Developing a validity argument for a simulation-based model of entrustment in dispensing skills (MEDS) assessment framework

Abstract

Background: Integrated assessment of multiple competencies at once, including Entrustable Professional Activity (EPA)-based assessment, is emerging as an effective approach to competency-based evaluation of health professionals, however, there is an absence of validated assessment frameworks in entry level pharmacy education. Accurate, timely and meaningful assessment of pharmacists in medication supply tasks underpins good professional regulation of pharmacists, and the safety of the public.

Objective: We aimed to develop an assessment framework and establish a validity argument, containing multiple sources of evidence, for use in the integrated assessment of pharmacy student's competency in all aspects of the supply of prescribed medicine(s).

Methods: A two-phase prospective study was conducted. Phase 1 involved the development and content validation of the Model of Entrustment in Dispensing Skills (MEDS) assessment framework using a literature review, a think-aloud study and expert consultation. In phase 2, a pilot study was conducted with faculty and expert assessors (n=10) to test the framework with a sample of Year 4 undergraduate pharmacy student simulations (n=21) to provide an assessment of their performance in an integrated medicine dispensing task. Subsequent analysis involved psychometric evaluation of rating scales and usability testing.

Results: Validity evidence was collected and organised across the two study phases. The MEDS framework has good evidence of content validity supported by the rigorous development and consultation process, as well as case sampling, with 88% of national practice-based competencies represented across the two simulations. Reliability coefficients were high and acceptable, supporting strong agreement across domains, students, and simulations; as well as a strong correlation between the EPA and total score (spearman correlation rho 0.725, *p*-value <.001);

Conclusion: This study describes a valid and rigorous approach for the implementation and interpretation of an integrated simulation-based assessment tool for determining pharmacy student's progress towards entrustment for independent medication supply practice.

Keywords: Simulation; assessment; validation; pharmacy; entrustable professional activities; competency.

Conflict of interest and financial disclosure

The authors whose names are listed above certify that they have NO conflict of interest with, affiliations with, or involvement in any organisation or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript

Introduction

The registration of individual pharmacists is centred on their ability to demonstrate their competence to practice. Competencies are the defined knowledge, skills, behaviours and attitudes that an individual develops through education and experience, while competence, captures a complete repertoire of competencies.[1] Accurate, timely and meaningful assessments are essential during health professional training and in clinical practice, to provide information about conformance with professional practice standards and the readiness of a health worker to provide effective patient care.[2, 3] Literature supports the evaluation of medicine dispensing tasks using uniform, structured assessments which break down elements of the task into smaller units that are evaluated separately.[3,4] however, a lack of integrated assessments with validated measurement scales may hinder the development of holistic skills in medicine dispensing. Integrated assessment requiring proficiency in multiple competencies simultaneously has emerged as an important approach and is being increasingly used across health professional education to evaluate competence and measure performance. Furthermore, a growing body of evidence supports implementation of Entrustable Professional Activities (EPAs) in health professional education, including assessment that results in summative decisions about the entrustment of individuals to act under specified levels of supervision.[5] However, despite the movement towards integrated and authentic performance evaluations (focus on the application of knowledge in real life settings),[6] the literature is riddled with shortcomings in validation efforts to support such assessments.[7,8]

The development of defensible assessments, that enable sound justification about decisions made regarding a trainee, can be challenging.[9] Validation is the process of collecting and interpreting evidence to support the decisions made during an assessment. Validation is also used to understand the strengths and limitations of an assessment tool.[2] The process of validation has evolved significantly and our understanding of validity theory, as it relates to assessment in health professional education, has become increasingly complex, with a range of frameworks offering unique yet overlapping views.[10,11,12] The CAPE (Center for the Advancement of Pharmacy Education) Educational Outcomes (2013)[13] and the National Competency Standards Framework for Pharmacists (2016),[14] including standards for advanced practice, are but two examples of the increased emphasis on standards of performance for pharmacists during all stages of training and practice including advancement in degree programs, remediation, registration and lifelong learning. As we move towards increased reliance on

performance-based assessments for decisions about health professionals' readiness and ability to perform in the workplace, there is a need for validated assessment frameworks which align with these outcomes and standards, to defend the processes used and decisions made. This is particularly important for high stakes assessments which are used to make important decisions about an individual's readiness to progress to the next stage of practice.[2,15-17]

Traditionally psychometric models have been used to measure validity.[18] Content validity, criterion validity and construct validity are the key parameters that have historically been used to judge the quality of assessments. However, the isolated application of such approaches to the validation of current trends in complex assessment models has been increasingly questioned because it may give an incomplete evaluation of the quality of the assessment overall.[11] More recently, alternative validity frameworks have been proposed by Messick[19] and Kane,[20] and adopted in the field of health professional education. These frameworks focus on collecting evidence from a number of different sources, including psychometric models, to compile a validity argument to support the proposed score interpretations[11]. The type of evidence collected depends on the assessment instrument in question and its intended application.

