Currents in Pharmacy Teaching and Learning 14 (2022) 701–711

Contents lists available at ScienceDirect

Currents in Pharmacy Teaching and Learning

journal homepage: www.sciencedirect.com/journal/currents-in-pharmacy-teaching-and-learning

Research Paper

Validity evidence for summative performance evaluations in postgraduate community pharmacy education

Marnix P.D. Westein [a,*], Andries S. Koster [b], Hester E.M. Daelmans [c], Carlos F. Collares [d], Marcel L. Bouvy [e], Rashmi A. Kusurkar [f]

[a] Department of Pharmaceutical Sciences, Utrecht University, Royal Dutch Pharmacists Association (KNMP), Research in Education, Faculty of Medicine Vrije Universiteit, Amsterdam, the Netherlands
[b] Department of Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands
[c] Master's programme of Medicine, Faculty of Medicine Vrije Universiteit, Amsterdam, the Netherlands
[d] Maastricht University Faculty of Health Medicine and Life Sciences, Maastricht, the Netherlands
[e] Department of Pharmaceutical Sciences, Utrecht University, Utrecht, the Netherlands
[f] Research in Education, Faculty of Medicine Vrije Universiteit, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Supervisor
Workplace-based assessment
Performance evaluation
CanMEDS
Decision making
Shadow assessment system

ABSTRACT

Introduction: Workplace-based assessment of competencies is complex. In this study, the validity of summative performance evaluations (SPEs) made by supervisors in a two-year longitudinal supervisor-trainee relationship was investigated in a postgraduate community pharmacy specialization program in the Netherlands. The construct of competence was based on an adapted version of the 2005 Canadian Medical Education Directive for Specialists (CanMEDS) framework.
Methods: The study had a case study design. Both quantitative and qualitative data were collected. The year 1 and year 2 SPE scores of 342 trainees were analyzed using confirmatory factor analysis and generalizability theory. Semi-structured interviews were held with 15 supervisors and the program director to analyze the inferences they made and the impact of SPE scores on the decision-making process.
Results: A good model fit was found for the adapted CanMEDS based seven-factor construct. The reliability/precision of the SPE measurements could not be completely isolated, as every trainee was trained in one pharmacy and evaluated by one supervisor. Qualitative analysis revealed that supervisors varied in their standards for scoring competencies. Some supervisors were reluctant to fail trainees. The competency scores had little impact on the high-stakes decision made by the program director.
Conclusions: The adapted CanMEDS competency framework provided a valid structure to measure competence. The reliability/precision of SPE measurements could not be established and the SPE measurements provided limited input for the decision-making process. Indications of a shadow assessment system in the pharmacies need further investigation.

* Corresponding author.
E-mail addresses: m.p.d.westein@uu.nl (M.P.D. Westein), a.s.koster@uu.nl (A.S. Koster), hem.daelmans@amsterdamumc.nl (H.E.M. Daelmans), c.collares@maastrichtuniversity.nl (C.F. Collares), m.l.bouvy@uu.nl (M.L. Bouvy), r.kusurkar@amsterdamumc.nl (R.A. Kusurkar).

https://doi.org/10.1016/j.cptl.2022.06.014

Available online 24 June 2022

Introduction

Assessing competence remains complex in healthcare education despite the use of competency frameworks.[1–5] While the high-stake decision to let a trainee graduate to independent practice is highly dependent on the assessment of performance in the (clinical) workplace, the validity and reliability of supervisor judgement is not always known.[6,7] In this study, quantitative and qualitative research methods were combined to investigate the validity, reliability, and impact of summative performance evaluations (SPEs) on decision making in postgraduate community pharmacy specialization in the Netherlands.

Competency-based education has become dominant in the education of healthcare professionals, such as physicians, nurses, and pharmacists.[2,8–10] Worldwide, >50 competency frameworks have been identified for pharmacists.[11] In the Netherlands, in 2012 it was decided to adapt an existing framework for medical specializations, the Canadian Medical Education Directive for Specialists (CanMEDS), for the use in pharmacy master and postgraduate specializations programs in order to ease coordination across multiple health care education programs.[12,13] The central CanMEDS role (medical expert) was, therefore, adapted to pharmaceutical expert.[13,14]

Despite the wide acceptance of competency frameworks, there have also been areas of critique.[3,15,16] Competencies are defined as observable abilities of a health professional related to a specific activity, and competency frameworks use classification schemes to link general competencies or roles to directly observable behaviors. However, it is argued that competence cannot be broken down into a series of discrete competencies, and that not all aspects of competence are observable and measurable.[17] In a systematic review, the validity and reliability of measurements with nine core competency measurement tools for nurses and physicians were found to be moderate to high.[18] Still, the validity of measurement of competencies cannot be assured and should always be investigated in a given context.

An important aspect of postgraduate competency-based education is that learning and assessment occur predominantly in the (clinical) workplace. Workplace-based assessments focus on measuring concrete activities, while performance evaluations give an overview of measurements during a certain period of time.[19] Supervisors (preceptors) combine their clinical tasks with giving feedback to trainees and evaluating their performance.[8,20] As these assessments are subjective by nature, they require adequate sampling to produce reliable results.[21] It is essential to understand how supervisors evaluate the performance of trainees, as different supervisors may focus on different aspects of performance, infer different reasons for the same behavior, conceptualize competence in different ways, exhibit different levels of stringency, and interpret scales in different ways.[22]

Next, the decision to let trainees graduate to independent practice relies heavily on supervisors' workplace-based performance evaluations. However, a study by Duitsman et al[23] revealed that program directors in medical education are influenced by their own beliefs about learning and education in valuing supervisors' feedback, and sometimes put more value on a supervisor's remarks than on feedback provided in assessment tools.

Directed by the above-mentioned limitations of competency-based performance evaluations, we identified the following research questions to investigate the degree to which validity and reliability evidence supports the interpretations of SPE measurement for the use of decision making in postgraduate community pharmacy education in the Netherlands: (1) what is the relationship between the content of the SPE measurement and the construct of competence as perceived by supervisors, (2) to what degree are the response processes of supervisors consistent with the intended interpretation of SPE scores, (3) to what degree does the internal structure of the SPE measurement conform to the construct of the adapted CanMEDS competency framework, (4) to what degree does the reliability/precision of the SPE measurement support the interpretation of scores for decision making, and (5) how do the SPE measurements impact the decisions made by the program director?

