



# EPA-based assessment: Clinical teachers' challenges when transitioning to a prospective entrustment-supervision scale

Lieselotte Postmes, Femke Tammer, Indra Posthumus, Marjo Wijnen-Meijer, Marieke van der Schaaf & Olle ten Cate

To cite this article: Lieselotte Postmes, Femke Tammer, Indra Posthumus, Marjo Wijnen-Meijer, Marieke van der Schaaf & Olle ten Cate (2021) EPA-based assessment: Clinical teachers' challenges when transitioning to a prospective entrustment-supervision scale, *Medical Teacher*, 43:4, 404-410, DOI: [10.1080/0142159X.2020.1853688](https://doi.org/10.1080/0142159X.2020.1853688)

To link to this article: <https://doi.org/10.1080/0142159X.2020.1853688>



© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 11 Dec 2020.



Submit your article to this journal [↗](#)



Article views: 1877







View related articles [↗](#)



View Crossmark data [↗](#)

## EPA-based assessment: Clinical teachers' challenges when transitioning to a prospective entrustment-supervision scale

Lieselotte Postmes<sup>a</sup> , Femke Tammer<sup>b</sup>, Indra Posthumus<sup>c</sup>, Marjo Wijnen-Meijer<sup>d</sup> ,  
Marieke van der Schaaf<sup>a</sup>  and Olle ten Cate<sup>a</sup> 

<sup>a</sup>Center for Research and Development of Education, University Medical Center Utrecht, The Netherlands; <sup>b</sup>Clinical Genetics Department, Radboud UMC, Nijmegen, The Netherlands; <sup>c</sup>Amphia Ziekenhuis, Internal Medicine, Breda, The Netherlands; <sup>d</sup>School of Medicine, TUM Medical Education Center, Technical University of Munich, Munich, Germany

### ABSTRACT

**Background:** This study explores the challenges clinical teachers face when first using a prospective entrustment-supervision (ES) scale in a curriculum based on Entrustable Professional Activities (EPAs). A prospective ES scale has the purpose to estimate at which level of supervision a student will be ready to perform an activity in subsequent encounters.

**Methods:** We studied the transition to prospective assessment of medical students in clerkships via semi-structured interviews with twelve purposefully sampled clinical teachers, shortly after the introduction of a new undergraduate EPA-based curriculum and EPA-based assessment employing a prospective ES scale.

**Results:** While some clinical teachers showed a correct interpretation, rating strategies also appeared to be affected by the target supervision level for completion of the clerkship. Instructions to estimate readiness for a supervision level in the future were not always understood. Further, teachers' interpretation of the scale anchors relied heavily on the phrasing.

**Discussion:** Prospective assessment asks clinical teachers to make an extra inference step in their judgement process from reporting observed performance to estimating future level of supervision. This requires a change in mindset when coming from a retrospective, performance-oriented assessment method, i.e., reporting what was observed. Our findings suggest optimizing the ES-scale wordings and improving faculty development.

### KEYWORDS

Clinical; undergraduate; work-based

## Introduction



Medical education programs are increasingly applying entrustable professional activities (EPAs) and entrustment-focused assessments (Shorey et al. 2019). While introduced for postgraduate training (ten Cate and Scheele 2007), undergraduate medical schools in several countries are implementing EPAs (AFMC EPA Working Group 2016; Englander et al. 2016; Jucker-Kupper 2016; ten Cate et al. 2018). EPAs describe the work that needs to be carried out in health care (ten Cate 2020). They are defined as units of essential professional practice (e.g. obtaining a history and performing a physical examination) for qualified professionals and they require the integration of knowledge, skills and attitudes gained through training (Chen and ten Cate 2018). Assessment within the EPA framework is directed toward entrustment decisions and thus aligns with the typical judgement processes of clinical teachers, who are used to making every-day informal, implicit entrustment decisions regarding learners, in the course of workplace-based training. When doing so, they must estimate and decide how much supervision a student requires and what their readiness is to participate in clinical care in a safe way (Crossley et al. 2011). EPA-based workplace assessments can make these everyday entrustment decisions explicit

### Practice points

- Prospective entrustment-supervision (ES) scales require estimating suitable levels of supervision in the future, rather than retrospectively reporting how much supervision students required.
- The clinical teachers interviewed were generally unaware of the difference between retrospective and prospective assessment and regularly used the prospective scale to report observed performance without judging readiness for future responsibility.
- The extra inference step of estimating an appropriate supervision level for future performance asks for a changed mindset of raters and for faculty development.

and prepare for summative entrustment decisions that permit a more formal decrease of supervision, which is sometimes called a Statement of Awarded Responsibility (STAR), thus allowing students more autonomy to act in patient care (ten Cate et al. 2015, 2016).

