE score for that individual. Participants were informed of their total score as well as the mean score and the range of scores for the form of the OSCE they completed.

Acceptability and Feasibility. The acceptability of the OSCE instruments and procedures was ascertained as a part of a study of the acceptability of the total program for assessing the continuing competence of practicing pharmacists. In this study, a 52-item questionnaire was developed and mailed to all 306 pharmacists who completed the assessment of practice knowledge. The 96 pharmacists who participated in the pilot test of the OSCEs were asked to answer questions specifically related to the OSCEs. The logistic and economic feasibility of the OSCEs were investigated by an analysis of data gathered during the developmental, pre-test, and pilot test phases for the OSCEs.

RESULTS AND DISCUSSION

Harden and Gleeson(15) recommended that any procedure to be used to assess clinical competence should be evaluated in terms of three attributes of the assessment procedure. These attributes are the reliability (dependability), validity, and practicality of the procedure. These attributes will be discussed in that order for the OSCEs that were pilot tested.

Dependability. The procedure used to estimate reliability and the associated standard error of measurement is dependent upon the frame of reference used to interpret the scores. In a norm-referenced score interpretation in which an individual’s score is interpreted in relation to the scores of others, reliability coefficients like Cronbach’s alpha are used. In a criterion-referenced score interpretation in which the individual’s score is interpreted in relation to a prespecified standard, dependability coefficients are used (24).

In the present case, the candidates’ scores on the OSCE form they took were interpreted in terms of the pre-specified standard for that form. Presented in Table II are the results of the estimation of dependability and the associated standard error for both Form A and Form B. These estimates were computed using a p x s (pharmacist x station) generalizability analysis, and are based on 47 candidates for Form A and 47 candidates for Form B.6,10

As shown, the general purpose indices of dependability (25) were 0.659 for Form A and 0.701 for Form B. The corresponding standard errors of measurement were, respectively, 0.056 (5.6 percent) and 0.060 (6.0 percent). These values are comparable to others found in health-related fields (26).

Examination of the variance components reveals that there was a fair degree of variability among stations. This finding is again similar to the findings for performance assessments in other fields (20,26,27).

Later in the project, for each individual, separate scores were calculated for each of the 12 competency areas assessed (Table I). These scores were not reported to the participants but clearly would be essential feedback if this assessment approach is adopted by the College.

References

7 Later in the project, for each individual, separate scores were calculated for each of the 12 competency areas assessed (Table I). These scores were not reported to the participants but clearly would be essential feedback if this assessment approach is adopted by the College.

8 Candidates selected the session they would attend. Consequently, while there were session differences, these differences were attributable to the differences in pharmacists who attended each session and not peculiar to the session itself. Consequently, the candidates from different sessions were combined into one sample for each form.

9 Station scores were considered in this analysis. An alternate design is to have checklist behaviors nested within stations crossed with candidates. However, the number of items per station was not equal and for some stations, the number was quite small. Consequently, the nested design was not analyzed.

10 The authors would like to acknowledge contributions related to the generalizability analysis provided by Dr. Arthur Rothman, University of Toronto, Toronto, Ontario.
Validity. Four types of validity evidence were collected for the pilot OSCEs. First, the validation panel pharmacists and the pharmacists who took part in the pretest agreed that: (i) the station problems were relevant to pharmacy practice; (ii) the behaviors in the checklist for each station were relevant to the problem to be solved; and (iii) as a set, the behaviors in the checklist included the behaviors necessary for an acceptable solution. Further, of the 77 (out of 96) pharmacists who completed an OSCE and who returned a follow-up acceptability questionnaire, more than 80 percent agreed that the OSCEs included realistic problems. However, despite attempts to make the stations realistic, a few of these pharmacists commented about the artificiality of the context of the assessment and the limitations placed on time and resources. Second, the checklist behaviors as reclassified by the content validation panels were mapped onto the competencies to be assessed as set forth in the assessment blueprint initially established for the total assessment program(1). It was hoped that the set of behaviors, aggregated across stations, would reflect the target percentage weights contained in the assessment blueprint. For example, as shown in Table I, competency one, “Demonstrate Knowledge of Prescription and Nonprescription Drug Therapy,” 10 percent of the total possible score should have been classified as relevant to competency one. However, neither Form A or Form B complied well with this or the other 11 target percents. The lack of agreement can be explained by the way in which the OSCE forms were developed. There was an attempt to ensure that the stations in each OSCE form were meaningful and representative. As mentioned earlier, each checklist behavior was initially included because of its relevance to the solution of the problem to which it was referenced. The checklists were to include a comprehensive listing of possible responses (positive and negative) to the problem presented. As a consequence, when aggregated across stations, the desired representation of competencies was not achieved. However, it is possible to statistically adjust the competency scores to achieve the desired “contribution” of each to the total score for the OSCE. 1

