

# What are Supervisor Perceptions of Clinical Placement Sign-off Forms for Medical Students?

Patrick J McGown<sup>1</sup>, Molly M Nichols<sup>1</sup>, Jennifer Forshaw<sup>2</sup>, Antonia Rich<sup>3</sup>, David Harrison<sup>3</sup>, Celia A Brown<sup>1</sup>, Amir H Sam<sup>1</sup>

<sup>1</sup>Imperial College School of Medicine, Imperial College London, London, UK; <sup>2</sup>Department of Paediatrics, Chelsea & Westminster Hospital NHS Foundation Trust, London, UK; <sup>3</sup>Research Department for Medical Education, University College London, London, UK

Correspondence: Patrick J McGown, Email [pjmcgown@doctors.org.uk](mailto:pjmcgown@doctors.org.uk)

**Background:** Medical students typically must pass an end-of-placement supervisor evaluation for each clinical rotation to progress to the next level of the course. However, evidence suggests these evaluations show weak correlation with other reliable educational assessments. This study aims to determine supervisor perceptions of the sign-off process to inform possible improvements.

**Methods:** Clinical supervisors were recruited to provide feedback on their perceptions of the end-of-placement sign-off process. Responses were received from 59 of the 161 (36.6%) Year 3 clinical placement supervisors in hospitals and General Practitioner (GP) practices affiliated with Imperial College School of Medicine. A mixed methods approach was used, combining statistical and thematic analyses.

**Results:** Perceived positives of the single supervisor sign-off model included accountability, feasibility and continuity of supervision. Reported negatives were inflexibility, limited contact time, psychometric concerns and altered student learning behaviours with different staff. GP supervisors reported significantly more contact time with students than hospital-based supervisors ( $p < 0.001$ ), and were more likely to feel they had adequate supervision time to make informed ratings of student knowledge ( $p < 0.001$ ) and skills ( $p < 0.001$ ). Significantly more supervisors in Medicine (62%) and Surgery (64%) desired a change in the single supervisor sign-off model than GP supervisors (27%) ( $p = 0.04$ ).

**Conclusion:** This study provides novel and rich qualitative data on supervisor perceptions of the clinical sign-off process which have implications for undergraduate medical education curriculum reform. Hospital-based supervisors were less positive towards the sign-off process, likely due to modern working practices of shared patient loads, shorter rotations, and reduced contact time with students. Different sign-off requirements should be considered for GP supervisors versus those in hospital settings; possible alterations include multi-source feedback, embedded in-placement assessment, optimisation of the sign-off form, and additional formalised supervision time in supervisors' work schedules.

**Keywords:** clinical placement supervisors, clinical hospital-based supervisors, undergraduate medical education, clinical assessment, clinical supervisor ratings, clinical sign-off assessments, supervisor perceptions

## Background

Progression through medical school curricula involves the completion of different clinical rotations, in line with situated learning theory, whereby learning is promoted through engagement in the working environment.<sup>1</sup> Clinical competence in these rotations is normally assessed by an end-of-placement supervisor evaluation,<sup>2</sup> which acts as an interim barometer of student progress prior to summative end-of-year examinations. The usual assessor is a consultant-level supervisor who is a highly experienced content expert, able to recognise unsafe levels of performance.<sup>3</sup> On sign-off forms, overall competence is typically subdivided into constituent sub-domains such as professionalism, knowledge and clinical skill, which are then combined to arrive at an overall outcome.<sup>4</sup>

Despite widespread use, the supervisor ratings system has long been criticised for being outdated and suboptimal.<sup>5</sup> Assessors are a major cause of measurement error,<sup>6</sup> and multiple supervisors are required given the high numbers of students per year-group, which introduces significant measurement error from inter-rater variability. Assessors are

seldom briefed on the student curriculum<sup>7</sup> and are often not accustomed to the marking framework.<sup>8</sup> Additionally, raters may concentrate on contrasting components of performance,<sup>9</sup> may not appreciate expected student standards,<sup>10</sup> and may even score the same performance differently on the same rating scale.<sup>11</sup> These issues can be somewhat mitigated through assessor training, which may reduce differences in marking stringency.<sup>11</sup> However, in practice the use of untrained assessors is common.<sup>12</sup>