Validation science would be improved by greater evidence informing the consequences and implications of assessment. Kane's model[20] proposes that assessors collect evidence across four inferences: scoring, generalisation, extrapolation and implications.[21] Scoring includes the scoring procedures and rubric and refers to the process of moving from an observed performance to an observed score. Generalisation refers to evidence for how well the observed sample represents the range of all possible performances. Implications refers to the process of moving from scores to decisions about individuals, such as readiness to progress to next stage of practice. Finally, extrapolation refers to how well students are likely to perform in future contexts, for example, does this assessment score predict performance in clinical practice?[11] The inferences in Kane's validity framework have been recently used to examine validity evidence for a simulation-based assessment of clinical competence[21].

Integrated assessments are central to competency-based education models that dominate curricula of health care professional programs.[12] According to Koster et al (2017), assessment formats in pharmacy curricula should move from simple isolated assessments to more integrated, complex assessment formats.[2] In this study, we aimed to establish a validity argument for a simulation-based assessment

framework known as the Model of Entrustment in Medicine Dispensing (MEDS), for its use in the integrated assessment of competency in medication supply. The framework is designed to assess students' ability to perform and integrate all skills necessary to dispense a medication for a patient that is both safe and appropriate in the circumstances. This study includes fourth year undergraduate students in Australia, which would most closely be associated with the end of third year of the PharmD curriculum, prior to students entering the experiential learning rotations. To achieve our aim, we sought to collect and interpret evidence from multiple sources including qualitative research data; blueprinting of our assessment, and preliminary psychometric data.

Methods

Phase 1: Development and content validity of the framework

For the purpose of this study, medicine dispensing tasks include the review of a prescription and the preparation, packaging, labelling, record keeping and supply of the prescribed medicine with adequate counselling to a patient, or another person who is responsible for the administration of the medicine to that patient. A comprehensive literature review (n=20) was conducted to identify the key components of medicine dispensing and factors that contribute to medication errors during this process.[22] A scoping review was also conducted to explore competency-based assessment of pharmacists (n=36),[23] and the use of integrated simulation-based assessments in health professional assessment literature, and their applicability to pharmacy education (n=20).[24] This process also incorporated a review of various rating systems including checklists and global rating scales.

A qualitative think-aloud study examined the patterns and processes of pharmacists' clinical reasoning when making decisions about the safety and appropriateness of prescribed medicines.[25] Medication supply behaviours and reasoning processes were extracted from the data and categorised thematically to reflect the steps of the dispensing process. Themes then informed the key components of the first version of the MEDS assessment framework, consisting of six domains and 26 items. Three members of the research team (senior academic in health behaviour and medical communication; pharmacist and professor in pharmacy practice; medical educator and psychometrician, each academics with more than ten years' experience in health professional assessment) reviewed the first version and consensus was reached on the most appropriate and practical rating system, a 4-point Likert-type scale based on previous evidence for use.[26]

The first version of the MEDS assessment framework was presented consecutively to nine individuals representing a variety of disciplines including pharmacy, nursing and medicine and with a range of expertise in health professional training and assessment, by satisfying at least one of the following criteria: a) experience in health professional assessment and conversant with current theories and trends in competency-based assessment; or b) experience in medicine dispensing, including supervision of trainees in medicine dispensing activities. These experts reviewed the framework and participated in a semi-structured interview to provide feedback on the structure, item content, and rating system. Issues of clarity and suggestions for improving the framework were documented, and the first version of the MEDS assessment framework was refined iteratively in response to feedback during regular consultations with the research team over the consultation period. Phase 1 was conducted over a two-year period, including 18 months of literature review and think-aloud study, followed by six months of consultation. The first version of the framework developed in phase 1 was subsequently piloted in phase 2.