The EPA framework implies an entrustment-supervision (ES) rating scale. Anchors in this scale differ from traditional

**CONTACT** Lieselotte Postmes  [l.m.l.postmes-2@umcutrecht.nl](mailto:l.m.l.postmes-2@umcutrecht.nl)  Center for Research and Development of Education, University Medical Center Utrecht, P.O. Box # 85500, 3508 GA Utrecht, The Netherlands

© 2020 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

rating scales since they describe readiness for a certain degree or level of supervision. In this scale, clinical teachers' trust in a student for a specific activity is reflected and scored holistically (Hauer et al. 2014). ES-scales can be classified into retrospective (focus on the past) and prospective (focus on the future) scales (ten Cate, Carraccio, et al. 2020). Whereas retrospective ES-scales report the actual level of supervision provided during a specific activity (observed behaviour), prospective ES-scales require an estimation of readiness for a specified level of supervision, focusing on future performance. An example of an ES-scale that contains a retrospectively formulated scale anchor is the O-score, as designed for surgical education. Supervisors score the amount of supervision they had to provide in the theatre, such as 'I needed to be in the room just in case' (Gofton et al. 2012; MacEwan et al. 2016; Rekman, Gofton, et al. 2016; Rekman, Hamstra, et al. 2016). An illustration of a prospective ES-scale using an accompanying description is the following: 'Based on my observation today, I suggest for this EPA this trainee may be ready after the next upcoming review to perform with supervision level X' (ten Cate et al. 2015).

The term prospective scale and *prospective assessment* are new in educational literature. The purpose itself is not. Medical students have always implicitly been judged with a prospective purpose. By graduating students at the end of their training for the MD degree, institutions and educators implicitly entrust students with all future tasks and responsibilities that accompany their new position (ten Cate, Carraccio, et al. 2020). While several studies describe prospective scales (Mink et al. 2017; Elias et al. 2018; Valentine et al. 2019), studies comparing retrospective and prospective scales are still scarce (Wijnen-Meijer et al. 2013; Cutrer et al. 2020). In the study of Cutrer et al., untrained raters in UME used a retrospective and prospective scale concurrently. Quantitative analysis suggests that the scales are indeed different. Interviews with eight raters revealed raters notice the difference between the two scales. They expressed preferences depending on the situation, but no uniform pattern could be identified (Cutrer et al. 2020). In a generalisability study by Wijnen-Meijer et al. (2013) raters were asked to score facets of competence in a retrospective assessment and to give a recommendation of supervision for *unfamiliar clinical situations or tasks in the future*, basically reflecting a prospective scale. This study involved candidates at the end of medical school in a simulated environment. Clinician raters were extensively informed using a frame of reference training. Retrospective ratings appeared to be more reliable than prospective ratings (Wijnen-Meijer et al. 2013). This could indicate that the use of a prospective rating scale is more challenging for clinical teachers because it requires to interpret and weigh less visible features (van Enk and ten Cate 2020).

Clinical teachers assessing students for entrustment decisions need to put the new rating scale into practice and, therefore, play an essential role in successful implementation. We sought to understand what challenges clinical teachers in undergraduate medical education (UME) face when they first start using a prospective ES scale. Our research question is: How do clinical teachers handle the initial use of EPA-based assessment employing a prospective ES scale and what challenges do they encounter?

## Methods

We designed a descriptive qualitative study to explore the clinical teachers' transition to the use of prospective assessment. In-depth semi-structured individual interviews were carried out to get a clear idea on the early perceptions and experiences of UME clinical teachers.

## Setting

The study was conducted at the UME program of the University Medical Center Utrecht (UMC Utrecht) in the Netherlands. At the time, UMC Utrecht had recently introduced an EPA-based clinical curriculum in the fourth year of a six-year curriculum with cohorts of about 300 students. Faculty informed students and clinical teachers about the changes via meetings and written instructions.

In total, the curriculum at the UMC Utrecht has five broad EPAs to be mastered in the final year, with clerkships that focus on smaller, nested, mostly discipline-specific EPAs in the preceding years. An absolute, i.e. criterion-referenced, target entrustment level was set for each nested EPA necessary to complete a clerkship. For example, during a clerkship of neurology, basic neurological history taking and physical examination must be mastered at a level of indirect supervision (i.e., without a supervisor present, but quickly available if needed). In this study, we focus on these smaller, nested EPAs. The flow of EPA-related assessment and entrustment decision includes brief observations (comparable to miniCEX and DOPS procedures), and case-based discussions (ten Cate and Hoff 2017). We refer to these as EPA-based workplace assessments (EPA-WBAs). Clinical teachers should report EPA-WBAs with an estimated level of autonomy the student is deemed ready for regarding that specific EPA: a recommendation for future supervision.