Furthermore, the format of undergraduate clinical attachments in hospitals has markedly changed since the introduction of the supervisor sign-off system. Rather than the archaic method of one longitudinal placement,<sup>13</sup> it is now routine to undergo a range of multiple different sub-speciality rotations. Consequently, there are shorter rotations and frequent changes of supervisor,<sup>14</sup> with less supervisor-learner contact for individual supervisors in any one rotation.<sup>2</sup> This reduces assessment validity as supervisor ratings are based on a smaller number of interactions,<sup>15</sup> leading some critics to argue that the advantages of rater-based clinical assessments are offset by psychometric drawbacks.<sup>16</sup>

Notably, in studies which investigated the correlation between clinical ratings from single assessors and end-point assessment scores, low or non-significant correlation coefficients have been found.<sup>17,18</sup> Additionally, ratings by resident doctors (non-consultants) demonstrate better correlation to other assessment scores than ratings by consultant-level assessors,<sup>19,20</sup> which calls into question the prevailing method of consultant-led end-of-placement sign-off assessments and suggests that changes should be made to the current consultant-led process.

In educational theory, any proposed changes to curricula or assessments should include input from key stakeholders.<sup>21</sup> Current supervisors are well placed to elucidate existing problems within the sign-off process and their perceptions might inform possible solutions. Therefore, this study aims to ascertain clinical supervisor perceptions of the end-of-placement single supervisor sign-off process, to aid in determining what changes should be made. A systematic literature review was conducted (Figure 1), using the PubMed database and OVID Medline with the following search terms: “supervisor perceptions” OR “faculty perceptions” AND *medical student* OR *undergraduate* AND “sign-off” OR *assessment* OR *evaluation* OR *rating*.

Two single-centre quantitative studies were identified, indicating there is a dearth of literature on supervisor perceptions of the clinical evaluation of medical students. Both studies only used Likert scale questions to gather supervisor perception data. Study characteristics, evaluation methods and outcomes are summarised in [Appendix 1](#).

In one study, supervisors perceived their assessments of students were fair and involved adequate observation time.<sup>23</sup> However, they also reported themselves to be fail-adverse, a finding which is widespread in assessment literature; assessors are less confident in making fail decisions than passing decisions,<sup>24</sup> and may award passing grades to avoid the consequences of expressing concerns over learner performance.<sup>25</sup>

The second study compared supervisor perceptions of evaluating students over a longitudinal placement lasting a whole academic year, versus a shorter clinical rotation. It found supervisors considered longitudinal supervision better for assessment validity, quality of clinical skills evaluation, and the ability to provide feedback,<sup>26</sup> although no comparison with end-point assessments occurred.

Overall, the studies are limited in their exploration of supervisor perceptions. Neither collected qualitative data, and it is unclear what changes supervisors would advocate making to the sign-off process, which merits further exploration.

We therefore formulated the following research questions:

1. What are the perceived positives of the single supervisor sign-off process?
2. What are the perceived negatives of the single supervisor sign-off process?
3. What changes to the end-of-placement sign-off process could be considered based on supervisors’ opinions?

## Methods

### Institutional Context (Imperial College London, UK)

Supervisors of third-year medical students were chosen as this year has the longest clinical placements at our institution. In this year students rotate through three different clinical placements: “Medicine”, “Surgery”, and “General Practice (GP)”. Placement duration is eight weeks for GP, and either eight weeks or two four-week mini-placements in Medicine and Surgery. A named consultant supervisor provides an in-person sign-off rating for each student at the end of their