Phase 2: Pilot - Validation study

Ethics

Ethics approval was obtained from the University Newcastle Human Research Ethics Committee (HREC) [Approval No H-2018-0083], for the publication of assessment data used in this study. Informed consent was collected from all participants prior to their participation in the study.

Study setting

The University of Newcastle Transition to Practice course (subject) includes a medicine dispensing assessment for 4th Year undergraduate Bachelor of Pharmacy (Honours) students. This simulation-based assessment requires integration of all multidimensional aspects of medicine dispensing (medication history taking, prescription review, drug selection and labelling, drug information consultation and patient counselling) in a single standardised patient interaction. Each medicine dispensing task involves a standardised patient interaction consistent with community pharmacy practice, from initial patient contact, to safe and appropriate supply of medication. Each pharmacy student was required to complete two different medication supply scenarios, each involving an experienced standardised patient, trained to function consistently and efficiently. The simulations take place in a demonstration pharmacy equipped with a range of actual and placebo medicinal products,

computer dispensing software and equipment such as barcode scanners. For the purpose of this study students are provided with a mock prescription which replicates all the authentic features of an actual prescription. Pharmacy students are expected to lead the patient interaction, elicit a complete medication history, review and interpret the prescription in the context of relevant patient history and drug information resources, decide on the safe and appropriate supply of medication, and provide adequate patient counselling. Following the simulated interaction, students participate in a post-simulation discussion with the assessor. During this time, students are asked to respond to structured questions from the assessor about the dispensing task, in order to provide verbal defence of decisions made during the medicine dispensing process. Students have approximately 12 minutes to complete both the simulation task and post-simulation discussion and have 10 minutes between the completion of one scenario and the beginning of the next. For the purpose of the validation study, these sessions were conducted as formative assessments, and students were provided with verbal feedback immediately after each simulation.

Case/ content development

The cases were based on actual patient encounters but modified to meet the relevant competencies of final year pharmacy students. The first simulation involved the supply of a new oral anticoagulant to a 42-year-old female patient with a diagnosis of venous thromboembolism. The second focused on the supply of erythromycin suspension to a 6-year-old male with suspected whooping cough (pertussis) which intentionally had a dispensing error that students were required to manage. Case content was blueprinted by strategically planning the assessment content against learning objectives, using the first version of the assessment framework and the resources which informed its development. The scenarios were reviewed by the research team until consensus was reached on the detail. The assessments occurred over four days in the second half of the 2018 academic year.

Student participants, assessors and assessor training

All students enrolled in Transition to Practice were invited to participate in the research study. Twenty-eight students consented to being included in the research, and twenty-one students were included in the final data set as they had completed two simulations and had recordings of sufficient quality to be assessed by the external assessors. This sample size was adequate according to an a priori sample size calculation. Each simulation was video-recorded and coded with the participants' identification numbers to maintain anonymity.

Ten assessors, including three faculty members and seven purposely recruited external assessors, provided assessment of selected students using the MEDS assessment framework during the assessment sessions. External assessors, who used video-recorded simulations, were recruited from across Australia and represented pharmacists with experience in evaluation of pharmacy student and/or pharmacist performance including at least one of the following: a) intern training provision, b) pharmacy student or intern pharmacist preceptor in the workplace, c) pharmacy board examiner, or d) undergraduate or postgraduate pharmacy teaching and assessment.

Each assessor received structured one-to-one briefing in the design and use of the MEDS assessment framework as well as an overview of the research study. The briefing included the expected pharmacy care associated with each of the scenarios. Using a nested study design, assessors and students were randomly assigned to three groups and assessors independently assessed a group of seven students, each performing two simulations. Each student video was assessed by at least three assessors. Assessors were instructed to provide their assessment individually without collaborating or sharing views on student performance. All assessors used the same paper-based MEDS assessment tool and recorded their scores in a pre-formatted Excel spreadsheet.

Usability testing

All assessors who provided student evaluations were asked to evaluate the usability of the simulation-based MEDS assessment framework. A 6-item questionnaire based on a four-point Likert scale (strongly agree/ agree/ disagree/ strongly disagree) was developed from questionnaires employed in previous studies[27]. The questionnaire was completed by assessors immediately following their assessment experience to evaluate the usability of the MEDS framework; assessors were also given the opportunity to provide feedback via open ended survey response or verbally via follow-up phone call.

Statistical analysis

Measures of validity and reliability were obtained by analysing student assessment data using SAS software version 9.4 (SAS Institute Inc. Cary, NC, USA).