Clinical teachers were instructed to relate to the current performance of students and to anticipate an appropriate future level of supervision. In the assessment form, this process was guided by the prospective ES-scale using an accompanying description that emphasizes its prospective nature: 'Based on this EPA-WBA, I *estimate* that this student will be able to perform this activity at supervision level X' (translated from Dutch). The prospective ES-scale was translated, adapted and operationalized from Chen's ES scale (Chen et al. 2015) and included four main categories: (1) presence allowed, no participation; (2) direct supervision; (3) indirect supervision and (4) limited supervision. There were some subcategories for supervision levels 2 and 3, resulting in 7 possible supervision levels (see Figure 1). An e-portfolio aggregates all assessments. For more details about the EPA curriculum at UMC Utrecht, see ten Cate et al. (2018).

## Participants

We used assessor data from UMC Utrecht students' e-portfolio system to purposefully sample clinical teachers with sufficient early experience with the new entrustment decision-making procedure. We sampled clinicians from various disciplines, affiliated hospitals and with and without involvement in the curriculum implementation in the undergraduate clerkships. We looked for clinicians with some experience with EPA-WBAs, some experience with summative decision making, and both.

Based on this (EPA-WBA) assessment, I estimate that this student will be able to perform this activity at supervision level:

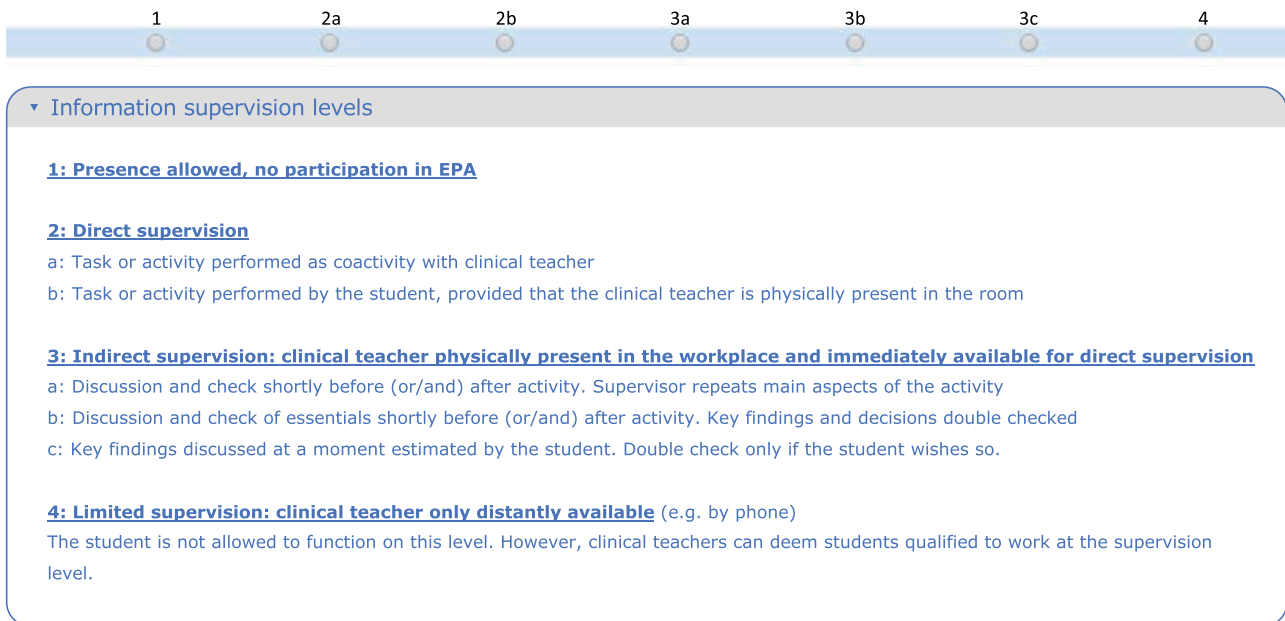


Figure 1. Prospective ES-scale used for evaluation in EPA-WBA (translated print screen).

### Data collection

In January and February 2017, during one-hour sessions, LP and IP interviewed clinical teachers using a semi-structured protocol. After analyzing the first 12 interviews, we felt that the last interviews had offered no new perspectives and explanations and concluded to have reached data saturation. Data collection took place 4–5 months after launching the new EPA-based curriculum, although local experiences with the new curriculum varied from 1.5 to 5 months. In order to capture clinical teachers' experience and views of the EPA-related concepts, we asked interviewees to (i) give their definitions of EPAs, supervision levels, EPA-WBA, and STARs, (ii) describe the coherence of these assessment elements, (iii) describe how they used entrustment-based assessment, (iv) give their opinion about the clarity, usability and value of entrustment-based assessment, and (v) provide any suggestions for improvement. Screen images of the e-portfolio were shown during the interview to prompt and clarify reactions.