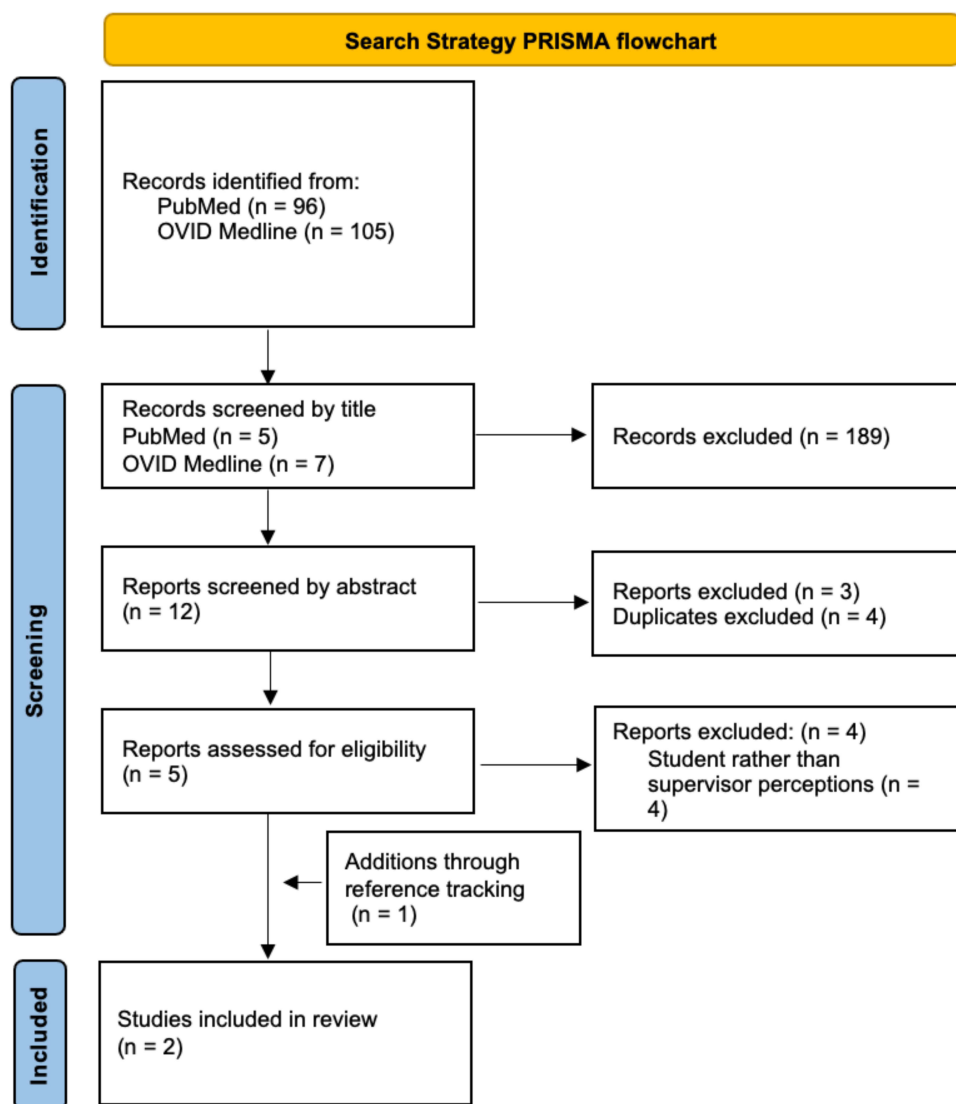


Figure 1 shows the research search strategy flowchart, adapted from Page et al.<sup>22</sup>

attachment (on four-week attachments, students will have a different supervisor for each mini-placement and therefore will receive two sign-off ratings). Supervisors are not made aware of student performance on prior rotations.

The sign-off form ([Appendix 2](#)) uses a four-point ordinal scale with the categories “Below Expectations”, “Borderline”, “Meets Expectations” and “Above Expectations”. An overall grade is awarded, as well as a rating for each of the sub-domains (Clinical Skills; Practical Skills; Communication; Knowledge; Professionalism). Students must achieve either “Meets Expectations” or “Above Expectations” across all ratings to successfully pass the placement. If this is not achieved, completion of a remedial period is required before they are eligible to attempt the high-stakes end-of-year examinations.

Several techniques are used to safeguard assessment validity and reliability. All supervisors undergo assessor training on expected student standards and how to complete the assessment form, including frame-of-reference training. Additionally, students are pre-assigned assessors, removing the possibility of assessor selection bias.<sup>25</sup>

## Data Collection

An anonymous web-based mixed-methods questionnaire was used to collect data ([Appendix 3](#)). Supervisors (all consultants or fully qualified GPs) were recruited at training days and via email using pre-existing institutional mailing

lists, with two reminder emails sent to boost response rates. All 161 Year 3 clinical placement supervisors in hospitals and GP practices affiliated with Imperial College School of Medicine were considered eligible.

Open-ended questions provide better depth of response but may lead to responder fatigue if too many are required,<sup>27</sup> therefore Likert scale questions were also included to complement findings. Seven-point scales were used to provide respondents with more scope to differentiate between values whilst maintaining validity.<sup>27</sup>

It was explained to supervisors that participation was voluntary and no personal or identifiable data would be collected. They were provided with a Participant Information Sheet and informed that completion of the questionnaire was evidence of consent. Participants were made aware that once the questionnaire was completed it would not be possible for them to withdraw retrospectively as all data was recorded anonymously.

## Ethical Approval

The study was conducted in line with the British Educational Research Association's 2018 guidelines.<sup>28</sup> Ethical approval was obtained from Imperial College's Education Ethics Review Process (reference EERP2223-060).