Due to clustering of the assessment results in the data, for both individual students and assessors, inter-rater agreement/ correlation tests were interpreted via the Intraclass Correlation Coefficient (ICC).[28] Inter-rater agreement was calculated for EPA rating, total score and for each question (domain). Inter-rater agreement was examined using mixed regression modelling.[28] Ordinal mixed modelling was used for the EPA score and five out of the six domains (logistic modelling was performed for domain 4 as this question only had two response options). Fixed effects for simulation number (1 or 2), period (first simulation performed, second simulation performed – defined here as the instance) and simulation order (subjects who performed simulation 1 then simulation 2 (1-2) or vice versa (2-1)), and random intercepts for assessor, and student were modelled.

The same regression models that were used to calculate ICC were also used to examine simulation effect. Estimates from linear mixed models are presented as the difference in scores between simulation 2 compared to 1, between second instance compared to first instance and between 2-1 compared to 1-2. Estimates from ordinal mixed models are presented as odds ratios: i.e. the odds of having a higher score from simulation 2 compared with simulation 1, odds of having a higher score for second instance compared to first instance, and odds of having a higher score for subjects who performed in the order 2-1 compared to 1-2.

Spearman correlation was used to examine the association between EPA ratings and total score; we have included all records, ignoring potential correlation due to multiple ratings and simulations. The correlation coefficient (rho) was interpreted as: 0-0.19 very weak; 0.2-0.39 weak; 0.4-0.59 moderate; 0.6-0.79 strong; and 0.8-1 very strong.

Results

Phase 1 – Development and content validity of the assessment framework

Content validity

The first version of the MEDS assessment framework was developed iteratively through expert consultation. This initial version of the framework consisted of a six-dimension global rating system with an entrustment scale, developed specifically for the purpose of the assessment of pharmacists in medicine dispensing tasks, outlined in Table 1. The quality of performance points and entrustment scale are defined based on patient safety factors and ability to progress to the next stage of practice. Results

of the phase 2 pilot study (outlined below) led to the separation of domain five into two separate components, which led to a total of seven domains in the final version of the MEDS assessment framework (Appendix 1). This change was supported by feedback from the expert assessors and stakeholders from the original consultation process.

Table 1: Summary of piloted version of MEDS assessment framework

Phase 2 – Validation of the assessment framework

A pool of 21 pharmacy students was included in the final data set and the performance of each student in their two simulated interactions were assessed by at least three different assessors, generating 178 assessment data points for use in our study.

Inter-rater reliability (ICC)

As shown in Table 2, there was a very strong agreement among assessors for EPA ratings made across students within simulation 1 (ICC 0.838), and a strong agreement for simulation 2 (ICC 0.784). At the individual domain level, inter-assessor agreement across both simulations was strong for domain 3 (ICC 0.684) and moderate for domains 1, 2, 5 and 6 (ICC 0.446-0.555). When examined by individual simulation, most ICCs indicated strong correlations.

Table 2: Inter-rater agreement (ICC) for the EPA and domain section ratings (1-6)

Correlation between EPA score and total score

The Spearman correlation showed, for all records, there was a strong positive correlation (rho 0.725, *p*-value <0.001) between the level of entrustment assigned on the EPA scale and the overall numerical score in the simulation. These results are shown in Table 3 and Figure 1.

Table 3: Correlation between EPA rating and total score

Figure 1: Correlation between EPA rating and total score

Simulation effect on EPA rating, total scores and domain scores

Effects for simulation number (1 or 2), period (first instance, second instance) and simulation order (1-2 or 2-1) were examined. No significant difference was observed in overall EPA ratings between simulations, although the simulation order was found to be associated with significantly different scores. Students who performed simulation 1 followed by simulation 2 had significantly higher odds of a higher EPA score in simulation 1.