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>Mean square</th>
<th>Variance component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacists (p)</td>
<td>46</td>
<td>0.16436</td>
<td>0.00584 (8.8%)</td>
</tr>
<tr>
<td>Stations (s)</td>
<td>19</td>
<td>0.64488</td>
<td>0.01271 (19.2%)</td>
</tr>
<tr>
<td>Residual</td>
<td>874</td>
<td>0.04758</td>
<td>0.04758 (71.9%)</td>
</tr>
<tr>
<td>Total</td>
<td>906</td>
<td></td>
<td>0.06613</td>
</tr>
<tr>
<td>Form B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharmacists (p)</td>
<td>46</td>
<td>0.21910</td>
<td>0.00838 (10.5%)</td>
</tr>
<tr>
<td>Stations (s)</td>
<td>19</td>
<td>0.98781</td>
<td>0.01992 (24.9%)</td>
</tr>
<tr>
<td>Residual</td>
<td>874</td>
<td>0.05163</td>
<td>0.05613 (64.6%)</td>
</tr>
<tr>
<td>Total</td>
<td>989</td>
<td></td>
<td>0.07993</td>
</tr>
</tbody>
</table>

\[ \varphi_A = \frac{.00584}{.00584 + .00127} = .659 \]
\[ \varphi_B = \frac{.00838}{.00838 + .00199} = .701 \]

The third type of validity evidence can be seen from the results of the empirical comparison of the performance of the students with the performance of the practicing pharmacists. It was hypothesized that the pharmacists would score higher than the students due to their greater experience in practicing pharmacy. The mean and standard deviation of the total OSCE score for each group for each OSCE form is shown in Table III. The results from the corresponding two-way analysis of variance are summarized in Table IV.

<table>
<thead>
<tr>
<th>Group</th>
<th>Number</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>13</td>
<td>61.6</td>
<td>7.5</td>
</tr>
<tr>
<td>Pharmacists</td>
<td>48</td>
<td>66.5</td>
<td>9.4</td>
</tr>
<tr>
<td>Students</td>
<td>48</td>
<td>62.0</td>
<td>10.2</td>
</tr>
</tbody>
</table>

remains, though, the problem of too few items and the attendant problems associated with error of measurement when reporting the scores for some of the competencies. Attention will need to be given to these issues in the next revision of the OSCEs to ensure more adequate representation of the competencies by developing station problems which, as a set, contain the desired distribution of relevant behaviors both within and across stations.

The third type of validity evidence can be seen from the results of the empirical comparison of the performance of the students with the performance of the practicing pharmacists. It was hypothesized that the pharmacists would score higher than the students due to their greater experience in practicing pharmacy. The mean and standard deviation of the total OSCE score for each group for each OSCE form is shown in Table III. The results from the corresponding two-way analysis of variance are summarized in Table IV.

Working at the 0.05 level of significance, there was a significant difference between groups (across forms); there was no significant difference between forms (across groups) and the interaction between forms and group was not significant.

Finally, evidence of the validity of the use of the OSCEs can be seen in agreement between the categorization of pharmacists into passing and failing subgroups using results from both the knowledge assessment and the OSCEs(3).

11 For example, the competency scores could be found as follows: First,
\[ Z_{ic} = \frac{X_{ic} - X_c}{S_c} \]
where \( Z_{ic} \) is the standardized score for competency \( c \) for each individual within a station,
\[ X_{ic} \] is the observed score for individual \( i \) on the competency \( c \) items within that station,
\[ X_c \] is the mean score for competency \( c \) within that station, and
\( S_c \) is the corresponding standard deviation for competency \( c \).

These standard scores are then aggregated to form the total competency score for each individual across all stations, \( Y_{ic} \). Since \( Y_{ic} \) scores are not in standard score form, they are re-standardized. Denoting these re-standardized scores by \( Y_{ic} \), the scores to be reported to the individuals, \( T_{ic} \), are found from
\[ T_{ic} = \mu_{ic} + \sigma_{ic} Y_{ic} \]
where \( \mu_{ic} \) and \( \sigma_{ic} \) are, respectively, the desired mean and standard deviation for competency \( c \). Most often the same mean and standard deviation will be used for all competencies so as to increase the interpretation of the reported scores and differences among these scores.