## Data Analysis

### Statistical Analysis

Mean and Standard Deviation (SD) were calculated for each Likert scale question. Ordered chi-squared testing was performed to determine statistical differences between responses by specialty (medicine-surgery; surgery-GP; GP-medicine), and by setting (hospital-GP) with surgery and medicine combined to form a hospital cohort.

## Thematic Analysis of Qualitative Questionnaire Responses

NVivo12 software<sup>29</sup> was used to facilitate coding. Template codes were pre-identified ([Appendix 4](#)) based on the literature review. Qualitative analysis of free-text questionnaire responses started on receipt of the first questionnaire, with emergent coding of unforeseen data.

Repeat analysis of codes was performed to ensure consistency and promote refinement into overarching themes.<sup>30</sup> The questionnaire was closed when thematic saturation was reached; this was determined as the point at which no new additional codes were generated from three consecutive questionnaire responses.<sup>31</sup> Saturation was reached within subgroups after analysing 32 questionnaires overall (11 medicine, 10 surgery, 11 GP).

## Results

Responses were received from 59 of the 161 Year 3 clinical placement supervisors in hospitals and GP practices affiliated with Imperial College School of Medicine (24/61 medicine; 19/49 surgery; 16/51 GP). The overall response rate of 36% is considered adequate within the literature on online questionnaires.<sup>32</sup>

## Statistical Analysis

[Table 1](#) shows summary statistics (mean and SD) for each Likert scale questionnaire item on a seven-point scale where 1 = strongly disagree and 7 = strongly agree, arranged by specialty. It also shows the results for Chi-squared testing of responses between specialties (medicine-surgery; surgery-GP; GP-medicine).

The Holm multiple correction method<sup>33</sup> was used to correct for multiple statistical tests (12 tests: four Likert scale questions for each of the three specialty-specialty comparisons). The adjusted significant p-value was calculated at <0.008 for each sub-test.

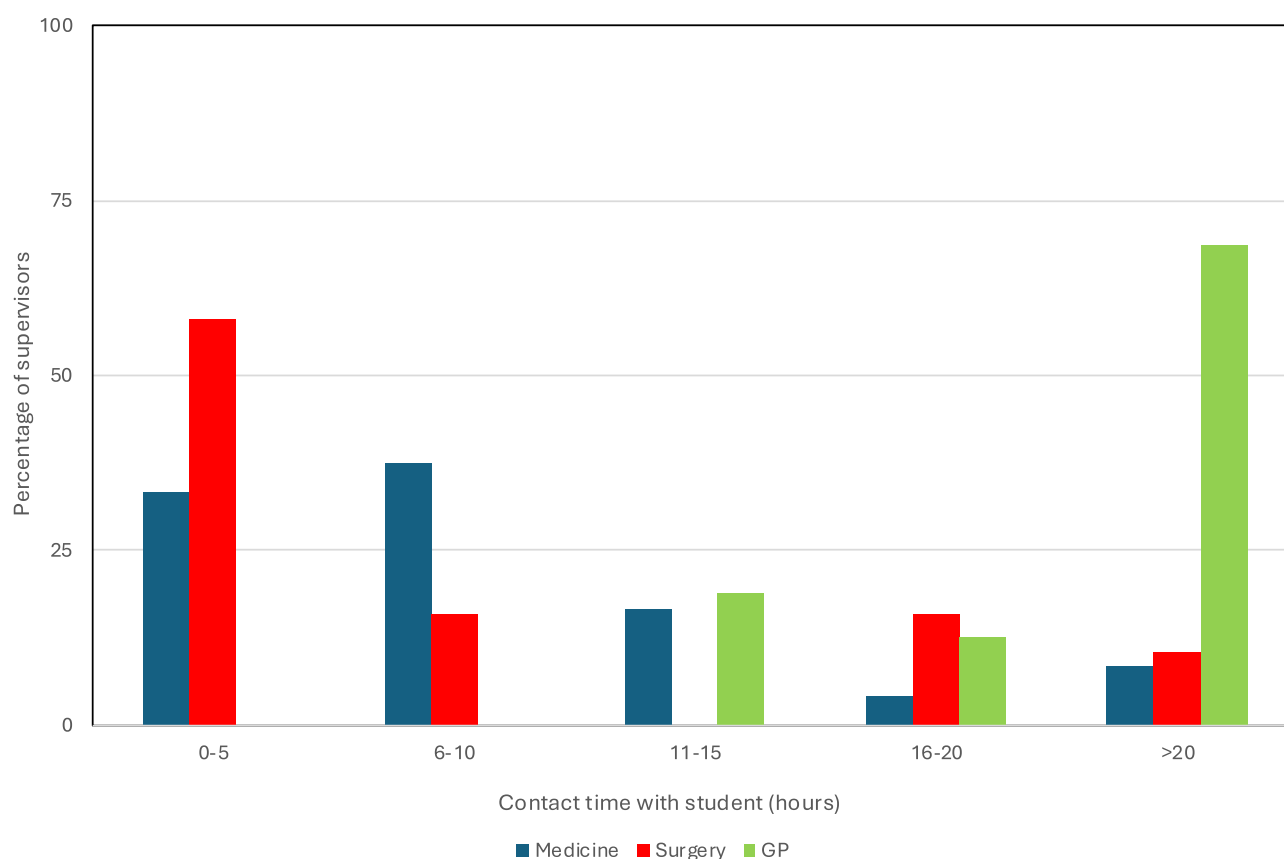
Responses from supervisors in surgery and medicine were not statistically significantly different. However, a statistically significantly higher proportion of GP supervisors perceived they had adequate time to assess students' knowledge and skills. Additionally, statistically significantly higher proportions of GP supervisors perceived their knowledge and skills ratings would correlate with end-of-year assessment scores than medicine supervisors. This same trend was evident between GP and surgery supervisors, but was not statistically significant.