Methods

This study had a case study design. An exploratory approach was taken, with parallel gathering and analysis of quantitative and qualitative data to investigate various aspects of validity in accordance with the Standards for educational and psychological testing.[24,25] Quantitative data were used to examine validity evidence based on internal structure and validity evidence based on reliability/precision of measurements with the SPE tool. Qualitative data were used to examine validity evidence based on test content, response processes, and test consequences. This study was performed within the postgraduate community pharmacy specialization program for pharmacists in the Netherlands.[14] Ethical approval was granted by the NVMO Ethical Review Board (2018.6.9).

Setting

The curriculum consists of two years of workplace-based learning in which trainees are situated in a single pharmacy. Trainees are employed at the pharmacy and deliver pharmaceutical care as part of their training, but are not allowed to bear the final responsibility at the pharmacy.[13] In 2012, the curriculum was modernized based on the CanMEDS 2005 framework, which was adapted to the community pharmacy setting.[14] Within the competency framework, 28 competencies were defined (see eAppendix 1), and a program of assessment in the workplace was implemented in combination with centralized courses and assignments. Each year up to 140 trainees enter the specialization program.

Since 2012, a certified supervisor needs to be present in the pharmacy. The supervisor has to complete a two-day training in which the supervisor is trained to supervise, assess, and evaluate trainees according to the program of assessment. Within the program of assessment, the supervisor has the following tasks: (1) giving feedback during daily routines and assessing the performance of the trainee on 36 Entrustable Professional Activities (EPAs); (2) evaluating performance development of the trainee during trimonthly progress evaluations; and (3) judging the performance with an SPE tool at two milestones, the end of year one and the end of year

two.[14] Pharmacy owners are allowed to act as supervisor concurrently. Unlike most medical specializations in the Netherlands, community pharmacy specialization does not receive government funding. Pharmacy owners establish trainees' employment contracts and commonly pay for the educational costs of the specialization of the trainees.

After the supervisor finalizes a SPE, the program director judges if the trainee has completed all tasks within the program and decides if the performance is adequate. The judgement of the program director at the end of year one is a formal advice on the progress of the trainee. At the end of year two, the program director decides if the trainee is ready to finish the program, in other words is eligible to be registered as a community pharmacist. The program director's decision is in principle based on the workplace-based assessments as well as the centralized courses and assignments, which are collected in a digital portfolio (see Table 1). When the program director has doubts about a trainee's level of competence, a clinical competency committee is consulted to conduct an independent portfolio review.

A specialist registration committee is responsible for certifying the supervisor, approving the pharmacy, and registering the trainees.[14]

Quantitative data collection and analysis

The anonymous SPE scores of 342 trainees, who started the specialization between January 2012 and September 2015, were extracted from their digital portfolios in July 2018. At the end of years one and two, supervisors assigned a score of 1 to 4 (1 = insufficient, 2 = moderate, 3 = adequate, 4 = good) during the SPE for each of the 28 competencies according to the trainees' level of performance. These scores aim to indicate the level of competence for the seven CanMEDS roles.

The competency scores were treated as ordered categorical variables. To examine the validity evidence based on internal structure, first the scores for each CanMEDS role were calculated and the internal consistency (reliability) was examined for the underlying competencies using McDonald's omega.[26] Secondly, the correlations were determined between the CanMEDS roles using Spearman's correlation coefficient (ρ). Thirdly, Confirmatory Factor Analysis was performed to examine if the inter-relationships among the items supported the inference of competence through the seven-factor model (see model A in Fig. 1).[27] Scores were set as categorical variables and as an estimator Weighted Least Squares Mean and Variance (WLSMV) was used. The following indices were used to estimate model fit: Chi-square (Chi,^{2} degrees of freedom N, and *P* value), the Root Mean Square Error of Approximation (RMSEA, good <0.05), the Comparison of Fit Index (CFI, good >0.95), and the Tucker-Lewis Index (TLI, good >0.95).[28] A chi-square test for difference testing was used to compare the fit of the seven-factor model with the fit of three alternative models (see the models B, C and D in Fig. 1).[29]

To examine the reliability/precision of the SPE measurement, we analyzed the sources of variance using generalizability theory.[30–32] A crossed measurement design was used: p / t * c, in which p stands for the trainee-supervisor-pharmacy complex (universe of possible allowed conditions = infinite), t stands for the year of study (universe of possible allowed conditions = 2), and c stands for competency (universe of possible allowed conditions = 28). These three facets and the interactions between them together with the error variance (e) explain the total variance (variance components: p, t, c, p x t, p x c, t x c, and p x t x c x e). As the trainees and supervisors worked in one pharmacy and each supervisor had scored only one trainee, trainees and supervisors were both nested within the pharmacy setting. Thus the variances caused by the trainees, the supervisors, and the pharmacy settings could not be calculated separately. The generalizability (G) and the dependability (D) coefficients were calculated to estimate the reliability of the SPE measurement in both a norm-oriented and a domain-oriented perspective. For low stakes decisions a score between 0.70 and 0.80 is considered acceptable and a score >0.80 is considered good. For high stakes decisions a reliability estimate of >0.90 is desired. In addition, we calculated individual reliability estimates for each role based on the individual estimates for the standard error of measurement from the seven-factor model.[33]

Descriptive analysis was performed with IBM SPSS statistics, version 24 (IBM Corp.), McDonald's omega was calculated with Jasp, version 0.13.1 (University of Amsterdam), Confirmatory Factor Analysis was performed with MPlus, version 8 (Muthén & Muthén), and analysis of the sources of variance was performed with EduG, version 6.1 (Educan).

Qualitative data collection and analysis

Semi-structured interviews were held with supervisors and the transcribed data was analyzed to examine the relationship between the content of the SPE measurement and the construct of competence it intended to measure, and to examine the response processes of

Table 1
Overview of portfolio data accessible to the program director.

Setting	Accessible tool (type of data)
Workplace	Personal development plans (narratives)
	EPA assessments (scores and narrative feedback)
	360-degree feedback (scores and narrative feedback)
	Progress evaluations (scores and narrative feedback)
	SPE (scores and narrative feedback)
Centralized courses and assignments	Centralized courses (attendance yes/no)
	Centralized activities (pass/fail, narrative feedback)

EPA = entrustable professional activities; SPE = summative performance evaluations.

