Abstract

Introduction: Pharmacy technician education is typically at pre-degree level and comprises instruction in a range of scientific and clinical disciplines. The assessment of practical laboratory skills often utilizes attainment-referenced methods which are not always appropriate for vocationally-focused programmes.

Methods: We introduced an objective structured practical examination to assess student competency in three key areas (accurate weighing, calibration of a pH meter and performing a dilution). Students were scored using weighted criterion-based assessment criteria plus an overall global performance rating, which enabled determination of cut scores using a borderline regression method. Student opinions were elicited by online questionnaires using a five-point Likert scale.

Results: The move towards OSPEs did not significantly alter the distribution of student results from previous years (mean $\pm$ s.d, OSPE *vs*. legacy: 77 $\pm$ 19% *vs*. 73 $\pm$ 21%), which suggests that academic integrity has been maintained. There was a high level of consistency in Likert score responses (Cronbach's α = 0.823), with students clearly favouring the OSPE approach.

Conclusions: The move to OSPE-based assessment has proven to be successful and provides a basis for development of similar assessment strategies in the Pharmacy Technician programme.

Introduction

Within any pharmacy programme, chemistry is a key component of the early curriculum and with it comes a range of laboratory activities that develop psychomotor skills and lay the foundations for more advanced work in pharmaceutics, extemporaneous dispensing and aseptic manipulations (Sosabowski & Gard, 2008). Assessment of laboratory work within the chemistry curriculum at this level often involves a practical examination that includes a mixture of activities and recall-based interpretation of data (Gott & Duggan, 2002). This is not necessarily an authentic assessment method as it can be possible to achieve a passing grade without demonstrating practical competency (Gericke, Högström & Wallin, 2022). For example, in a question where the student is asked to describe and carryout out a dilution, they could achieve a passing grade by providing a good theoretical account of the procedure and performing an accurate calculation; the practical aspect could be ignored. A common alternative is to assess learners through written submissions such as laboratory reports; however, for large cohorts, it is impractical to mark written submissions consistently, even with rubrics and moderation between assessors. The impact of so-called 'essay mills', or more recently artificial intelligence chatbots such as ChatGPT, make it increasingly difficult to discern the authenticity of a student's written submission even with the assistance of anti-plagiarism software (Lee, 2023).

A further alternative is to use criterion-based assignments, in which students must satisfy specific assessment criteria rather than achieve a pass mark based on relative standards (Newton, 2011). This has clear advantages for professions where competency must be demonstrated for reasons of public safety (Pereira *et al*., 2018). However, one potential disadvantage is that learners must achieve all assessment outcomes in order to pass. For example, if a module consists of ten learning outcomes, and each one of these is assessed,

they must demonstrate understanding of all ten outcomes – the equivalent of scoring 100% in an examination. This assessment format is often open-book, so while students may have to exhaustively address all assessment outcomes, they can directly draw on information from textbooks, websites and class notes. Despite this, many struggle with the analytical and evaluative skills required, and so success in criterion-referenced assessment requires continuous formative feedback; approaches such as the cognitive apprenticeship model (Lyons *et al*., 2017) can be used to good effect here, maximizing the likelihood of achieving assessment outcomes at the first attempt.

In reviewing assessment approaches used in pharmacy education, a third option which can be viewed as a compromise between the two previous examples is the objective structured practical examination (OSPE) (Ahmed *et al*., 2011), a modification of the objective structured clinical examination (OSCE) pioneered by Harden *et al*. (1980). At its core, the OSPE assesses students by taking into account four factors:

(1) Assessment of the process and the product: rather than assess competency by viewing the final result, each step of the process is observed and assessed under controlled conditions.
(2) Breadth of skills assessed: the skills required for professional practice can be specifically assessed by an OSPE activity. This is not always possible through 'cook book' style practical work.
(3) Student's approach to the assessment: the approach to a task cannot be judged from typical assessment methods, yet this is an important transferable and transversal skill. This can be assessed in a OSPE through a global performance rating.
(4) Objectivity: although mark schemes, rubrics *etc*. ensure an acceptable level of objectivity during assessment, different assessors will apply the same rubric in slightly different ways.