**Table 1** Likert-Scale Responses to Questionnaire Statements by Specialty

Specialty	Statement 1 Mean/7 (SD)	Statement 2 Mean/7 (SD)	Statement 3 Mean/7 (SD)	Statement 4 Mean/7 (SD)
Medicine (n=24)	4.04 (1.59)	3.92 (1.63)	4.17 (1.18)	4.08 (1.29)
Surgery (n=19)	3.79 (1.67)	3.79 (1.61)	4.21 (1.40)	4.26 (1.33)
GP (n=16)	6.06 (0.84)	5.75 (1.52)	5.25 (0.66)	5.19 (0.63)
Inter-Specialty Chi-squared test	Statement 1 p-value	Statement 2 p-value	Statement 3 p-value	Statement 4 p-value
Medicine-Surgery	0.619	0.801	0.912	0.659
Surgery-GP	<0.001*	<0.001*	0.015	0.021
GP-Medicine	<0.001*	0.002*	0.004*	0.005*

**Notes:** Likert scale: 1 = strongly disagree; 7 = strongly agree. \* = statistically significant result. Mean Likert scale responses (/7) to statements in the supervisor questionnaire by specialty, and chi-squared testing of responses between specialties. Statement 1: "I feel that I have adequate supervision time to assess students' *knowledge* on their clinical placement". Statement 2: "I feel that I have adequate supervision time to assess students' *skills* on their clinical placement". Statement 3: "I believe that my assessment of students' *knowledge* reflects their likely performance in end-of-year examinations". Statement 4: "I believe that my assessment of students' *skills* reflects their likely performance in end-of-year examinations".

Figure 2 illustrates the reported supervisor-student contact hours by specialty. GP supervisors had significantly increased contact time with learners, with 68.8% of GPs reporting over 20 contact hours with each student across their placement compared to 9.3% of hospital-based supervisors (GP-medicine:  $p < 0.001$ ; GP-surgery:  $p < 0.001$ ).



**Figure 2** shows clinical supervisors' self-reported contact time with students by specialty.

Additionally, on Chi-squared testing there was a statistically significant disparity ( $p=0.041$ ) between supervisors' desire to make changes to the single supervisor sign-off system. Those based in hospital settings were more in favour of making changes to the process (medicine 62% in favour; surgery 64% in favour) than GP-based supervisors (27% in favour).

## Thematic Analysis of Qualitative Questionnaire Responses

The full list of codes and themes is shown in [Appendix 5](#). The key themes identified during coding were divided into *Positives* and *Negatives* of the single supervisor sign-off model:

### 1. *Positives of the single supervisor sign-off model:*

- a) Accountability
- b) Continuity
- c) Feasibility

### 2. *Negatives of the single supervisor sign-off model:*

- a) Effect on learning
- b) Inflexibility
- c) Limited contact time
- d) Psychometric concerns
- e) Sign-off form shortcomings

## Positives of the Single Supervisor Sign-off Model

### Accountability

Having a single named supervisor was perceived to improve robustness of assessment with “accountability” and “oversight and scrutiny of student performance” compared to having multiple assessors. A named supervisor was seen to improve students’ “motivation”, “engagement” and attendance at the supervisor’s teaching sessions, whilst pre-assigning supervisors was considered beneficial to prevent students from only seeking sign-offs from individuals perceived to be lenient.