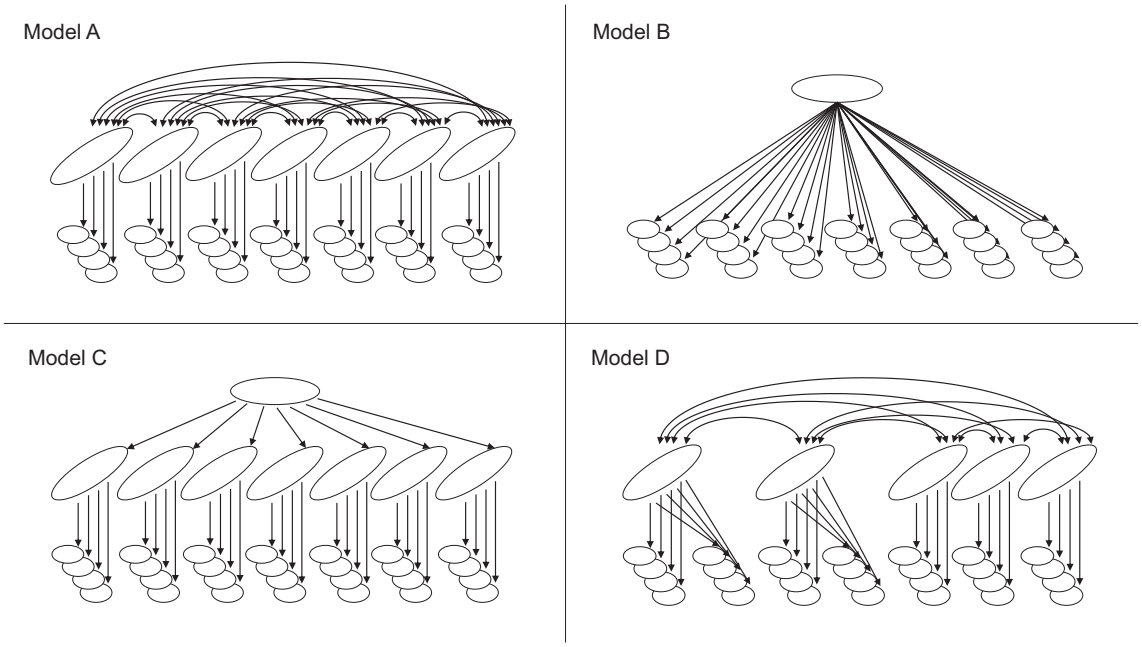

Fig. 1. Confirmatory factor analysis models.
Model A: Seven-factor model used to evaluate the performance of trainees and the alternative models tested. Model B: 1-factor model: 28 competencies, directly measuring competence, without distinguishing between Canadian Medical Education for Specialists (CanMEDS) roles. Model C: second-order-7-factor model: 28 competencies, using seven CanMEDS roles which converge to competence on a second order. Model D: 5-factor model described by Kassam et al[29]: 28 competencies, use of combined CanMEDS roles for pharmaceutical expert and scholar and for communicator and collaborator, and the separate roles for health advocate, manager, and professional subscales.

the supervisors in relation to the intended interpretation of scores. The first researcher deductively developed a set of questions and codes based on Kane's validity framework and van der Vleuten's model for programmatic assessment to examine the inferences made by supervisors.[19,34,35] Two research team members gave feedback on the set of questions and codes. These questions (see eAppendix 2) and codes were finalized in the full research team. In order to obtain recent experiences with the SPE measurement, supervisors whose trainees had successfully completed their training within six months prior to the supervisors' interview were selected for this study. Supervisors were approached by email and telephone to join the study. They received a description of the study by email and provided an informed consent. In order to prepare for the interview, the participants received the questions in advance through email. The first researcher interviewed the participants by telephone at a predetermined date and time. Interviews were audio recorded and then transcribed verbatim. Afterwards, supervisors received a transcript to check for errors. Prior to coding the interviews were anonymized. After conducting the first two interviews, two researchers coded both interviews using the predetermined codes (directed content analysis).[36] The researchers compared notes and discussed the differences until they reached consensus on the coding. The first researcher then conducted and coded the remaining interviews based on the agreed strategy. Data collection started with 15 interviews. If the last two interviews revealed new sub-themes, as judged by the first researcher, two additional interviews would be conducted. Otherwise, data sufficiency was considered as attained. Next, the first researcher analyzed the coded data and discussed them within the team to identify themes and subthemes.[36] Coding was done using Atlas.ti, version 8 (Scientific Software Development GmbH).

A semi-structured interview was held with the program director by two researchers to get an insight into how the SPE scores affected the final decision making by the program director. This interview was also audio recorded and transcribed verbatim. The program director received the transcript to check for errors. The data were coded deductively and themes were generated.

Results

In this section, we first report the characteristics of the trainees, whose scoring data were analyzed, and the characteristics of the interviewees. Next, we report the validity evidence based on the Standards for Educational and Psychological Testing (based on test content, response processes, internal structure, reliability/precision, and test consequences).[25] Where relevant, reference is made to the quotes supplied in eAppendix 3.

Characteristics of the trainees

From the 342 trainees who had started their training in the given period, 304 (89%) had completed it, 23 (7%) trainees had

discontinued their training, and 15 (4%) trainees were still in training at the moment of data extraction. Most trainees were female (n = 237, 69%). On average, the trainees were 27 years old at the beginning of their postgraduate training (range = 24 to 42 years). During the year one and year two SPEs, all the trainees were judged suitable by their supervisor to progress to year two and capable of completing the program at the end of year two, respectively. From the 23 trainees that had discontinued their training, 15 had done so before receiving their first year SPE. Another four had done so before receiving their second year SPE. Although no records of the reasons of trainees for discontinuing the program were kept, the specialist registration committee considered the main reasons to be early termination of their employment, personal circumstances (e.g. illness), and choice to work outside community pharmacy.

Characteristics of the interviewees

Supervisors were interviewed between December 2018 and April 2019. As the 14th and 15th interview revealed no new subthemes, data sufficiency was considered reached after 15 interviews. To reach this number of interviews, thirty-two supervisors had been approached: eight supervisors could not be reached and nine supervisors were not willing to participate in the interviews (mainly because of 'lack of time'). There were interviewees from both urban and rural pharmacies. Nine interviewees were male and six were female. There were six interviewees who had supervised a trainee for their first time, and nine interviewees who had supervised one or more trainees prior to their last trainee in the current or previous curriculum. The interviewed supervisors were considered to be representative of the variability within the supervisor population. The program director, who was responsible for the specialization program between January 2017 and June 2020 was interviewed in January 2019.