If a total OSCE score across all competencies is needed, then the TS scores are first weighted by their respective percentages (Table I) and then aggregated to form a total OSCE score for each individual \( Y_{Tic} \). Again, the \( Y_{Tic} \) scores need to be re-standardized. Denoting these re-standardized scores by \( Y_{Tic} \), the total OSCE scores to be reported to the individuals, \( TOS_{ic} \), are found from
\[ TOS_{ic} = \mu_{ic} + \sigma_{ic} Y_{Tic} \]
where \( \mu_{ic} \) and \( \sigma_{ic} \) are, respectively, the desired mean and standard deviation for the OSCE scores. It is recommended that the OSCE mean and standard deviation be the same as the mean and standard deviation used for the competency scores so as to increase the interpretation of both the OSCE and competency scores.
One of the 52 questions on the acceptability survey asked respondents to rate the OSCE’s on 12 attributes such as adequate time, convenient location, and acceptability as a test of basic professional performance capability. Each attribute was rated using a six-point scale ranging from “Strongly Agree” to “Strongly Disagree.”

Table IV. Analysis of variance (percent total OSCE score) students-by-pharmacists

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>ms</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form (F)</td>
<td>1</td>
<td>331.47</td>
<td>3.62</td>
</tr>
<tr>
<td>Group (G)</td>
<td>1</td>
<td>339.39</td>
<td>4.36*</td>
</tr>
<tr>
<td>F x G</td>
<td>1</td>
<td>3.21</td>
<td>0.04</td>
</tr>
<tr>
<td>Residual</td>
<td>117</td>
<td>91.67</td>
<td></td>
</tr>
</tbody>
</table>

*p<0.05.

Practicality. Re-licensure assessments must be acceptable to those to be assessed and about whom a decision is to be made. At the same time, they must be feasible to administer. Of the 77 (out of 96) pharmacists who completed an OSCE and who returned the 52-item acceptability questionnaire, approximately 80 percent of the community pharmacists and approximately 70 percent of institutionally-based pharmacists agreed that the OSCEs were acceptable to basic professional performance. The community pharmacists in the pilot test sample strongly supported the relevance of the OSCEs in terms of their work settings but the pharmacists in hospital and long-term care settings were less sure. The latter groups suggested that the OSCEs may be more useful in identifying professional weaknesses rather than making actual relicensure decisions based on those identified weaknesses.

It appears that administration of OSCEs is feasible. First, at the individual pharmacist level, more than 80 percent of the pharmacists in the pilot test who returned the acceptability survey found the number of problems included in each OSCE acceptable. Thus it would appear that the time required of a pharmacist to complete an OSCE is reasonable.

To determine the total number of days that may be needed to complete an OSCE assessment of pharmacists under the competency assurance model being investigated, the results of the knowledge assessment(3) were extrapolated to the total provincial pharmacist population for 1992. As a result, it was projected that between 110 and 140 pharmacists would not meet the standard set for the knowledge test and, consequently, be required to complete an OSCE assessment. If, as was done in the pilot study, 36 pharmacists were assessed in one day, then three to four days would be required to assess the projected number.

Turning to costs, the 1992 cost per pharmacist administered an OSCE in the pilot study was approximately $250 (Can.). This figure represents assessment costs only and does not include the cost of development of the OSCEs. In addition, it does not include out-of-pocket expenses (e.g., travel, accommodation) incurred by the pharmacists assessed. If such costs were added, the total would be higher. For comparison, Vu et al. (23) reported a cost of $560 (U.S.) to develop and administer a performance assessment that included 18 stations involving an assessment time of 2.5 hours.

As a result, it was projected that between 110 and 140 pharmacists would be required to complete an OSCE assessment of pharmacists. The latter groups suggested that the OSCEs may be more useful in identifying professional weaknesses rather than making actual relicensure decisions based on those identified weaknesses.

CONCLUSIONS

As in other professions, pharmacy is searching for improved techniques to assess students and practitioners. This quest has gained a greater sense of urgency as undergraduate curricula place more emphasis on competency-based and problem-based instruction and assessment. In addition, there is increasing interest in periodic assessments of practitioners’ continuing competence. Traditional assessment strategies (multiple choice, short answer, and essay questions), although much maligned, can be useful in such assessments if carefully conceived, constructed, and applied. There is a need, however, for additional refinements in our assessment techniques, particularly in high-stakes assessments. As a result, performance assessments are receiving considerable attention in both general and professional education. One form of performance assessment, the OSCE, is currently being studied and applied in a number of the health professions. Recently, it has been suggested(28) that, due to the resources required to develop and implement OSCEs, consortia of pharmacy groups should be established to investigate their application to pharmacy education and assessment.