### Continuity

The enhanced supervisor-student contact time and “consistency” of observation was felt to aid the identification of gaps in student knowledge, which could facilitate tailored teaching and individualised feedback. Likewise, supervisors articulated that the greater direct contact time with learners provided a more accurate view of competency.

### Feasibility

Supervisors believed a single supervisor model made it more feasible to provide all supervisors with training, which in turn would safeguard ratings against assessor variance. Furthermore, a single assessor could complete the sign-off form efficiently without needing to review multiple mini-assessments from other parties. Finally, the single sign-off model was considered more achievable and less stressful for students than seeking out multiple assessors.

## Negatives of the Single Supervisor Sign-off Model

### Effect on Learning

Interestingly, supervisors thought that having a named supervisor had a negative impact on student learning behaviours with other staff members. Students were viewed as “sign-off focused”, and willing to forgo learning opportunities with doctors who were not their allocated supervisor. When the named supervisor was absent, it was perceived that students often displayed “little effort to attend or impress other members of the team”.

Additionally, respondents articulated that feedback content and style differed between supervisors, which could impact learners more in a single supervisor model as feedback is limited to one voice. Having multiple students

simultaneously assigned to each supervisor was also seen as a potential disadvantage, with dominant students masking the competency of others.

### **Inflexibility**

Several logistical difficulties with the single sign-off model were highlighted by supervisors, including scheduling issues and a lack of flexibility when supervisors were absent due to annual leave, sickness, or the need to provide emergency clinical cover. Timetabling clashes between university and hospital teaching agendas were also reported, with centralised university teaching often precluding student attendance on clinical placement.

The in-person sign-off requirement was criticised for reducing depth of feedback due to time-pressure. Multiple supervisors desired the ability to complete sign-off forms remotely to overcome this.

### **Limited Contact Time**

Short rotations and mixed student attendance levels were perceived to lead to “brief” or “variable contact with individual students”, to the extent that “feedback can be subjective” or “prone to observation bias”. Supervisors acknowledged that variation in case presentations, informal teaching opportunities and student motivations were further drawbacks in making accurate rating judgements within a limited contact time.

### **Psychometric Concerns**

Variation in rating stringency, differing perceptions of acceptable standards, and assessor bias were seen as threats to sign-off rating reliability. A lack of triangulation via multiple assessor ratings was felt to compound this. One supervisor commented that the single supervisor model was no longer fit for purpose due to “modern teamworking practices in hospitals where inpatient care is shared between a number of consultants...so no one person has oversight across the placement”.

Validity concerns were also widespread with the absence of formalised quantitative assessments of knowledge or skills during placements, and supervisors’ reluctance to fail students due to repercussions. Numerous hospital supervisors acknowledged that grades awarded were often based on student attendance, punctuality, and enthusiasm, rather than knowledge or skill.

Suggested changes included incorporating supervised learning events such as Case-Based Discussions into sign-off criteria, or introducing a formalised in-placement assessment with a standardised OSCE or written examination.

### **Sign-off Form Shortcomings**

Numerous supervisors articulated that practical skills should be removed from the sign-off form, with a separate assessment performed by the clinical skills team/teaching fellows who were acknowledged to spend more time with learners. Feedback from others in supervisory roles was also recognised as important, and multiple respondents desired the creation of a multi-source feedback component to increase reliability of assessment.

Finally, there were mixed opinions regarding whether clinical placement performance could be expected to predict examination performance. Some viewed clinical ratings as a better indicator of future performance as a resident doctor than a predictor of end-of-year assessment scores.

## **Discussion**

This study adds rich qualitative detail to the existing literature on supervisor perceptions. Variation in clinical assessment scores should be due to differences between those being assessed, but rater error is known to exert a considerable influence.<sup>2</sup> This psychometric quality was a source of concern for some supervisors, with large numbers of hospital-based supervisors referencing the possibility of subjectivity and bias in a single supervisor assessment model.