Validity evidence based on test content

Most supervisors were satisfied with the SPE tool and said that the 28 competencies covered all aspects that were relevant for a community pharmacist (quote 1.1, eAppendix 3). Some supervisors said that the competencies described for the role of manager were more focused on working in a pharmacy than on fully managing one. These supervisors felt several management skills were neglected, such as personnel management, financial management, and leadership in the pharmacy team (quote 1.2, eAppendix 3). In accordance with the critique aimed at the manager role, some supervisors desired more attention to the collaboration of the trainee with or within the pharmacy team, in the role of collaborator (quote 1.3, eAppendix 3). The program director made no remarks about the validity of the SPE measurement based on test content.

Validity evidence based on response processes

The interviews showed that supervisors held different views on scoring the competencies. Some supervisors were unwilling to gives scores of 1 (= insufficient) or 2 (= moderate), because they regarded these as demotivating or because they were conscious of judging their own level of performance in a negative manner (quote 2.1, eAppendix 3). Also, some supervisors felt the competencies were not adequately defined, which seemed to impede these supervisors from scoring the competencies as less than adequate (quotes 2.2, eAppendix 3). Meanwhile other supervisors were reluctant to give a score of 4 (= good) at the year one SPE, because this would leave no room for improvement in year two (quote 2.3, eAppendix 3).

In accordance to the above findings, the program director reported to experience a large variation in the way supervisors judged the performance of trainees and acknowledged the subjectivity of the SPE scores. The program director experienced that some supervisors were strict and others lenient. Some supervisors focused on written feedback, while most gave oral feedback. Also, the level of detail in written feedback varied. In the experience of the program director, differences between supervisors were reflected in (the lack of)

Table 2
Summative performance evaluation scores of the trainees for each CanMEDS role.

Role (Range)	SPE year 1 (n = 317)			SPE year 2 (n = 308)		
	Mean (SD)	Score < 3 (%)	McDonald's Omega (95% CI)	Mean (SD)	Score < 3 (%)	McDonald's Omega (95% CI)
Pharmaceutical Expert (1–4)	3.55 (0.45)	4.4	0.85 (0.81–0.88)	3.91 (0.22)	0	0.76 (0.69–0.97)
Communicator (1–4)	3.47 (0.46)	6.6	0.82 (0.77–0.86)	3.83 (0.31)	1.3	0.81 (0.75–0.86)
Collaborator (1–4)	3.40 (0.45)	9.5	0.81 (0.77–0.84)	3.77 (0.35)	1.3	0.81 (0.76–0.86)
Scholar (1–4)	3.39 (0.47)	9.1	0.78 (0.73–0.82)	3.76 (0.35)	1.6	0.76 (0.68–0.82)
Health Advocate (1–4)	3.31 (0.46)	11	0.81 (0.76–0.86)	3.75 (0.36)	1	0.81 (0.76–0.85)
Manager (1–4)	3.42 (0.46)	8.8	0.77 (0.72–0.81)	3.79 (0.32)	1.9	0.73 (0.67–0.79)
Professional (1–4)	3.58 (0.44)	3.8	0.84 (0.80–0.87)	3.88 (0.25)	0.3	0.76 (0.67–0.84)

SPE = summative performance evaluations.

variation in the SPE scores they assigned (quote 2.4, eAppendix 3). It troubled the program director when supervisors reported that trainees' progress was as planned while the digital portfolio was missing information described in Table 1 (quote 2.5, eAppendix 3). Meanwhile, the program director was not aware of having a preconceived judgement towards certain supervisors.

While the curriculum was designed to prepare community pharmacists for independent practice, it became apparent during the interviews that some supervisors accepted an end level of the specialization program in which the trainee was not completely fit for independent practice. Sometimes, supervisors would advise the trainee to continue working under the guidance of a senior community pharmacist. Or they would advise to start working as a community pharmacist in a small community pharmacy or a chain-owned pharmacy (where the management is partly done at a central office). Such an advice was given as verbal or written comments and was not necessarily reflected in the SPE score (quote 2.6, eAppendix 3).

Validity evidence based on internal structure

Table 2 shows the descriptive statistics for the year one and year two SPE scores for each CanMEDS role. In both years, supervisors scored the roles of pharmaceutical expert and professional the highest, and the role of health advocate the lowest. Few trainees had an average score that was less than adequate (< 3) on one or more roles, especially in year two. The internal consistency was acceptable for each CanMEDS role with McDonald's omega ranging from 0.73 to 0.85. The Spearman's correlations between all CanMEDS roles were significant ($P < .01$) in both years. Correlations were high for the year one SPE, with the highest between manager and professional (r 0.75) and the lowest correlation between the roles of communicator and scholar (r 0.58). Correlations were moderate to high for the year two SPE, with the highest between health advocate and manager (r 0.68) and the lowest correlation between scholar and professional (r 0.48).

Confirmatory factory analysis suggested that the seven-factor model (Model A in Fig. 1) had a good model-fit as shown in Table 3. Due to the high correlations between competencies and high correlations between roles, the alternative one-factor model (Model B), second order seven-factor model (Model C), and five-factor model (Model D), also had a good model-fit. However, Chi-square difference testing demonstrated a better fit for the original seven-factor model compared to the alternative models at both year one and year two, which led us to accept the seven-factor model.

Validity evidence based on reliability/precision

Generalizability theory was used to calculate the sources of variance for the SPE scores as described in Table 4. The variance, caused by the trainees-supervisor-pharmacy complex (p) explained 19% of the total variance. The interaction between p and the year of study (t) explained 18% of the total variance, the interaction between p and the competencies (c) explained 21% of the total variance. The error variance was 31%. The reliability of the SPE scores for comparing the performance between trainee-supervisor-pharmacy sites was acceptable (G-coefficient 0.78), but insufficient to interpret the performance at the site of training for making high-stakes decisions (D-coefficient 0.65). With each trainee nested in a context of one pharmacy and one supervisor, the score variances caused by trainees, supervisors, and pharmacies could not be calculated separately. It was not possible to establish the reliability of the SPE scores for comparing the performance between trainees alone. The results for the individual reliability estimates show that for each role the reliability decreased, and error increased for the most proficient trainees. The results indicate fair reliability of measurements within a single pharmacy setting for comparing competencies and years of scoring of trainees, but not across sites for comparing the

Table 3
Fit indices for the models tested using confirmatory factor analysis.