In the OSPE, it is (usually) the same person who assesses all candidates for a particular assessment criteria.

From the student perspective, OSPEs have the advantage that they are specific to the curriculum delivered at an institution (*vs*. an externally set assessment) which allows thorough preparation through formative feedback. As OSPEs have multiple assessors, the potential for biased judgements is reduced and there is greater transparency in the assessment process. As multiple skills can be assessed at a single OSPE station, the process is efficient and reduces the overall examination time for candidates. From an institutional point-of-view, OSPEs provide valid and reliable assessment decisions that satisfy external stakeholders (Shirwaikar, 2015).

In designing an OSPE, an important consideration is the statistical framework in which the candidates' performance will be viewed. For large cohorts, the borderline group method (Boursicot, Roberts & Pell, 2007) has been successfully used as an alternative to the modified Angoff scheme. However, in smaller cohorts, the likelihood of having a sufficient number of borderline candidates to confidently set the standard is low, and so the borderline regression model is preferred (Schoonheim-Klein *et al*., 2009). In this approach, each OSPE station must have appropriately weighted item marks and an overall global performance mark (*e.g*. fail, borderline, clear pass, good and outstanding) (Wood, Humphrey-Murto & Norman, 2006).

Educational Setting & Activity

The Diploma for Pharmacy Technicians is a two-year, competency-based program that is mapped to the UK General Pharmaceutical Council's Initial Education & Training Standard for Pharmacy Technicians (Boughen & Fenn, 2020). It is at assessed at Level 3 of the Regulated Qualifications Framework which represents pre-degree-level qualifications. In our institution, we follow an integrated spiral curriculum in which the basic sciences are delivered and

assessed in semester one of the first year, after which students move through a systems-orientated approach to clinical pharmacy (Mawdsley & Willis, 2018).

Laboratory skills may be viewed as less important for contemporary pharmacy technician training – there is now very little extemporaneous dispensing in UK pharmacies, except in some hospital pharmacies and dedicated businesses (termed 'Specials manufacturers'). Therefore, UK Pharmacy Technicians are unlikely to use practical, laboratory-like skills in their day-to-day practice. However, as a proportion of our graduates seek employment in non-clinical roles (*e.g.* in the pharmaceutical industry), we continue to embed a set of core laboratory skills that are transferable to the compounding environment, or as preparation for more advanced roles (Burnett, Dooley & Wall, 2003). These were historically delivered as a series of standard practical activities with students working in pairs but submitting individual laboratory reports.

The development of the OSPE began by referring to the module assessment outcomes and extracting the core competencies that we felt could be assessed *via* OSPE. We selected three core skills – *viz*. accurate weighing (OSPE 1), preparation of a dilution (OSPE 2) and calibration of a pH meter (OSPE 3) – to be assessed in this format. These skills were selected as they are required for the pharmaceutics module in the second year (where extemporaneous dispensing skills are taught and assessed). The grading and characteristics of each global proficiency for each OSPE station was agreed by a quorum of six subject experts (two chemists, two practicing pharmacists and one academic pharmacist). To trial our approach, we used a group of students from the first year of a sister program in Applied Science who had a similar background and level of experience. One group undertook the legacy assessment (standard practical activities in pairs but individual laboratory reports (referred to as 'paired laboratory work'), a second group undertook the same standard practical activities but did so individually,

submitting laboratory reports (referred to as 'individual laboratory work'), and a third group undertook the proposed OSPEs. This approach allowed us to refine the logistics of the process but to also to account for confounding variables in our analysis (see later). Based on these findings, we proceeded with the move to OSPEs for the next academic year.