The sign-off model was perceived more favourably by GP supervisors, which tallies with advantages cited in the existing literature on longitudinal assessment;<sup>34</sup> community-based placements are more longitudinal in nature than hospital rotations and thus provide increased learner observation time<sup>14</sup> and this improves inter-rater reliability<sup>35</sup> and accuracy of performance ratings.<sup>26</sup>

In hospitals there are numerous barriers to making competency judgements, which exert less effect in GP settings; the more varied environment with frequent interruptions,<sup>36</sup> multiple individuals acting in supplemental supervisory roles when the allocated supervisor is not present,<sup>37</sup> and multiple grades of students in the environment with different learning goals.<sup>36</sup>

Perceptions that student behaviour differed in the absence of their named assessor suggests ratings awarded by supervisors are under-sampling true performance. This is alluded to in the wider literature in a deception study where statistically different competency ratings were awarded to students between expected and unexpected raters, with higher scores awarded when students knew they were being assessed.<sup>19</sup>

Despite its perceived advantages, there was mixed support for the single supervisor sign-off method overall, with calls for more formalised, defensible, and validated mechanisms to aid in the sign-off process. The statistically significant split between opinions of hospital- and GP-based supervisor suggests the sign-off process need not be uniform between community and hospital-based placements.

## Possible Changes to the Sign-off Process

### Use of Multi-Source Feedback in Hospital Settings

The use of multiple assessors is beneficial for assessment reliability<sup>2</sup> and reduces the effect of leniency-bias/generosity error that is hypothesised to lead assessors to rate positively across clinical ratings.<sup>38</sup> Multi-source feedback may also help to alleviate some of the assessment focused student behaviours noted by supervisors, where students were perceived as only attending teaching from the clinician tasked with their sign-off, at the expense of other learning opportunities.

As part of promoting multi-source feedback, hospital-based supervisors suggested including the opinions of junior team members within the sign-off process. This could be advantageous as students spend more time with resident doctors than consultants,<sup>20</sup> and residents may more accurately evaluate student performance than consultant assessors.<sup>19</sup> This surprising phenomenon may be rooted in Schmidt and Moust's concept of social and cognitive congruence (1995). With closer proximity in age and position in the learning hierarchy, junior assessors are likely more aware of expected levels of student competence than consultants, equipping them to make better performance judgements. The effect could be accentuated by student behaviours; junior team members are more approachable from a student's perspective,<sup>39</sup> and students may therefore be more willing to ask them questions which display a lack of knowledge.

Nevertheless, there are downsides to the use of multi-source feedback. Accountability and oversight of student performance could be lost with multiple raters, so having a primary named supervisor may still be required. A study in postgraduates indicates four to 11 raters are required for 80% inter-rater reliability,<sup>40</sup> but this reduces feasibility and necessitates a sizeable increase in institutional administration load. Furthermore, student workload/stress levels might increase when tasked with undertaking multiple sign-offs, and larger volumes of required assessments may act as a barrier to learning rather than an aid.<sup>41</sup> To improve feasibility students could be asked to self-select raters, however assessor selection bias would reduce validity.<sup>38</sup>

### Increase Supervisor-Learner Contact Time on Hospital Placements

Increasing assessor-learner contact time would improve both reliability and validity of assessment, as the brief contact time with learners in hospitals was perceived to lead to the normalisation of "snapshot" competency ratings. This fits with the dominant discourse within the literature, with concerns regarding student supervision time being more pronounced in hospital settings.<sup>42</sup> Longer contact time leads to more accurate competency ratings; in a cohort study where four-week supervisory periods were compared with two-week periods, correlation between clinical knowledge ratings awarded by supervisors and students' end-point assessment scores was superior for longer supervision duration.<sup>43</sup>

Opinions on how to safeguard increased supervisor-learner contact time were united; supervisors desired a contractual change so that supervision time was built into their clinical schedules. However, this may be difficult to achieve due to competing clinical pressures.

## Using Different Assessors for the Practical Skills Sign-off

The assessor being present is one of the cornerstones of assessment validity,<sup>44</sup> yet questionnaire responses suggest that practical skills observation is rarely performed by supervisors in their day-to-day roles. Supervisors were mindful that this lack of observation affected assessment validity, and desired removal of the practical skills component from the sign-off form, suggesting a separate sign-off should be carried out by the clinical skills team.