Model	Chi-square	df	Chi-square / df	CFI	TLI	RMSEA (95% CI)	Chi-square test
Year 1							
7-factor (Model A)	603.6	329	1.84	0.981	0.978	0.051 (0.045–0.058)	
1-factor (Model B)	837.2	350	2.39	0.966	0.963	0.066 (0.061–0.072)	$\chi^2_{21} = 200.5$ ($P < .001$)
2nd order 7-factor (Model C)	641.7	343	1.87	0.979	0.977	0.052 (0.046–0.059)	$\chi^2_{14} = 48.5$ ($P < .001$)
5-factor (Model D)	703.9	340	2.07	0.974	0.971	0.058 (0.052–0.064)	$\chi^2_{11} = 94.4$ ($P < .001$)
Year 2							
7-factor (Model A)	495.4	329	1.51	0.980	0.977	0.041 (0.033–0.048)	
1-factor (Model B)	606.8	350	1.73	0.969	0.967	0.049 (0.042–0.055)	$\chi^2_{21} = 120.7$ ($P < .001$)
2nd order 7-factor (Model C)	522.5	343	1.52	0.979	0.976	0.041 (0.034–0.048)	$\chi^2_{14} = 36.2$ ($P < .001$)
5-factor (Model D)	547.4	340	1.61	0.975	0.972	0.044 (0.038–0.051)	$\chi^2_{11} = 59.8$ ($P < .001$)

CFI = comparison of fit index; df = degrees of freedom; RMSEA = root mean square error of approximation; TLI = Tucker-Lewis index.

Table 4
Sources of variance for the summative performance evaluation scores.

Source of variance	Number of items	Variance estimate	% of total variability
trainee-supervisor-pharmacy (p)	307	1236.5	19
year of study (t)	2	582.6	8.9
competency (c)	28	136.2	2.0
p x t		585.8	18
p x c		1341.2	21
t x c		12.5	0.1
p x t x c x error (e)		968.4	31

performance of trainees or for making high-stakes decisions.

Validity based on test consequences

When considering the overall suitability of trainees at the end of years 1 and 2, supervisors said they mainly focused on the SPE scores, needing them to be at a certain level. Most supervisors said that trainees needed to perform all competencies at an adequate (= 3) to good (= 4) level (quote 3.1, eAppendix 3). Some supervisors emphasized specific aspects of competence such as reflectivity and team-collaboration, while other supervisors had a more subjective or holistic view (quote 3.2, eAppendix 3). Supervisors said that trainees would be considered unsuitable if the level of performance on one or more roles was insufficient to moderate. On the other hand, one supervisor remarked that the performance in the role of scholar was by definition adequate due to the fact that trainees had graduated from university.

While the program director acknowledged the importance of the supervisors' judgement within the program, she also called it a weakness in the system. According to the program director, her own role in the decision-making process was to look at all aspects of the trainees performance within the program, thereby balancing the judgement of the supervisor (quote 3.3, eAppendix 3). The program director checked if all assignments had been performed and if the performances were adequate. A staff member assisted by systematically checking the portfolio data (see Table 1). When inconsistencies within the portfolio were identified, the program director would contact the trainee. If needed, she contacted the supervisor and sometimes also the teacher(s) from the centralized courses to get a better understanding of the situation and to discuss the inconsistencies found in the portfolio. After making inquiries with the trainee, supervisor, and teacher(s), if the program director still had doubts about the suitability of a trainee, she would turn to the clinical competency committee for advice (quote 3.4, eAppendix 3).

It was not uncommon for trainees to get delayed during their specialization when tasks were not completed on time. However, the program director commented that in seven years, out of the approximately 700 trainees that had joined the program, none had been judged incompetent on the basis of their performance at the local pharmacy. A few trainees had been judged incompetent based on their classroom assignments. The program director noticed that in practice, supervisors sometimes overrode the formal assessment program and played a decisive role in the discontinuation of the training. They did this by telling trainees they were unsuitable to become a community pharmacist and/or by ending their employment (quote 3.5, eAppendix 3).

Discussion

Our study underlines the importance of having an in-depth understanding of the validity and reliability of supervisor judgement and its impact on decision making in competency-based education.[35,37] The content of the SPE measurement in the postgraduate community pharmacy education program had an acceptable relationship to the construct of competence for the community pharmacist based on the adapted CanMEDS framework, and the relationships among the competencies and roles conformed to the CanMEDS seven-factor construct. However, the response processes of supervisors were not consistent with the intended interpretation of SPE scores. While the reliability/precision of the SPE measurement for inferring trainee competence could not be determined fully, it appeared insufficient for making high stakes decisions. The SPE scores had limited impact on the decisions the program director made about trainees' readiness for independent practice.

Validity evidence for using the adapted CanMEDS framework

Many countries have developed their own national competency frameworks for pharmacists.[11] Also, some countries have used the FIP global frameworks for foundation and advanced pharmacy practice to guide their national frameworks.[11,38,39] Our results confirm previous findings in Canada, that it is feasible to use an adapted version of the CanMEDS framework for defining competence in community pharmacy education.[40]

Supervisors said that an insufficient or moderate score on a single CanMEDS role would be enough to render a trainee unsuitable for independent practice. In accordance, the SPE measurements did not converge to a single competency factor, and we found a good model fit for the seven-factor construct. Earlier studies had reported difficulties in measuring constructs based on competency frameworks, but in our study alternative models had a lower quality fit.[17,29]

Some supervisors felt that certain management and collaboration skills were neglected in the framework. In accordance with this finding, some supervisors said they rated their trainees competent at the end of year two, while advising against managing a pharmacy

independently. Because van de Pol et al[41] found that community pharmacists in the Netherlands regard pharmacy management tasks less important than the delivery of cognitive pharmaceutical services and quality assurance, the relevance of this finding needs further study.

Reliability/precision when comparing trainees

Supervisors differed in their standards for scoring competencies. Some supervisors reported to be reluctant or hesitant to give scores below 3 (= adequate) for various reasons, while other supervisors were inclined to score competencies below 4 (= good) at the end of year one as to leave room for improvement in the second year. The program director confirmed the variance between supervisors. In accordance, previous research has shown that interrater variance can have a large effect on the reliability/precision of competency-based assessments.[22]

Due to the curricular restrictions (each trainee scored by a single supervisor in a single pharmacy), the interrater variance and contextual variance could not be separated from the score-variance of trainees. When we used the total score-variance of trainees, supervisors and site of training to calculate the reliability/precision of the SPE measurements, we found it to be insufficient for making high-stakes decisions. The four-point scale itself is less reliable for SPE scores at the high end of the scale. We suggest that the sensitivity and wording of the items can be improved to promote more heterogeneous levels of endorsement for participants with the highest latent scores.