For the roll-out of the OSPEs, our first year pharmacy technician students were briefed about the assessment process and they undertook three sessions of formative laboratory work through which they gained experience of the skills that would be summatively examined through the OSPEs. For the examination period, the teaching laboratory was set-up with three OSPE stations, and at each station there was an examiner who observed the candidate's performance and recorded their achievement on a score sheet. Candidates were provided with a set of instructions alongside any materials required. Prior to the examination, candidates were randomly divided into groups of three, with each student assigned to start at a particular OSPE station. When the time elapsed (ten minutes per station), a buzzer sounded that signalled the end of that OSPE and the candidates rotated to the next station.

Data Collection & Analysis

For the pilot, Group 1 (legacy: paired laboratory work, individual reports; n = 12) and Group 2 (legacy: individual laboratory work, individual reports; n = 12) were graded using an established rubric that awarded an overall percentage mark. Group 3 (OSPE; n = 12) were given an overall percentage mark based on their performance at all three stations. As the sample size was small, we selected Quade's analysis of covariance to assess differences in individual performance taking into account paired performance as a covariant.

In the main study, we used historical data from students who had undertaken the legacy assessment (paired laboratory work, individual reports; n = 63) and compared this to the OSPE cohort (n = 65). To satisfy the assumptions of the independent samples t-test, data were first

transformed using the inverse hyperbolic sine function. Finally, we evaluated learner engagement through a short online questionnaire that reported results on a five-point Likert scale. The internal constancy of the Likert responses was assessed by Cronbach's alpha and overall consensus scores for each question were calculated using a modification of Shannon entropy (Tastle & Wierman, 2007) which converts ordinal data into a dimensionless measure of dispersion (Eqn. 5 in reference). All statistical analyses were undertaken using IBM SPSS Statistics for Windows, version 28.0.0.1.1 (IBM Corp., Armonk, N.Y., USA). Violin plots were constructed using GraphPad Prism for Windows, version 9.5.1 (GraphPad Software, San Diego, California USA). Written informed consent was obtained from all participants in line with our institutional ethics policy.

Evaluation

In the pilot study, we sought to compare the legacy assessment method with the OSPE, taking into account the impact of working in pairs *vs*. working alone. Analysis by Quade's ANCOVA revealed no significant difference between the groups (two-tailed *p*-value 0.504; $F = 0.470$), which suggests that results obtained from the OSPE are in line with those obtained by the legacy assessment. Moreover, the move away from working in pairs to working individually does not appear to have any confounding impact. The performance of candidates in each of the OPSE stations is summarised in Table I.

For the main study, candidate performance at each of the OSPE stations was largely homogeneous with the distribution of scores following similar distributions (Figure 1). The median score for the second OSPE was lower than the two other stations; this is likely due to the increased complexity of this station (calculations plus practical manipulatives) *vs*. the two other stations (only practical manipulatives). However, the use of a borderline regression

model to establish the cut score means that the increased difficulty of this station is accounted for in assigning pass/fail to candidates.

Comparison of legacy student performance *vs*. OSPE revealed no significant difference between the two cohorts (two-tailed *p*-value 0.373; t = 0.896). Of course, with the legacy cohort, practical work was undertaken in pairs (although all students submitted individual reports), whereas the OSPE cohort undertook practical work individually. While we cannot exclude a confounding impact from the switch to individual practical work, we demonstrated in the pilot study that this had no significant effect, albeit at a lower statistical power. From this, we can infer that OSPEs yield a similar spread of results to traditional assessment methods and does not significantly bias the pass rate (mean $\pm$ s.d, OSPE *vs*. legacy: 77 $\pm$ 19% *vs*. 73 $\pm$ 21%). To highlight the impact of any outliers on this inference, we constructed a Bland-Altman plot (Figure 2) in which the difference in scores for the two assessment methods is on the *y*-axis and the average of the assessment scores on the *x*-axis. This identified only a single borderline outlier, with all other student scores clustering about the mean and well-within the limits of agreement ($\pm$ 1.96σ). This would seem to confirm that the OSPEs are as valid an assessment instrument as the legacy approach.