### Data analysis

FT and LP transcribed the interviews verbatim and used inductive, open coding, aiming to find themes in teachers' assessment perspective, use and challenges. Via an iterative process, FT and LP discussed the themes at least four times over two months, until reaching consensus. Transcripts were coded separately and reviewed afterwards by the other researcher to ensure consistency and reliability. Differences were discussed and resolved. Coding was conducted using Dedoose(R), version 7.6.17 (SocioCultural Research Consultants, LLC). The quotes were translated aiming to retain the original Dutch meaning as carefully as possible.

### Ethical considerations

The Netherlands Association for Medical Education Ethics Review Board approved the study (case number 800).

### Results

#### Sample

Our sample included twelve clinical teachers, seven males and five females, from six different specialities. Six were from UMC Utrecht and six from affiliated teaching hospitals. The experience and prior knowledge at the time of the interview varied. Four had rated fewer than five EPA-WBAs, four five to ten, and four more than ten. Five had made fewer than five summative decisions, three had 5–15, and four had more than 15. Four were actively involved in the design and implementation of the new curriculum; five had only attended informational meetings about the new curriculum; three relied solely on the information provided to them through the local coordinator, directly through the students, assessment forms, or instructional emails.

#### Themes

Open coding of the transcripts resulted in codes, categorized into clusters according to three central themes: how clinical teachers interpreted the scale anchors, handled the prospective ES scale and experienced the challenges of the prospective nature of the ES scale.

**Clinical teachers' interpretation of the scale anchors** describes aspects mentioned by clinical teachers about how they interpreted the scale anchors (see Figure 1).

Clinical teachers felt able to interpret the global categories (direct supervision, indirect supervision, limited supervision). We noticed that they relied heavily on written descriptions of the displayed dummy e-portfolio to describe their interpretation of different supervision levels. As one of the participants reflects:

I always have to look up the [supervision levels'] exact meanings. To tell you the truth, I don't always remember well. Let's have a look. Main aspects, uh, yes, so you check from more to less [supervision], right? (Clinician 12)

Several participants described distinguishing between the more detailed sublevels (a, b and c; see Figure 1) as

being more difficult. We found three explanations. Most mentioned difficulties due to terminology or phrasing:

[supervision levels] 3b and 3c, for example, what exactly is the difference? Yes, the words are different. [...] Repeats main aspects [3a]. Yes, key findings [3b&c]. That difference isn't always entirely clear. (Clinician 6)

A second explanation observed was a mismatch between formal supervision levels and the preference of clinical teachers. For example, one interviewee commented that the amount of supervision she provided and wanted to rate was located somewhere between level 1 b and 2a. Finally, some interviewees experienced the scale as a continuum and subsequently had difficulties drawing the line between the categories:

Well, I find that quite difficult sometimes. 3a and 3b, for example, discuss and check the/an activity shortly before and/or after the student. Supervisor repeats main aspects of the activity [3a]. It is a sliding scale, right. It is always a sliding scale, from 3a to 3b. (Clinician 9)

**Handling the prospective ES scale** as the second theme includes factors that influence the judgement process and how clinical teachers decide on a recommended supervision level for students.

Clinical teachers mentioned a wide variety of factors used in the judgement process triggered by the prospective ES scale, e.g. trust, students' self-assessment, knowledge, and clinical reasoning skills. This is illustrated by various quotes:

So, I kind of have a basic question of: do I entrust this [EPA] to you if it would concern my own family? [...] sometimes I judge: I just don't trust the person that much. And then the grade starts to shift in my mind. (Clinician 11)

So, you can also just ask the student, like: how did you do? And do you think you can now do this on your own and afterwards only call me? [...] 'So, most [students] can estimate well themselves what they are good at. Students are becoming doctors, so they always take that very seriously. (Clinician 1)

[...] especially if someone also takes the steps of the clinical reasoning process well and I notice that he or she just really understood it, I tend to lower supervision, so a higher grade [supervision level], or whatever you want to call it [...]. (Clinician 3)

During the interviews, it became clear that the target entrustment level necessary to complete a clerkship influenced how clinical teachers handled the prospective ES scale. The target entrustment level was well known among all respondents. This was due to clinical teachers' experience with the summative assessment: 'I should declare them competent for 3a' (Clinician 4), but also because their students referred to it often: '[...] the other day, a student said to me: you should rate that as a 3a.' (Clinician 6).