As part of this investigation of a model for the assessment of the continuing competence of pharmacy practitioners, two matched forms of an OSCE instrument were developed and pilot-tested. Results indicated that an assessment consisting of 20 stations involving an assessment time of 2.5 hours produces moderate total score dependability estimates.

In the competency assessment model under investigation, the results of the OSCE assessments are to be used to confirm sub-standard performance on the general screen of competence (a comprehensive knowledge assessment). Evidence supporting inferences drawn from the OSCE results is provided in this study by a series of expert panel reviews that revealed that the checklist behaviors were all relevant to the competencies assessed. Furthermore, an empirical comparison showed that pharmacists (the “experts”) outperformed final year pharmacy students (the “novices”). As well, there is substantial agreement between the knowledge screen test and the OSCE when identifying candidates who passed and failed(3). Finally, the pharmacists who took part in the development of the OSCEs and who were assessed by them were both supportive of their use as part of a comprehensive model that is fair and believable.

The accuracy of the SPs and SMDs used in this study were found to compare very favorably with those reported in other studies and is expected to improve with subsequent revisions in station content and procedures(29).

The OSCEs were found to be technically and logistically feasible, but, as “more authentic” assessments of competence, were more costly than traditional assessment procedures.

While seen as somewhat artificial, the OSCE model received considerable support from the pharmacists involved in this pilot test. Community pharmacists were more supportive than hospital pharmacists, reflecting the reality that, for the most part, the stations were more community-based.

This pilot test has provided evidence that the OSCE technique can be applied to high-stakes assessments in pharmacy. As has been suggested elsewhere (20,30) the dependability is generally quite good but the variability across stations is somewhat large indicating that performance is task specific. As a result, longer OSCE instruments comprised of more stations may be required to obtain more acceptable (i.e., higher) dependability coefficients and, consequently, lower standard errors of measurement. Alternatively, Newble and
Swanson(31) have suggested that “a reasonably reliable assessment of clinical competence could be attained in about 4 hours of testing time which was divided between an OSCE, constructed of practical and clinical stations, and a written test.” In the competency assessment model being investigated as part of this study, the authors concluded that an individual’s competence could be evaluated as a result of both a three hour assessment of practice knowledge and, in some cases, a 2.5 hour, 20 station OSCE.


References
(25) Ibid., p.108

APPENDIX A.

PARTI

Example of OSCE Station Development Worksheet

Author: ___________________ Editor: ___________________

Date: ___________________ Date: ___________________

I. Competency/Objective
1 — Demonstrate knowledge of prescription and nonprescription drug therapy
   a) pharmacological actions
   b) dosage — usual length of therapy

II. Station Requirements
Patient — adult
Patient medication profile (only prescription on record is the current prescription for 60 ranitidine 150 mg — i bid — dispensed 2 days previously)
Telephone

Compendium of Pharmaceuticals and Specialties
Goodman and Gilman

III. Scenario/Background Information
A patient with peptic ulcer disease, who has just started taking ranitidine, wants to know more about the drug, how it works and how long he should keep taking it.

IV. Standardized Patient Instructions
Phone the candidate at (_#_) and ask for the pharmacist.
Identify yourself as Norma Turnbull. Say, “I picked up this prescription for ranitidine 2 days ago, and I’ve been taking it. But I can’t remember what the pharmacist told me about the drug. What is this drug and what does it do? Also, how long do I have to keep taking it? If asked: the doctor says you have a stomach ulcer — you haven’t noticed any effects yet — you are taking antacid after each meal

V. Written Information
Prescription/s, profile/s, to be provided on appropriate form/s (to be attached if required).
PART 2
Example of OSCE Station Development Checklist

CANDIDATE #_____________________________

ASSESSOR _________________________________

Scoring Answers/Checklist

Candidate responds:

1. What are you being treated for? (What did your doctor say the prescription was for?)
2. The ulcer is a breakdown of the lining of the stomach, accompanied by excess acid in the stomach.
3. Ranitidine decreases excess stomach acid.
4. Some acid will still be produced, as it is needed for digestion.
5. Four weeks of ranitidine therapy is usually long enough to allow the ulcer to heal; it sometimes takes longer.
6. After 4 weeks your doctor will advise you whether or not you should continue therapy.

PART 3
Example of OSCE Station — Candidate Instructions

N. Turnbull will call you for some information about a prescription for ranitidine. Please respond as you would in practice.

You will have 5 minutes to complete this assignment. When you hear the signal please proceed to the next station.

Communication Effectiveness

Candidate:

7. Establishes rapport
8. Uses clear explanations free of medical jargon
9. Stresses important information
10. Is polite, tactful and receptive
11. Checks if patient understands information

Score/Results