The results of the student evaluations showed strong internal consistency (Cronbach's α = 0.823), indicating a good level of reliability in these evaluations (Tavakol & Dennick, 2017). Results (Table II) show strong consensus in the positive responses for Q1–3 & Q5, which is encouraging. For Q4 – that they felt the Examiners' judgements were fair – there was a lower consensus score. We consider two explanations for this. Firstly, one of the examiners was a member of faculty unknown to the students; this could potentially introduce an element of uncertainty/mistrust. In support of this, two quotes from students were

"*what if they didn't see how much I weighed out*"

"*I don't think they could see the level [of liquid] from where they were standing*"

These are valid concerns and in the most recent OSPEs, we now only use faculty that are known to the students and ensure that they over-emphasize the observation process by using statements like "can I see the reading on the balance please". A second explanation for the lower consensus score in Q4 is that students may not have fully understood the grading process. In the first iteration of the OSPEs, students were briefed about the grading process, but they had not seen the candidate record sheets (essentially the rubrics) against which they were being assessed. Some faculty view a rubric in the same light as a mark scheme for an exam and are resistant to sharing this with students before the assessment. However, as discussed by Jackson & Larkin (2002), allowing students to view rubrics helps them progress towards a goal, and can therefore be considered as part of the formative assessment process. This being the case, we now make partially redacted candidate record sheets available to students prior to the OSPEs.

Conclusion

We have shown that OSPEs are a means of securing authentic, criterion-referenced assessment decisions and can be an efficient replacement for traditional assessment methods. For this particular group of students, the early introduction of OSPEs has the added benefit of preparing them for OSCEs later in the program. This approach is also consistent with the upper levels of Miller's triangle, were pharmacy professionals are required to demonstrate 'shows how' and 'does' characteristics (McFadyen & Diack, 2017). The move to an OPSE approach was supported by the majority of the faculty, although some were cautious of the change, particularly with respect to the time commitment involved. However, when we compared the effort-hours for marking laboratory reports *vs*. the time required for OSPEs, there was clear support for the latter. Other criticisms, such as stress to the student, role of the examiner and

fragmentation of the assessment have been addressed elsewhere (Harden, 2015). Of course, any change to the curriculum must have buy-in from all stakeholders – staff, students and external groups such as employers. To address the latter, we issue certificates of competency and guide our students through development of a skills-based CV (résumé). This provides employers with a more precise understanding of the skill set a potential employee possesses, rather than just the title of a qualification.

REFERENCES

[1] Sosabowski, M.H. & Gard, P.R. (2008) Pharmacy Education in the United Kingdom. Am J Pharm Educ., 72(6):130. https://doi.org/10.5688%2Faj7206130

[2] Gott, R. & Duggan, S. (2002) Problems with the Assessment of Performance in Practical Science: Which Way Now? Cambridge Journal of Education, 32(2), 183-201. https://doi.org/10.1080/03057640220147540

[3] Gericke, N.; Högström, P. & Wallin, J. (2022) A Systematic Review of Research on Laboratory Work in Secondary School. Studies in Science Education, 25,1-41. https://doi.org/10.1080/03057267.2022.2090125

[4] Lee, H. (2023). The Rise of ChatGPT: Exploring its Potential in Medical Education. *Anatomical Sciences Education*, 00, 1– 6. https://doi.org/10.1002/ase.2270

[5] Newton, P.E. (2011) A level Pass Rates and the Enduring Myth of Norm-referencing. Research Matters, 2, 20-26. Available at: https://www.cambridgeassessment.org.uk /Images/567580-a-level-pass-rates-and-the-enduring-myth-of-norm-referencing.pdf

[6] Pereira, A.G.; Woods, M.; Olson, A.P.; Van Den Hoogenhof, S.; Duffy, B.L. & Englander, R. (2018) Criterion-based Assessment in a Norm-based World: How Can We Move Past Grades? Academic Medicine 93 (4), 560-564. https://doi.org/10.1097/ACM.0000000 000001939

[7] Lyons, K.; McLaughlin, J. E.; Khanova, J. & Roth, M. T. (2017) Cognitive Apprenticeship in Health Sciences Education: A Qualitative Review. Adv. In Health Sci. Educ. 22, 713-739. https://doi.org/10.1007/s10459-016-9707-4