The use of a target supervision level had several effects. On the one hand, it clarified requirements for students at the end of the clerkship and it made clinical teachers aware of what they needed to aspire to: 'And that we have to strive for, that is the level they [students] have to reach [...].' (Clinician 5). On the other hand, the target entrustment level influenced the judgement process, leading to implicit expectations of what supervision level should be rated. One interviewee even stated: 'We focus our rating around 3a because that is the norm [...].' (Clinician 10). Finally, the target entrustment level served an essential role for clinical teachers who had trouble interpreting the

ES-scale or felt that they lacked information. As a clinical teacher who had trouble with both explains:

If I think: ok, this student is functioning in a way that I normally find sufficient [...] I score a 3a [the target supervision level for the specific EPA] and don't worry about it. (Clinician 5)

One of the clinical teachers who noticed that her colleagues were affected by the target supervision level, reflected upon the advantages and disadvantages of the target supervision level:

I think this [omitting the target entrustment level] results in a fairer answer with more variance. On the other hand, then you [as a clinical teacher] don't know what the norm is [...] How do you then determine where to place someone on that scale? (Clinician 12)

During the interviews, it was clear that the ease with which clinical teachers handled the supervision scale varied. Clinicians who could clearly describe how they handled the supervision scale and came to their judgement generally were 'closer to the information source'. Furthermore, those who tended to focus less on terminology seemed less troubled by their decision. As one clinician argued:

But maybe from 3a to b, to c, it gets more difficult, I mean: what is indirect supervision or limited supervision? I don't even think it's... It's not even super important. I think that it is mainly semantics. [...] I am a little more open about that. (Clinician 10)

**Challenges of the prospective nature of the ES-scale** as the last theme focusses explicitly on challenges that arise due to the prospective nature of the ES-scale.

The instruction for the recommendation of future supervision was that the supervision level should reflect what a student could be entrusted within similar (future) activities, not necessarily the provided amount of supervision. In the interviews, few clinicians referred to these instructions but several described a judgement process that did take future supervision into account such as the clinical teacher who imagined the student doing the same activity without their supervision:

[...] should I be present next time or can he do this on his own? And that's how I determine the [supervision] level. (Clinician 2)

However, it became clear that some clinicians rated observed performance or even gave ratings for recommended supervision based on habits of educational practice, rather than thoughtful individual observation:

I always give a 3 [indirect supervision]. [...] the student usually checks the child, while I am talking to the parents. (Clinician 6)

Also, one interviewee thought she was required to rate observed performance but did end up rating how much supervision she thought was necessary:

I think it [EPA in highly specialized patient population] should always be done under [direct] supervision, but then they [students] would never exceed a 2 [direct supervision], though I see very good clerkship students here' [...] 'I did give a 3 [indirect supervision] once, even though I was physically there [...] I considered him [the student] capable of doing that without me being there. (Clinician 7)

She falsely believed she was not allowed to rate this way because her students had not *actually* performed at this supervision level.



Only one clinical teacher, actively involved in the design and implementation of the new curriculum, could describe the confusion among his colleagues clearly:

I describe it as whether you deem a student qualified to perform a certain task. It does not have to be done, but that you deem him qualified to do so. There is confusion about this amongst my colleagues [...] they say that it's a 2 [direct supervision] because they [the students] have not done it on their own, they have done it under supervision. (Clinician 9)

## Discussion

This study explored clinical teachers' initial uses of a prospective ES scale in UME EPA-based assessment. The purpose of our study was to understand what challenges clinical teachers face. How would clinical teachers adapt to an assessment instruction that, instead of solely evaluating and reporting observed learner behaviour, ask to recommend a suitable level of supervision for future encounters? For many clinical teachers, this novel approach, generated by the use of EPAs and the focus on entrustment decisions, asks for a change in assessment behaviour.

Summarising our results, clinical teachers applied various strategies and approaches while using the ES scale. We found that they interpreted the ES scale anchors not always how they were intended. The prospective nature of the ES scale proved challenging, and instructions to estimate readiness for a supervision level in the future, were not always interpreted as intended.

### The nature of prospective judgement

We found that some clinical teachers just rated performed or observed performance, indicating difficulties in understanding the prospective nature. Young et al. reported similar findings (Young et al. 2020). However, raters did not appear to experience these problems in the studies from Wijnen-Meijer et al. (2013) and Cutrer et al. (2020) where the retrospective and prospective rating scales were provided concurrently. Offering both scales simultaneously might have clarified the prospective assessment question.

This warrants an in-depth analysis of the nature of the judgement that clinical teachers are asked to make. Norcini (2014) describes three categories. In the first, the *quality* of performance is judged. Traditional assessment, be it in written assessment, skills assessment, or workplace-based assessment, mostly falls within this category of valuing observed behaviour. Retrospective assessment asks for judgement within the *occurrence* category (e.g. 'did the learner show performance that required direct supervision?'). However, the nature of the judgement in prospective entrustment lies within the last category of *suitability* or *fitness*. This requires clinical teachers first to establish the performance quality, and second, to decide how much supervision is suitable for the (near) future. This type of judgement is similar to judgements made in daily practice, although this is often not reflected in traditional rating forms (Holmboe et al. 2018). A prospective ES scale requires clinical teachers to evaluate what risks they personally, or the health care, would face if the learner would granted permission to enact the EPA without, or with less, supervision (Damodaran et al. 2017).