For total scores, simulation 2 had a significantly lower average total score when compared with simulation 1. No significant differences were seen in average scores for Domains 1-5 between simulations, however simulation 2 had a significantly lower odds of having a high domain 6 score when compared to simulation 1. Similarly, the odds of having a high domain 6 score were significantly higher the second time the simulation was performed for either simulation, regardless of simulation order. A detailed description of how students were rated at each EPA level is described elsewhere.[29]

Case sampling

Our sampling strategy involved creating an assessment blueprint using guiding practice documents as described above. Across the six assessment domains, 65% (17/26) of all competency standards were included.[14] Furthermore, of 70 relevant (i.e. practice-based) Australian competencies, 88% (n=62) were represented across the two simulations, representing the essential knowledge, skills and attributes that underpin effective and safe medicine dispensing. Two out of three pre-defined age groups (paediatric, adult) were represented in the simulations, as were two broadly classified disease types deemed significant public health priorities, including communicable disease[30] and cardiovascular health.[31]

Usability

As shown in Figure 2, six out of seven assessors (86%) who evaluated the usability of the MEDS assessment framework perceived the tool to be flexible in rating student performance using their own

professional judgement; that it was easy to use; flexible for providing student feedback and effectively evaluated performance in medicine dispensing tasks.

The qualitative comments were focused on the face to face questioning, and while some assessors agreed that it adds value for making decisions ("a good means to assess understanding of the dispensing process, something which isn't readily assessed") others noted that often students misinterpreted or deviated from the question and that there may need to be further clarity in the way the questions are structured and asked, to obtain the most accurate evaluation of their clinical reasoning skills. Assessors identified that further clarification is required around some of the domains as performance criteria used to inform the EPA (e.g. "in Domain 4 – Clinical reasoning, "decision making about whether the script is legal should be separated from decisions about patient safety factors").

Figure 2: Evaluation of Usability Results

Discussion

Despite the importance of evaluating competency in medicine dispensing by pharmacy students, there is limited literature on the assessment of this competency. An integrated simulation-based assessment framework (MEDS) was developed to measure progress towards entrustability for students in medicine dispensing tasks. The framework not only represents promising progress towards the development of a validated, holistic approach to EPA assessment, but a useful tool for providing formative feedback to students in skill development for safe medication supply.

We sought to combine both analytic scoring and holistic rating approaches into our assessment framework. The MEDS assessment framework uses a global scoring approach to collect a holistic view of the performance of each student. Global rating scales can effectively measure certain constructs such as communication and decision making, which are complex and multidimensional.[25, 31-34] Although analytic scoring is associated with increased reliability, holistic scoring is preferred from a validity perspective as, the psychometric properties of global rating scales are often superior, with the ability to more effectively capture the nuanced components of expertise for clinical practice.[35-38] However, we also incorporated 'prompts' for each domain which was designed to inform assessors overall global ratings and feedback to students. Our data suggests that there is good correlation between the total domain score and EPA rating.

We have demonstrated with 10 assessors, that the MEDS assessment framework produces reliable results. Assessment tool ratings were found to be consistent for all outcomes, across all assessors both at the global and domain level. The inter-assessor reliability was also consistent between simulations suggesting that the tool is reliable for different simulations. Lowest levels of inter-rater reliability were observed for professional competencies that require expert assessors to impart a higher level of subjective judgment, based on different levels of understanding. For example, Domain 2 (providing patient-centred care) covers competencies related to provision of medicines information and advice, including an outline of the requirements for monitoring and assessing response to therapy. The assessment of this skill requires not only an objective measure of whether the information was provided to the patient, but also requires the assessor to apply their expert opinion, of how well the pharmacy student engaged in collaborative planning such as eliciting ideas and opinions of the patient, anticipating what to expect and demonstrating mutual decision making. Cognitive processes can be difficult to detect if not verbalised by the pharmacy student, and therefore must be inferred by the assessor through observation. Although assessor training was provided, assessor standardisation could be improved with more scenario-specific training to clarify the expectations within this domain for different scenarios. Interestingly, there are discussions in the literature suggesting that assessor variance may reflect expert judgement and therefore should not be considered solely as measurement error.[37-39]

Analysis of overall EPA rating between simulations shows it is possible there was an effect of the order in which students undertook the simulations. Students who performed simulation one first had higher EPA ratings than students who performed simulation two first. Furthermore, simulation 2 had a significantly lower overall numerical score than simulation 1. This result suggests simulation 2 may have been a more difficult scenario, which is in keeping with its complexity; pharmacy students were required to communicate with a third party (the patient's grandmother), and there was a dispensing error – the label generated from the dispensing software for the prescription had incorrect dosing details. This means that for students to make a safe and appropriate supply they needed to identify the error and make an intervention. Comparatively, scenario 1 was more straightforward, thus it is possible for students to make the correct supply by 'chance' that there were no clinical interventions required. It is possible that students who performed the more difficult simulation first took a 'confidence hit' which affected their performance in the subsequent simulation, while those who performed simulation 1 first were able to 'warm up' and understand the process and expectations before doing the more difficult