Impact on decision-making

In this study, the use of an SPE tool did not lead to a distinction between competent and incompetent trainees on the work floor, as supervisors rated the performance of trainees on the CanMEDS roles as adequate or good in the majority of cases and judged their trainees in all instances suitable to progress to year two and capable of completing the program at the end of year two. Some supervisors regarded their trainees as unfit for independently managing a community pharmacy, but judged them as competent nonetheless. In these cases, supervisors reported to have given oral feedback instead. Rich et al[42] reported that faculty provided trainees with more oral feedback than written documented feedback when they had concerns over the accuracy and generalizability of their judgements. Barriers found in literature for supervisors failing a poorly performing trainee are supervisor's professional or personal considerations, trainee related considerations, unsatisfactory supervisory training and tools, institutional culture, and lack of available remediation for the trainee.[43]

In the decision-making process, the program director systematically looked at other portfolio data to get a better picture of the workplace-based performance. While this was intended to balance supervisor subjectivity, it also lowered the impact of the SPE. Moreover, by looking at the assessments in the portfolio, the program director became a stakeholder in the generalization process of the workplace-based performance in addition to her role as decision maker.[37] Previously, it has been found that the judgement of program directors is influenced by their personal beliefs.[23,44] While in our study the program director was not aware of any bias towards the judgement of supervisors, the scores given to trainees did influence her opinion of the supervisors.

In postgraduate medical education, clinical competency committees have been introduced to make high-stakes decisions based on the assessment data collected in the portfolio, with the intention to improve the validity and reliability of decisions made.[45,46] In the community pharmacy specialization program, a clinical competency committee was available, but it played merely an advisory role. Strengthening the role of the clinical competency committee within the education program can perhaps improve the quality of the decision-making. However, discrepancies can still be found between a committee's review and a program director's decision.[44] Moreover, clinical competency committees remain dependent on the quality of the information available to inform their high-stakes decision and trainees may find the decisions of such committees less credible than their supervisors' opinions.[47]

Presence of shadow assessment

Castanelli et al[48] described a shadow system in assessment among supervisors, indicating a lack of alignment between the formal assessment program of the curriculum and the local procedures for assessing employees. In our study, we found suggestions for a similar misalignment. While both the quantitative and qualitative analyses pointed at a failure to fail trainees underperforming on their SPEs, 7% of the trainees in the study population ended their training in an untimely manner. The program director recollected that in some cases supervisors played a decisive role in trainees' ending their program.

A power imbalance between supervisors and trainees is considered a potential threat to the supervisory relationship. Therefore, non-hierarchical relationships are advocated.[49] Within the community pharmacy specialization program, the supervisors have substantial power, as trainees are employed by them, without external funding supplied. This could potentially influence the educational environment at the workplace negatively.[50] Since no exit-interviews were held within the program, we could not determine whether ending the program untimely resulted from a decision of the supervisor or the trainee.

Limitations

Our findings are limited to a single postgraduate specialization program for community pharmacists but have similarities with findings in other assessment programs in healthcare professions education.[47] We used fixed cutoffs for the fit indices to judge model fit of the adapted CanMEDS framework. However, there is criticism that fixed cutoffs still lack sufficient power to detect wrong

specifications of the proposed models. Tools for performing dynamic fit cutoff calculations are becoming increasingly available and could prove a direction for further CFA studies.[51] The design of the educational program in which every trainee was scored by a single supervisor was a major limitation for inferring the reliability/precision of the SPE measurements. To correct for interrater variance, multiple supervisors at each pharmacy performing the SPEs are required. Also, we could not relate the SPE scores to other constructs, which could have helped building the validity argument. Trainees' experiences with the SPE measurement were not investigated. Their views could add perspective to the validity evidence gathered in this study. Another limitation was the absence of information about the reasons for individuals discontinuing the program. Data from these cases could have introduced variability into the observed scores and shed light on local procedures for assessing trainees, besides the formal assessment system. Taking into account these limitations, this study can be an example for others attempting to assess competence in a community pharmacy setting. However, we recommend caution with generalizability of the findings.

Suggestions for practice

The results of this study revealed the weaknesses in the program of assessment with having a single supervisor performing the SPEs. From our experience in the community pharmacy context in the Netherlands (although informed by the greater health-professions literature), we suggest that the reliability of the performance evaluations may be improved by introducing multiple supervisors for rating trainee performance and by supplying additional training to supervisors with respect to the interpretation of SPE scores.[22,52] Additionally, response processes of supervisors can be directed by using rubrics to increase objectivity of the scoring options.[53]

In our study, we found several supervisors reluctant to fail trainees. A study by Yepes-Rios et al[43] identified three enablers that can be used to support supervisors' willingness to fail a failing trainee: (1) trigger their duty to patients, to society, and to the profession; (2) provide institutional support; and (3) generate alternate career opportunities for students after failing. However, tension between assessment of learning and assessment for learning among supervisors is unlikely to be eliminated.[47,54,55]

A different approach would be to have the clinical competency committee take up the role of aggregating portfolio data and making motivated decisions.[45,46,56,57] Supervisors could then focus fully on delivering feedback and guiding the learning process.[57] Furthermore, the shadow system in assessment which seemed present in some pharmacies warrants further investigation of the educational environment and the reasons for trainees discontinuing the education program.[48,49]

Conclusions

Our findings suggest that the adapted CanMEDS competency framework and the defined competencies provided a valid construct for measuring competence in this community pharmacy specialization program. The availability of a single supervisor for each trainee performing the SPEs was a major limitation in measuring the reliability/precision of the SPE measurements. Moreover, some supervisors appeared reluctant to fail trainees, and using the SPE tool did not lead to a distinction between competent and incompetent trainees. As a result, SPE measurements provided limited input for the decision-making process. Meanwhile, there were signs that supervisors influenced trainees' decisions to untimely end their training, aside from the formal assessment program.

Declaration of Competing Interests

None.

Disclosure(s)

None.

CRediT authorship contribution statement

Marnix P.D. Westein: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Andries S. Koster:** Conceptualization, Software, Validation, Formal analysis, Writing – review & editing. **Hester E.M. Daelmans:** Validation, Formal analysis, Writing – review & editing. **Carlos F. Collares:** Software, Validation, Resources, Writing – review & editing. **Marcel L. Bouvy:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Rashmi A. Kusurkar:** Conceptualization, Methodology, Software, Validation, Resources, Writing – review & editing, Supervision.