Although clinical teachers make frequent, implicit entrustment decisions of prospective nature in daily practice, this apparently does not mean they recognise that prospective assessment asks for the same. The clinical teachers of our study seemed to think that the nature of judgement they were asked to make, lied within the occurrence category. Therefore, prospective rating might require a different mindset, or at least an increased awareness.

### Prospective assessment and inferences

The concept of prospective assessment is new and the additional inference from a reported observed performance to an estimated future level of supervision opens up discussions about inferences within the judgement processes.

Krupat, critically discussing the EPA metric, voiced concerns that this metric reduces accuracy and interrater agreement (Krupat 2017). Additional layers of inference move from evaluating whether a task was performed competently to evaluating students' trustworthiness, which includes truthfulness, conscientiousness, and discernment of limitations (Kennedy et al. 2008; Krupat 2017). One could argue that the additional inference that clinical teachers should make when using a prospective ES scale might reduce reliability and subsequently affect the validity of the scores given to students.

Several empirical studies show that a retrospective ES scale with the question 'how much supervision was needed or provided?' can yield reliable scores (Gofton et al. 2012; George et al. 2014; Rekman, Gofton, et al. 2016; Weller et al. 2017). The study of Wijnen-Meijer, however, showed a lower reliability of the prospective rating scale (Wijnen-Meijer et al. 2013). While we understand these concerns, we feel that any WBA implies eventual entrustment with clinical care responsibilities. Ultimately, WBA is part of a process intended to culminate in graduation and licensure. To optimize the validity of WBA, assessors should include this endpoint as a consequence in their assessment considerations (Kane 2016a, 2016b; ten Cate, Schwartz, et al. 2020). Clinical teachers can judge the consequences of *ad hoc* decisions to entrust care to learners, by evaluating what happened after this act. Such evaluations constitute a source of consequential validity evidence (Downing 2003), which is not often explicitly valued. While future estimations are based on observed behaviour in the past, entrustment decisions do not simply correlate with rated capable behaviour only (Cutrer et al. 2020), as several other factors come into play, such as reliable behaviour, discernment of limitations or humility, integrity and truthfulness, and agency (Kennedy et al. 2008; ten Cate and Chen 2020; ten Cate, Carraccio, et al. 2020).

### Differences among interviewees

Interestingly, some of our interviewees seemed to have more difficulty with prospective rating than others. We discern several reasons that might explain this difference. The first group of explanations can be understood as missing or (mis)interpreting the provided information via meetings, e-mail, or the assessment forms themselves. That few clinical teachers referred to the instructions and that only one of our most informed interviewees was able to express the

prospective-retrospective difference during the interview, could indicate the importance of being well informed. Also, some clinical teachers had difficulties distinguishing supervision sublevels, which could be a consequence of how different questions and instructions were phrased. One may wonder whether valid assessment can be achieved if raters misunderstand scale anchors (Crossley et al., 2011).

Another explanation might lay in other experiences with general entrustment decisions, such as with licensing of doctors, which we argued, has a prospective purpose. This might explain why clinician 7, who appeared not to understand the concept of prospective assessment and falsely believed she should rate *actual* performance, nevertheless rated her students in a prospective way. She could envision students being able to see a patient on their own and knew students whom she would entrust to do so.

This study shows that clinical teachers experience assessment on a prospective ES scale as different from traditional assessment rating observed performance. Only clinical teachers most informed and most experienced with prospective assessments were able to grasp the retrospective-prospective difference and estimate future levels of supervision. The question arises whether more experience and providing better information alone will solve this problem. We argue that using prospective assessment requires new instructions for clinical teachers but also warrants training them to adopt a new approach to assessment. As the transition to prospective assessment can be challenging, continuous attention and evaluation are required.

### Implications for practice and future research

The correct interpretation of a prospective rating scale by clinical teachers is essential. We therefore recommend making the prospective nature of the ES scale as explicit as possible in the assessment form and the scale anchors and, if applicable, the accompanying question. Evaluation of the scale anchors and the accompanying question shortly after implementation is advisable to check whether the prospective nature has become clear to a clinical teacher. This study further endorses faculty development in the assessment and provision of information. Assessing on prospective scales requires a different type of judgement and clinical teachers do not automatically appear to be aware of this. Finally, faculty might consider the effects of a target supervision level while implementing an EPA assessment. We recommend that future studies on EPA-assessment consider whether they use a retrospective or prospective ES scale since their validity and reliability might differ. Follow-up studies on the validity and reliability of prospective ES scales are especially needed, and research settings with clinical teachers having prolonged experience can be valuable.