simulation. Regardless of the reasons, it is clear that there was a difference between scenarios, which was reflected in assessor ratings demonstrating that the MEDS framework has application in real-world performance-based assessments in which level of difficulty and student performance varies. These were formative assessments for students and were conducted at the very beginning of their Transition to Practice course (a subject representing 25% of the load of the final semester of B Pharmacy program). The study presented a new assessment approach using simulated patients and an integrated clinical task that had not been used previously for these students. Although attempt was made to provide sufficient student briefing, these factors provide some explanation as to why students' results may be lower than expected.

We have presented validity evidence from qualitative research data; assessment blueprinting including mapping and tracking professional competencies, and preliminary psychometric data. Our approach aligns with the current standard for assessment validation, which advocates for evidence to be collected from several sources. [2,10,11] Researchers remain uncertain regarding the types of data that contribute to each evidence source,[40] however there is a growing number of examples in healthcare assessment literature which support the approach we have taken. One well documented model is that of Kane[21]. We have collected evidence of validity through a number of avenues aligned with Kane's framework, but further work is needed to more comprehensively validate the framework for use in different contexts. Positive inter-rater correlations provide evidence for scoring, but further studies can continue to evaluate such evidence, especially using other analysis such as generalisability theory, which can assist in further exploring sources of error related to measurement inaccuracy.[41] Our sampling strategy and assessment blueprint, incorporating a large proportion (88%) of skills, knowledge and attitudes deemed relevant to medicine dispensing by pharmacists, contributes to evidence for generalisation. Our study has shown that scores obtained in this setting are able to discriminate between cases, evidenced by analysing the variation between different simulations, assessors and students. Our assessment framework aims to measure entrustability for independent practice. A strong body of research shows there is a need for robust evidence to support 'rules' for making inferences of competence and readiness for independent practice. The positive correlation between total numerical score and overall global rating contributes to validity evidence, however further studies which detect expert-novice differences across training are required to build on this. Although our assessment framework was able to detect differences between simulations, further studies are required to focus on extrapolation

evidence - the correlation between assessment performance and measures of ability in clinical practice. Translational evidence is lacking more broadly within simulation literature, and there is increasing interest in research to determine whether performance in simulation can produce downstream results for actual patient care.

The relatively small sample of final year pharmacy students and one university is an acknowledged limitation of this study. For resource and feasibility reasons, only two clinical scenarios were used here, but the use of multiple clinical scenarios would add rigour to the validity argument. Structured assessor training was provided individually, and to maintain the element of expert judgement it was decided not to implement further procedures to standardise marking. While the framework validation study offers preliminary indications about the framework's psychometric properties, it is not possible to claim these findings are generalisable to every medication supply situation or every level of pharmacist training. Usability comments relating to further separation of domains to clearly evaluate a scenario have led to review of the MEDS framework and further work is necessary to evaluate the revised version. The MEDS framework has only been evaluated in simulated consultations and while this is widely accepted and useful, it is necessary to observe how the MEDS framework performs in the practice setting.

Conclusion

The importance of using integrated assessment approaches in health professional education has been emphasised, and this study developed a simulation-based MEDS assessment framework to fill gaps in pharmacy education. Our results demonstrate initial validity evidence to support the use of this framework in measuring pharmacy student progress towards entrustment for independent practice, specifically for managing the supply of prescribed medication(s) in a community pharmacy. The MEDS framework may serve as a reliable approach to assessment and could be used in conjunction with other assessment tools to provide more information about level of supervision required in practice, as well as a useful tool for formative student feedback.

Acknowledgements

We would like to acknowledge the Clinical Research Design and Statistics (CR€DITSS) team who provided an analysis of the data collected in this study. We are grateful to the experts from various healthcare fields who contributed to the development and validation of the MEDS framework. We would

also like to thank standardised patients, students and pharmacists who agreed to use their video recorded consultations for this research.

This research was supported by grant funding from the Pharmacy Council of New South Wales.