Acknowledgments

We would like to thank the supervisors and the program director, dr. A. Floor, for their open-hearted participation in this study. We would like to thank the American Journal of Pharmaceutical Education for their permission to publish eAppendix 1.[14]

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cptl.2022.06.014.

References

1 Gruppen L, Frank JR, Lockyer J, et al. Toward a research agenda for competency-based medical education. *Med Teach*. 2017;39(6):623–630. https://doi.org/10.1080/0142159X.2017.1315065.
2 Lurie SJ. History and practice of competency-based assessment. *Med Educ*. 2012;46(1):49–57. https://doi.org/10.1111/j.1365-2923.2011.04142.x.
3 Holmboe ES, Sherbino J, Englander R, Snell L, Frank JR, Collaborators ICBME. A call to action: the controversy of and rationale for competency-based medical education. *Med Teach*. 2017;39(6):574–581. https://doi.org/10.1080/0142159X.2017.1315067.
4 Austin Z. Competency and its many meanings. *Pharmacy (Basel)*. 2019;7(2):37. https://doi.org/10.3390/pharmacy7020037.
5 Paradis E, Zhao R, Kellar J, Thompson A. How are competency frameworks perceived and taught?: an exploratory study in the context of pharmacy education. *Perspect Med Educ*. 2018;7(3):200–206. https://doi.org/10.1007/s40037-018-0432-y.
6 Tavares W, Rowland P, Dagnone D, McEwen LA, Billett S, Sibbald M. Translating outcome frameworks to assessment programmes: implications for validity. *Med Educ*. 2020;54(10):932–942. https://doi.org/10.1111/medu.14287.
7 Boulet JR, Durning SJ. What we measure … and what we should measure in medical education. *Med Educ*. 2019;53(1):86–94. https://doi.org/10.1111/medu.13652.
8 Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638–645. https://doi.org/10.3109/0142159X.2010.501190.
9 Katoue MG, Schwinghammer TL. Competency-based education in pharmacy: a review of its development, applications, and challenges. *J Eval Clin Pract*. 2020;26(4):1114–1123. https://doi.org/10.1111/jep.13362.
10 Goudreau J, Pepin J, Dubois S, Boyer L, Larue C, Legault A. A second generation of the competency-based approach to nursing education. *Int J Nurs Educ Scholarsh*. 2009;6:15. https://doi.org/10.2202/1548-923X.1685.
11 Udoh A, Bruno-Tomé A, Ernawati DK, Galbraith K, Bates I. The development, validity and applicability to practice of pharmacy-related competency frameworks: a systematic review. *Res Soc Admin Pharm*. 2021;17(10):1697–1718. https://doi.org/10.1016/j.sapharm.2021.02.014.
12 Scheele F, Teunissen P, Luijk SV, et al. Introducing competency-based postgraduate medical education in the Netherlands. *Med Teach*. 2008;30(3):248–253. https://doi.org/10.1080/01421590801993022.
13 Koster AS, Mantel-Teeuwisse AK, Woerdenbag HJ, et al. Alignment of CanMEDS-based undergraduate and postgraduate pharmacy curricula in the Netherlands. *Pharmacy (Basel)*. 2020;8(3):117. https://doi.org/10.3390/pharmacy8030117.
14 Westein MPD, de Vries H, Floor A, Koster AS, Buurma H. Development of a postgraduate community pharmacist specialization program using CanMEDS competencies, and entrustable professional activities. *Am J Pharm Educ*. 2019;83(6):6863. https://doi.org/10.5688/ajpe6863.
15 Boyd VA, Whitehead CR, Thille P, Ginsburg S, Brydges R, Kuper A. Competency-based medical education: the discourse of infallibility. *Med Educ*. 2018;52(1):45–57. https://doi.org/10.1111/medu.13467.
16 Whitehead CR, Kuper A, Hodges B, Ellaway R. Conceptual and practical challenges in the assessment of physician competencies. *Med Teach*. 2015;37(3):245–251. https://doi.org/10.3109/0142159X.2014.993599.
17 Lurie SJ, Mooney CJ, Lyness JM. Measurement of the general competencies of the accreditation council for graduate medical education: a systematic review. *Acad Med*. 2009;84(3):301–309. https://doi.org/10.1097/ACM.0b013e3181971f08.
18 Yaqoob Mohammed Al Jabri F, Kvist T, Azimirad M, Turunen H. A systematic review of healthcare professionals' core competency instruments. *Nurs Health Sci*. 2021;23(1):87–102. https://doi.org/10.1111/nhs.12804.
19 van der Vleuten CPM, Schuwirth LWT, Driessen EW, et al. A model for programmatic assessment fit for purpose. *Med Teach*. 2012;34(3):205–214. https://doi.org/10.3109/0142159X.2012.652239.
20 Kogan JR, Conforti LN, Iobst WF, Holmboe ES. Reconceptualizing variable rater assessments as both an educational and clinical care problem. *Acad Med*. 2014;89(5):721–727. https://doi.org/10.1097/ACM.0000000000000221.
21 van der Vleuten CPM, Schuwirth LWT, Scheele F, Driessen EW, Hodges B. The assessment of professional competence: building blocks for theory development. *Best Pract Res Clin Obstet Gynaecol*. 2010;24(6):703–719. https://doi.org/10.1016/j.bpobgyn.2010.04.001.
22 Dory V, Gomez-Garibello C, Cruess R, Cruess S, Cummings B, Young M. The challenges of detecting progress in generic competencies in the clinical setting. *Med Educ*. 2018;52(12):1259–1270. https://doi.org/10.1111/medu.13749.
23 Duitsman ME, Fluit CRMG, van der Goot WE, ten Kate-Booij M, de Graaf J, Jaarsma DADC. Judging residents' performance: a qualitative study using grounded theory. *BMC Med Educ*. 2019;19(1):13. https://doi.org/10.1186/s12909-018-1446-1.
24 Kajamaa A, Mattick K, de la Croix A. How to … do mixed-methods research. *Clin Teach*. 2020;17(3):267–271. https://doi.org/10.1111/tct.13145.
25 American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
26 Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105(3):399–412. https://doi.org/10.1111/bjop.12046.
27 Rios J, Wells C. Validity evidence based on internal structure. *Psicothema*. 2014;26(1):108–116. https://doi.org/10.7334/psicothema2013.260.
28 Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling*. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
29 Kassam A, Donnon T, Rigby I. Validity and reliability of an in-training evaluation report to measure the CanMEDS roles in emergency medicine residents. *Can J Emerg Med*. 2014;16(2):144–150. https://doi.org/10.2310/8000.2013.130958.
30 Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE guide no. 68. *Med Teach*. 2012;34(11):960–992. https://doi.org/10.3109/0142159X.2012.703791.
31 Crossley J, Davies H, Humphris G, Jolly B. Generalisability: a key to unlock professional assessment. *Med Educ*. 2002;36(10):972–978. https://doi.org/10.1046/j.1365-2923.2002.01320.x.
32 Mema B, Park YS, Kotsakis A. Validity and feasibility evidence of objective structured clinical examination to assess competencies of pediatric critical care trainees. *Crit Care Med*. 2016;44(5):948–953. https://doi.org/10.1097/CCM.0000000000001604.
33 Pontual AADD, Tófoli LF, Collares CF, Ramaekers JG, Corradi-Webster CM. The setting questionnaire for the Ayahuasca experience: questionnaire development and internal structure. *Front Psychol*. 2021;12, 679016. https://doi.org/10.3389/fpsyg.2021.679016.
34 Kane M, Crooks T, Cohen A. Validating measures of performance. *Educ Meas Issues Pract*. 1999;18(2):5–17. https://doi.org/10.1111/j.1745-3992.1999.tb00010.x.
35 Schuwirth LWT, van der Vleuten CPM. Programmatic assessment and Kane's validity perspective. *Med Educ*. 2012;46(1):38–48. https://doi.org/10.1111/j.1365-2923.2011.04098.x.
36 Hsieh H, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res*. 2005;15(9):1277–1288. https://doi.org/10.1177/1049732305276687.
37 Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. *Med Educ*. 2015;49(6):560–575. https://doi.org/10.1111/medu.12678.
38 *Pharmacy Education Taskforce: A Global Competency Framework*. International Pharmaceutical Federation; 2012. Accessed 9 June 2022 https://www.fip.org/file/1412.
39 *Advanced Practice and Specialisation in Pharmacy: Global Report*. International Pharmaceutical Federation; 2015. Accessed 9 June 2022 https://www.fip.org/file/1397.
40 Kennie-Kaulbach N, Farrell B, Ward N, et al. Pharmacist provision of primary health care: a modified Delphi validation of pharmacists' competencies. *BMC Fam Pract*. 2012;13:27. https://doi.org/10.1186/1471-2296-13-27.
41 van de Pol JM, Koster ES, Hövels AM, Bouvy ML. How community pharmacists prioritize cognitive pharmaceutical services. *Res Social Adm Pharm*. 2019;15(9):1088–1094. https://doi.org/10.1016/j.sapharm.2018.09.012.