### Strengths and limitations

This study has several limitations. The single setting with a specific assessment structure and a specific phrasing limits the external validity. Due to the study's recent EPA implementation, results may not be representative of settings in which clinician teachers have more experience with EPAs. Further, our purposeful sampling resulted in a group of

clinical teachers who were all somehow involved in high-stakes summative assessment. This group might feel more responsible for understanding the new assessments than clinical teachers solely involved in low-stakes EPA-WBAs. Challenges might be more extensive or different in clinical teachers solely involved in EPA-WBAs.

Our empirical study describes challenges with early prospective assessment in authentic daily practice. While we should be cautious to generalize, clinical teachers' judgement processes in our study reflect other accounts in the literature (Shorey et al. 2019). This study, therefore, provides important insights into the interpretation and use of the prospective ES scale. This provides practical relevance for curricula implementing new EPAs or instructing faculty members new to prospective assessment.

### Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

### Notes on contributors

**Lieselotte Postmes**, MD, is a PhD student in medical education at the University Medical Center Utrecht, the Netherlands.

**Femke Tammer**, MD, is a Clinical Genetics Resident at the Radboud UMC, Nijmegen, the Netherlands.

**Indra Posthumus**, MD, is a Geriatrician Resident at the Amphia Hospital, Breda, the Netherlands.

**Marjo Wijnen-Meijer**, PhD, is a professor of medical education and team leader of curriculum development at Medical Education Center, Technical University of Munich, Germany.

**Marieke F. van der Schaaf**, PhD, is a professor of Research and Development of Health Professions Education at University Medical Center Utrecht, the Netherlands.

**Olle ten Cate**, PhD, is a professor of medical education and a senior scientist at the Center for Research and Development of Education, University Medical Center Utrecht, the Netherlands.

### ORCID

Lieselotte Postmes  <http://orcid.org/0000-0002-9264-8626>

Marjo Wijnen-Meijer  <http://orcid.org/0000-0001-8401-5047>

Marieke van der Schaaf  <http://orcid.org/0000-0002-4810-5464>

Olle ten Cate  <http://orcid.org/0000-0002-6379-8780>

### References

- AFMC EPA Working Group. 2016. AFMC entrustable professional activities for the transition from medical school to residency. In: Touchie C, Boucher A, editors. Education for primary care. Ottawa (Canada): Association of Faculties of Medicine of Canada; p. 1–26.
- Chen HC, ten Cate O. 2018. Assessment through entrustable professional activities. In: Delany C, Molloy E, editors. Learning & teaching in clinical contexts: a practical guide. Chatswood: Elsevier Australia; p. 284–304.
- Chen HC, van den Broek WES, ten Cate O. 2015. The case for use of entrustable professional activities in undergraduate medical education. *Acad Med.* 90(4):431–436.
- Crossley J, Johnson G, Booth J, Wade W. 2011. Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Med Educ.* 45(6):560–569.
- Cutrer WB, Russell RG, Davidson M, Lomis KD. 2020. Assessing medical student performance of entrustable professional activities: a mixed