References

1. Nash R, Chalmers L, Brown N, Jackson S, Peterson G. An international review of the use of competency standards in undergraduate pharmacy education Pharmacy Education 2015;15(1):131-141.
2. Cook D, Hatala R. Validation of educational assessments: a primer for simulation and beyond Advances in Simulation 2016;1(31):1-12. https://doi.org/10.1186/s41077-016-0033-y
3. Haughey S, O'Hare R. Pharmacy OSCEs and Competency-Based Assessments Belfast, UK: Elsevier 2017.
4. Australian Pharmacy Council. Intern Year Blueprint Literature Review Canberra, ACT: Australian Pharmacy Council Limited 2017.
5. Marotti S, Yu Ting Sim, Macolino K, Rowett D. From entrustment of competence to entrustable professional activities. 2018; https://az659834.vo.msecnd.net/eventsairaueprod/production-shpa-public/bda34d002cfd423ca12e14cad6eb1710
6. LeBlanc VR, Tabak D, Kneebone R, Nestel D, MacRae H, Moulton C-A. Psychometric properties of an integrated assessment of technical and communication skills. The American Journal of Surgery 2009;197:96-101. https://doi.org/10.1016/j.amjsurg.2008.08.011
7. Cook D, Brydges R, Ginsburg S, Hatala R. Technology-enhanced simulation to assess health professionals: a systematic review of validity evidence, research methods and reporting quality Academic Medicine. 2013;88:872-883. https://doi.org/10.1097/ACM.0b013e31828ffdcf
8. Ryall T, Judd B, Gordon C. Simulation-based assessments in health professional education: a systematic review Journal of Multidisciplinary Healthcare. 2016;9:69-82. https://doi.org/10.2147/JMDH.S92695
9. Cees van der Vleuten, Schuwirth L, Scheele F, Driessen E, Hodges B. The assessment of professional competence: Building blocks for theory development. Best Practice and Research. Clinical Obstetrics and Gynaecology. 2010;24:703-719. https://doi.org/10.1016/j.bpobgyn.2010.04.001
10. Cizek G. Progress on validity: the glass half full, the work half done. Assessment in Education: Principles, Policy and Practice 2016;23(2):304-308. https://doi.org/10.1080/0969594X.2016.1156642

11. Cook D, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Medical Education. 2015;49(560-575). https://doi.org/10.1111/medu.12678

12. Koster A, Schalekamp T, Meijerman I. Implementation of Competency-Based Pharmacy Education (CBPE). Pharmacy 2017;5(10):1-16. https://doi.org/10.3390/pharmacy5010010

13. Medina M, Plaza C, Stowe C, et al. Center for the Advancement of Pharmacy Education 2013 Educational Outcomes American Journal of Pharmaceutical Education. 2013;77(8):Article 162. https://doi.org/10.5688/ajpe778162

14. Pharmaceutical Society of Australia. National competency standards framework for pharmacists in Australia Canberra 2016. https://my.psa.org.au/s/article/2016-Competency-Framework. Published July 2016. Accessed 11 September 2019.

15. Siddiqui ZS. Framework for an effective assessment: From rocky roads to silk route Pakistan Journal of Medical Sciences. 2017;33(2):505-509. https://doi.org/10.12669/pjms.332.12334

16. Franklin N, Melville P. Competency assessment tools: An exploration of the pedagogical issues facing competency assessment for nurses in the clinical environment Collegian. 2013;22:25-31. https://doi.org/10.1016/j.colegn.2013.10.005

17. Crossley J, Johnson G, Booth J, Wade W. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales Medical Education. 2011;45:560-569. https://doi.org/10.1111/j.1365-2923.2010.03913.x

18. Wass V, Cees Van der Vleuten, Shatzer J, Jones R. Assessment of clinical competence. The Lancet. 2001;357:945-949. https://doi.org/10.1016/S0140-6736(00)04221-5

19. Messick S. Validity. In: Linn R, ed. Educational Measurement. New York: American Council on Education and Macmillan 1989:13-103.

20. Kane M. Validation In: Brennan R, ed. Educational Measurement Westport Praeger; 2006:17-64.

21. Tavares W, Brydges R, Myre P, et al. Applying Kane's validity framework to a simulation based assessment of clinical competence Adv Health Sci Educ Theory Pract. 2018;23:323-338. https://doi.org/10.1007/s10459-017-9800-3

22. Croft H, Nesbitt K, Rasiah R, Levett-Jones T, Gilligan C. Safe dispensing in community pharmacies: how to apply the SHELL model for catching errors Clinical Pharmacist 2017;9(7):214-224.