42 Rich JV, Fostaty Young S, Donnelly C, et al. Competency-based education calls for programmatic assessment: but what does this look like in practice? *J Eval Clin Pract*. 2020;26(4):1087–1095. https://doi.org/10.1111/jep.13328.

43 Yepes-Rios M, Dudek N, Duboyce R, Curtis J, Allard RJ, Varpio L. The failure to fail underperforming trainees in health professions education: a BEME systematic review: BEME guide no. 42. *Med Teach*. 2016;38(11):1092–1099. https://doi.org/10.1080/0142159X.2016.1215414.

44 Schumacher DJ, Poynter S, Burman N, et al. Justifications for discrepancies between competency committee and program director recommended resident supervisory roles. *Acad Pediatr*. 2019;19(5):561–565. https://doi.org/10.1016/j.acap.2018.12.003.

45 French JC, Dannefer EF, Colbert CY. A systematic approach toward building a fully operational clinical competency committee. *J Surg Educ*. 2014;71(6):e22–e27. https://doi.org/10.1016/j.jsurg.2014.04.005.

46 Kinnear B, Warm EJ, Hauer KE. Twelve tips to maximize the value of a clinical competency committee in postgraduate medical education. *Med Teach*. 2018;40(11):1110–1115. https://doi.org/10.1080/0142159X.2018.1474191.

47 Schut S, Maggio LA, Heeneman S, van Tartwijk J, van der Vleuten CPM, Driessen E. Where the rubber meets the road — an integrative review of programmatic assessment in health care professions education. *Perspect Med Educ*. 2021;10(1):6–13. https://doi.org/10.1007/s40037-020-00625-w.

48 Castanelli DJ, Weller JM, Molloy E, Bearman M. Shadow systems in assessment: how supervisors make progress decisions in practice. *Adv Health Sci Educ*. 2020;25(1):131–147. https://doi.org/10.1007/s10459-019-09913-5.

49 Jackson D, Davison I, Adams R, Edordu A, Picton A. A systematic review of supervisory relationships in general practitioner training. *Med Educ*. 2019;53(9):874–885. https://doi.org/10.1111/medu.13897.

50 Schonrock-Adema J, Bouwkamp-Timmer T, van Hell EA, Cohen-Schotanus J. Key elements in assessing the educational environment: where is the theory? *Adv Health Sci Educ Theory Pract*. 2012;17(5):727–742. https://doi.org/10.1007/s10459-011-9346-8.

51 McNeish D, Wolf MG. Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychol Methods*. 2021. https://doi.org/10.1037/met0000425. Published online 25 October.

52 Moonen-van Loon JMW, Overeem K, Donkers HHLM, van der Vleuten CPM, Driessen EW. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. *Adv Health Sci Educ*. 2013;18(5):1087–1102. https://doi.org/10.1007/s10459-013-9450-z.

53 Panadero E, Jonsson A. A critical review of the arguments against the use of rubrics. *Educ Res Rev*. 2020;30, 100329. https://doi.org/10.1016/j.edurev.2020.100329.

54 Govaerts MJB, van der Vleuten CPM, Holmboe ES. Managing tensions in assessment: moving beyond either- or thinking. *Med Educ*. 2019;53(1):64–75. https://doi.org/10.1111/medu.13656.

55 Bok HG, Teunissen PW, Favier RP, et al. Programmatic assessment of competency-based workplace learning: when theory meets practice. *BMC Med Educ*. 2013;13:123. https://doi.org/10.1186/1472-6920-13-123.

56 Oudkerk Pool A, Govaerts MJB, Jaarsma DADC, Driessen EW. From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Adv Health Sci Educ*. 2018;23(2):275–287. https://doi.org/10.1007/s10459-017-9793-y.

57 Ramani S, Konings KD, Ginsburg S, van der Vleuten CPM. Relationships as the backbone of feedback: exploring preceptor and resident perceptions of their behaviors during feedback conversations. *Acad Med*. 2020;95(7):1073–1081. https://doi.org/10.1097/ACM.0000000000002971.