[8] Ahmed, K.; Miskovic, D.; Darzi, A.; Athanasiou, T. & Hanna, G.B. (2011) Observational Tools for Assessment of Procedural Skills: A Systematic Review. The American Journal of Surgery, 202 (4), 469-480. https://doi.org/10.1016/j.amjsurg.2010.10.020

[9] Harden, R.M. & Cairncross, R.G. (1980) Assessment of Practical Skills: The Objective Structured Practical Examination (OSPE). Studies in Higher Education 5 (2), 187-196. https://doi.org/10.1080/03075078012331377216

[10] Shirwaikar, A. (2015) Objective structure clinical examination (OSCE) in pharmacy education – a trend. Pharmacy Practice, 13 (4), 627. https://doi.org/10.18549%2 FPharmPract.2015.04.627

[11] Boursicot, K.A.M., Roberts, T.E. & Pell, G. (2007) Using Borderline Methods to Compare Passing Standards for OSCEs at Graduation Across Three Medical Schools. Medical Education, 41: 1024-1031. https://doi.org/10.1111/j.1365-2923.2007.02857.x

[12] Schoonheim-Klein, M., Muijtjens, A., Habets, L., Manogue, M., Van Der Vleuten, C. & Van Der Velden, U. (2009) Who Will Pass the Dental OSCE? Comparison of the Angoff and the Borderline Regression Standard Setting Methods. European Journal of Dental Education, 13: 162-171. https://doi.org/10.1111/j.1600-0579.2008.00568.x

[13] Wood, T.J., Humphrey-Murto, S.M. & Norman, G.R. (2006) Standard Setting in a Small Scale OSCE: A Comparison of the Modified Borderline-Group Method and the Borderline Regression Method. *Adv Health Sci. Educ. Theory Pract.* 11, 115–122. https://doi.org /10.1007/s10459-005-7853-1

[14] Boughen, M. & Fenn, T. (2020) Practice, Skill Mix, and Education: The Evolving Role of Pharmacy Technicians in Great Britain. Pharmacy 8 (2), 50 https://doi.org/10.3390/ pharmacy8020050

[15] Mawdsley, A. & Willis, S. (2018) Exploring an Integrated Curriculum in Pharmacy: Educators' Perspectives. Currents in Pharmacy Teaching & Learning 10 (3) 373-381 https://doi.org/10.1016/j.cptl.2017.11.017
[16] Burnett, D.; Dooley, M.J. & Wall, D. (2003), Cell-Based Therapies—A Role for Pharmacy Technicians. *Journal of Pharmacy Practice and Research*, 33, 296-298 doi: 10.1002/jppr2003334296.
[17] Tastle, W.J. & Wierman, M.J. (2007) Consensus and Dissention: A Measure of Ordinal Dispersion. International Journal of Approximate Reasoning, 45 (3), 531-545. https://doi.org/10.1016/j.ijar.2006.06.024
[18] Tavakol, M. & Dennick, R. (2011) Making Sense of Cronbach's Alpha. International Journal of Medical Education, 2, 53. https://doi.org/10.5116%2Fijme.4dfb.8dfd
[19] Jackson, C.W. & Larkin, M.J. (2002) Teaching Students to Use Grading Rubrics. Teaching Exceptional Children, 35 (1), 40-45. https://doi.org/10.1177/004005990203500106
[20] McFadyen, M.C.E. & Diack L. (2017) I Can Step outside My Comfort Zone. *Pharmacy*, 5(4):59. https://doi.org/10.3390/pharmacy5040059
[21] Harden, R.M. (2015) Misconceptions and the OSCE. Medical teacher 37 (7), 608-610. https://doi.org/10.3109/0142159X.2015.1042443

Table I Summary of OSPE Performance

	Pilot Study (*n* = 12)			Main Study (*n* = 65)		
	OSPE 1	**OSPE 2**	**OSPE 3**	**OSPE 1**	**OSPE 2**	**OSPE 3**
Median (95% CI) (%)	61 ± 2.8	51 ± 3.5	69 ± 4.1	65 ± 4.2	55 ± 3.9	65 ± 4.3
Cut Score (%)	44	41	45	43	44	46
Pass Rate (%)	90	78	88	88	74	77

Table II Summary of Student Questionnaire Responses.