23. Croft, H., Gilligan C., Rasiah R., Levett-Jones T., Schneider J. Current Trends and Opportunities for Competency Assessment in Pharmacy Education – A Literature Review. Pharmacy 2019 (Basel, Switzerland), 7; 67

24. Croft, H., Gilligan C., Rasiah R., Levett-Jones T., Schneider J. Integrated simulation-based skills assessment for evaluating pharmacist competence – A scoping review. Pharm Educ. 2019;19(1):381-396

25. Croft H, Gilligan C, Rasiah R, Levett-Jones T, Schneider J. Thinking in Pharmacy Practice: A Study of Community Pharmacists' Clinical Reasoning in Medication Supply Using the Think-Aloud Method Pharmacy. 2017;6(1):1-14. https://doi.org/10.3390/pharmacy6010001

26. Sullivan GM and Artino AR. Analyzing and Interpreting Data from Likert-Type Scales J Grad Med Ed. 2013;5(4):541-42. https://dx.doi.org/10.4300%2FJGME-5-4-18

27. Sok Ying Liaw, Rashasegaran A, Lai Fun Wong, et al. Development and psychometric testing of a Clinical Reasoning Evaluation Simulation Tool (CREST) for assessing nursing students' abilities to recognise and respond to clinical deterioration Nurse Educ Today. 2018;62:74-79. https://doi.org/10.1016/j.nedt.2017.12.009

28. Shrout P, Fleiss J. Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin 1979;86(420-428). https://doi.org/10.1037//0033-2909.86.2.420

29. Croft H, Gilligan C, Levett-Jones T, Rasiah R, Schneider J. Development and inclusion of an Entrustable Professional Activity (EPA) scale in a simulation-based medicine dispensing assessment. Curr Pharm Teach Learn 12(2) 203-212. https://doi.org/10.1016/j.cptl.2019.11.015

30. Australian Government Department of Health. Communicable Disease Control. 2018; www.health.gov.au/main/publishing.nsf/Content/ohp-communic-1. Accessed 15 September 2019

31. Australian Bureau of Statistics. National Health Priority Areas. 2019; www.abs.gov.au/ausstats/abs@nsf/Products/5317BAD6B8EEE19ACA25757C001EED30?opendocument. Accessed 15 September 2019

32. Pharmacy Connection. Making Decisions In the Best Interest of the Patient In: Ontario, Canada: Ontario College of Pharmacists 2018.

33. Easton K, Morgan T, Williamson M. Medication safety in the community: A review of the literature. Sydney: National Prescribing Service; June, 2009.

34. Scheffer S, Muehlinghaus I, Froehmel A, Ortwein H. Assessing students' communication skills: Validation of a global rating Adv Health Sci Educ Theory Pract. 2007;13(5):583-592. https://doi.org/10.1007/s10459-007-9074-2

35. Boulet JR. Summative assessment in medicine: the promise of simulation for high-stakes evaluation. Academic Emerg Med. 2008;15(11):1017-1024. https://doi.org/10.1111/j.1553-2712.2008.00228.x

36. Boulet J, Rebbecchi T, Denton E, McKinley D, Whelan G. Assessing the written communication skills of medical school graduates Adv Health Sci Educ Theory Pract. 2004;9:47-60. https://doi.org/10.1023/B:AHSE.0000012216.39378.15

37. Regeher G, MacRae H, Reznick R. Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination Acad Med. 1998;73:993-997. https://doi.org/10.1097/00001888-199809000-00020

38. Jonathan Ilgen, Irene W Y Ma, Rose Hatala, David A Cook. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ. 2015;49(2):161-173. https://doi.org/10.1111/medu.12621

49. Marceau M, Gallagher F, Young M, St-Onge C. Validity as a social imperative for assessment in health professions education: a concept analysis Med Educ. 2018;52:641-653. https://doi.org/10.1111/medu.13574

40. Cook DA, Zendejas B, Hamstra SJ, Hatala R, Brydges R. What counts as validity evidence? Examples and prevalence in a systematic review of simulation-based assessment. Adv Health Sci Educ Theory Pract. 2014;19(2):233-250. https://doi.org/10.1007/s10459-013-9458-4

41. Bloch R, Norman G. Generalizability theory for the perplexed: A practical introduction and guide: AMEE Guide No. 68. Medical Teach. 2012;34(11):960-992. https://doi.org/10.3109/0142159X.2012.703791