- methods comparison of Co-Activity and Supervisory Scales. *Med Teach*. 42(3):325–332.
- Damodaran A, Shulruf B, Jones P. 2017. Trust and risk: a model for medical education. *Med Educ*. 51(9):892–902.
- Downing SM. 2003. Validity: on meaningful interpretation of assessment data. *Med Educ*. 37(9):830–837.
- Eliasz KL, Ark TK, Nick MW, Ng GM, Zabar S, Kalet AL. 2018. Capturing entrustment: using an end-of-training simulated workplace to assess the entrustment of near-graduating medical students from multiple perspectives. *MedSciEduc*. 28(4):739–747.
- Englander R, Flynn T, Call S, Carraccio C, Cleary L, Fulton TB, Garrity MJ, Lieberman SA, Lindeman B, Lypson ML, et al. 2016. Toward defining the foundation of the MD degree: core entrustable professional activities for entering residency. *Acad Med*. 91(10):1352–1358.
- George BC, Teitelbaum EN, Meyerson SL, Schuller MC, DaRosa DA, Petrusa ER, Petito LC, Fryer JP. 2014. Reliability, validity, and feasibility of the zwisch scale for the assessment of intraoperative performance. *J Surg Educ*. 71(6):e90–e96.
- Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. 2012. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med*. 87(10):1401–1407.
- Hauer KE, ten Cate O, Boscardin C, Irby DM, Lobst W, O'Sullivan PS. 2014. Understanding trust as an essential element of trainee supervision and learning in the workplace. *Adv Health Sci Educ*. 19(3):435–456.
- Holmboe ES, Durning SJ, Hawkins RE. (editors) 2018. *A practical guide to the evaluation of clinical competence*. 2nd ed. Philadelphia, PA: Elsevier.
- Jucker-Kupper P. 2016. The “Profiles” document: a modern revision of the objectives of undergraduate medical studies in Switzerland. *Swiss Medical Weekly*. 146(February):14270.
- Kane M. 2016. Validation strategies. Delineating and validating proposed interpretations and uses of test scores. In: Lane S, Raymond MR, Haladyna TM, editors. *Handbook of test development*. 1st ed. New York: Routledge; p. 64–80.
- Kane MT. 2016. Explicating validity. *Assess Educ*. 23(2):198–211.
- Kennedy TJT, Regehr G, Baker GR, Lingard L. 2008. Point-of-care assessment of medical trainee competence for independent clinical work. *Acad Med*. 83(10 Suppl):S89–S92.
- Krupat E. 2017. Critical thoughts about the core entrustable professional activities in undergraduate medical education. *Acad Med*. 93(3):1.
- MacEwan MJ, Dudek NL, Wood TJ, Gofton WT. 2016. Continued validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): use in the simulated environment. *Teach Learn Med*. 28(1):72–79.
- Mink RB, Schwartz A, Herman BE, Turner DA, Curran ML, Myers A, Hsu DC, Kesselheim JC, Carraccio CL. 2017. Validity of level of supervision scales for assessing pediatric fellows on the common pediatric subspecialty entrustable professional activities. *Acad Med*. 93(2):283–291.
- Norcini JJ. 2014. Workplace assessment. In: Swanwick T, editor. *Understanding medical education: evidence, theory and practice*. 2nd ed. Chichester (UK): John Wiley & Sons, Ltd.; p. 279–292.
- Rekman J, Gofton W, Dudek N, Gofton T, Hamstra SJ. 2016. Entrustability scales: outlining their usefulness for competency-based clinical assessment. *Acad Med*. 91(2):186–190.
- Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. 2016. A new instrument for assessing resident competence in surgical clinic: The Ottawa Clinic Assessment Tool. *J Surg Educ*. 73(4):575–582.
- Shorey S, Lau TC, Lau ST, Ang E. 2019. Entrustable professional activities in health care education: a scoping review. *Med Educ*. 53(8):766–777.
- ten Cate O. 2020. When I say ... entrustability. *Med Educ*. 54(2):103–104.
- ten Cate O, Chen HC. 2020. The ingredients of a rich entrustment decision. *Med Teach*. DOI:10.1080/0142159X.2020.1817348.
- ten Cate O, Hoff RG. 2017. From case-based to entrustment-based discussions. *Clin Teach*. 14(6):385–389.
- ten Cate O, Scheele F. 2007. Competency-based postgraduate training: can we bridge the gap between theory and clinical practice? *Acad Med*. 82(6):542–547.
- ten Cate O, Chen HC, Hoff RG, Peters H, Bok H, van der Schaaf M. 2015. Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Med Teach*. 37(11):983–1002.
- ten Cate O, Hart D, Ankel F, Busari J, Englander R, Glasgow N, Holmboe E, Lobst W, Lovell E, Snell LS, et al. 2016. Entrustment decision making in clinical training. *Acad Med*. 91(2):191–198.
- ten Cate O, Graafmans L, Posthumus I, Welink L, van Dijk M. 2018. The EPA-based Utrecht undergraduate clinical curriculum: development and implementation. *Med Teach*. 40(5):506–508.
- ten Cate O, Carraccio C, Damodaran A, Gofton W, Hamstra SJ, Hart D, Richardson D, Ross S, Schultz K, Warm E, et al. 2020. Entrustment decision making: extending Miller's pyramid. *Acad Med*. DOI:10.1097/ACM.0000000000003800.
- ten Cate O, Schwartz A, Chen HC. 2020. Assessing trainees and making entrustment decisions: on the nature and use of entrustment and supervision scales. *Acad Med*. 95(11):1662–1669.
- Valentine N, Wignes J, Benson J, Clota S, Schuwirth LW. 2019. Entrustable professional activities for workplace assessment of general practice trainees. *Med J Australia*. 210(8):354–359.
- van Enk A, ten Cate O. 2020. “Languaging” tacit judgment in formal postgraduate assessment: the documentation of ad hoc and summative entrustment decisions. *Persp Med Educ*. DOI:10.1007/s40037-020-00616-x.
- Weller JM, Castanelli DJ, Chen Y, Jolly B. 2017. Making robust assessments of specialist trainees' workplace performance. *Br J Anaesth*. 118(2):207–214.
- Wijnen-Meijer M, Van der Schaaf M, Booij E, Harendza S, Boscardin C, Van Wijngaarden J, Ten Cate TJ. 2013. An argument-based approach to the validation of UHTRUST: can we measure how recent graduates can be trusted with unfamiliar tasks? *Adv Health Sci Educ Theory Pract*. 18(5):1009–1027.
- Young JQ, Sugarman R, Schwartz J, McClure M, O'Sullivan PS. 2020. A mobile app to capture EPA assessment data: Utilizing the consolidated framework for implementation research to identify enablers and barriers to engagement. *Perspect Med Educ*. 9(4):210–219.