Question	Relative Frequency*					C-Score
	SA	A	N	D	SD	
The OSPEs were well organized	0.43	0.53	0.03	0.03	0.00	0.78
I had enough time for each station	0.35	0.48	0.13	0.03	0.03	0.72
I felt prepared for the OSPEs	0.13	0.70	0.13	0.03	0.03	0.80
I feel that the judgements were fair	0.28	0.38	0.25	0.08	0.03	0.64
I prefer OSPEs to laboratory reports	0.75	0.18	0.08	0.00	0.00	0.80

*SA, strongly agree; A, agree; N, neutral; D, disagree; SD, strongly disagree

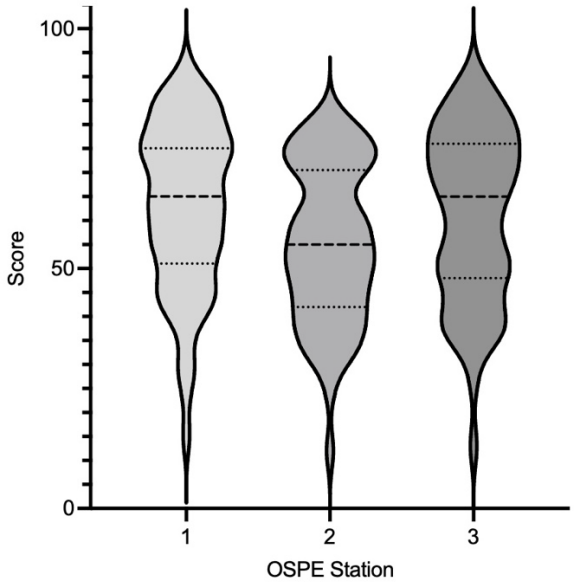

Figure 1 Summary of Scores for OSPE Stations. Dashed line (---) represents the median score; dotted line (···) represents the 95% confidence intervals.

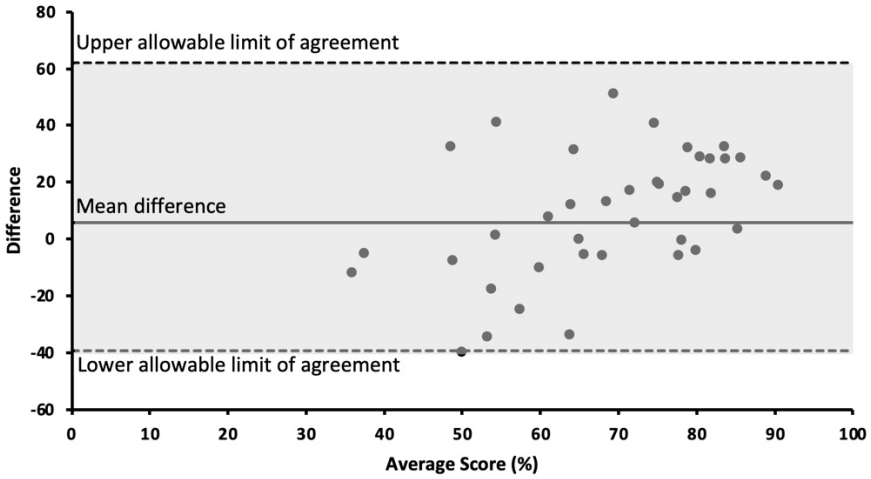

Figure 2 Bland-Altman Plot of Legacy *vs*. OSPE Performance. The mean difference between assessment methods (5.8%) is represented by the solid line. Dashed lines represent the upper and lower boundaries of agreement between the two assessment approaches.