## Embedded in-Placement Assessments

Proposed embedded in-placement assessments options included either a standardised quantitative assessment of knowledge (eg Single Best Answer Questions, SBAQs), a practical examination (eg an Objective Structured Clinical Examination, OSCE), or the completion of workplace-based assessments (WPBAs). This would mean the use of multiple sampling methods, which would have the benefit of increasing reliability through triangulation.<sup>45</sup>

Advantages of these assessments include the objectivity of SBAQs and a reduction in assessor variance with increased assessor numbers in OSCEs versus single supervisor assessments.<sup>44</sup> Post-hoc calculations of cut-scores would also safeguard against assessors' reluctance to fail.<sup>46</sup> Furthermore, assessment stimulates learning behaviours,<sup>47</sup> and targeted feedback from assessments is valuable for learners,<sup>48</sup> so embedded in-placement assessment could help drive student development.

Unfortunately, additional assessments pose logistical challenges. A question bank would be required where written questions would need to be continually written, whilst recruitment and training of examiners plus the creation of stations and marking schemes are needed for practical examinations. However, embedded in-placement assessment are commonly used in undergraduate medicine in other countries,<sup>17,19,49</sup> indicating feasibility concerns can be overcome.

## Remote Sign-Offs

Reflection on feedback is an important driver of deep learning,<sup>1</sup> however supervisors reported that students typically present in groups for the sign-off, and the need to complete forms for multiple learners in a timely manner restricted the value of feedback given. Additionally, the provision of detailed feedback for learners was sometimes limited by competing clinical demands. Therefore, some supervisors called for the ability to conduct remote sign-offs to improve feasibility of providing personalised feedback.

The drawback would be the lack of student involvement, with no possibility of dynamically addressing learner queries or gauging student self-awareness of performance. Students could also be overwhelmed by feedback which has not been previously mentioned in a face-to-face environment, and may be less willing to accept it as valid.<sup>47</sup> Thus, a remote sign-off would perhaps not be appropriate when wishing to give significant feedback to enact meaningful change. A hybrid model could present a possible sign-off method to explore; either with remote sign-offs plus the option of face-to-face feedback, or where remote feedback is permissible alone only if both supervisor and student are content not to meet for face-to-face feedback. This may reduce administrative burden whilst safeguarding the pedagogical benefits of face-to-face feedback if deemed preferable by either assessor or learner.

## Rating Scale Changes

Assessment reliability is threatened where rating scale anchors are not meaningful to the assessor.<sup>50</sup> The institution's anchors ("Above Expectations"; "Meets Expectations"; "Borderline"; "Fail") were criticised by some supervisors as being "vague" and unhelpful, and supervisors reported that expected levels of performance could be easily confused or forgotten, especially where staff were also supervising students from different years of the curriculum.

Descriptor training is carried out at the institution, but the descriptor framework may need improvement. Some supervisors articulated that they would prefer a pass-fail grading scale for each domain, rather than the discriminatory *Expectations* scale with multiple passing categories. The use of pass-fail scales does not affect summative assessment outcomes but has been shown to improve student satisfaction, by engendering a more collaborative learning environment with reduced competitiveness between students.<sup>51</sup> Alternatively, clinically anchored rating scales with entrustment terminology (rather than *Expectations* scales) have been shown to improve both inter-rater reliability,<sup>44</sup> and discrimination between levels of learner performance.<sup>52</sup>

## Limitations

There are a few limitations to this study. Generalisability is limited as only supervisor perceptions from one medical school were sampled, although this included several different affiliated hospitals and GP practices. Other institutions may have different sign-off requirements, placement durations, rating scales, and assessor training methods. Triangulation through collaboration with different institutions would be beneficial, with sharing of best practice.

A limitation in study design using questionnaire data is that there is a lack of adaptability in data collection,<sup>53</sup> although this is somewhat mitigated by also collecting qualitative data. We also note the potential for volunteer bias in questionnaire responses.

## Contribution to Knowledge in This Area

We could not locate any studies in the existing literature which explore clinical supervisor perceptions of the undergraduate end-of-placement supervisor sign-off process from a qualitative perspective. This study adds rich data on stakeholder perceptions to explore why there might be limitations with the predictive validity of supervisors' end-of-placement clinical ratings of medical students.

## Conclusion

Hospital-based supervisors were less positive towards the sign-off process compared to GP supervisors, likely due to modern working practices of shared patient loads, shorter rotations, and reduced contact time with students.

This study can contribute to broad and meaningful change in long-term medical education policy beyond the local context, with rich qualitative data informing possible changes to the widespread but outdated clinical placement sign-off process for medical students. Differing sign-off processes could be considered in GP and hospital-based settings. In hospital settings, a multiple assessor model with embedded in-placement assessments could be considered. The assessment of practical skills should be removed from the end-of-placement clinical supervisor sign-off, and instead should be performed by clinical skills tutors, teaching fellows or resident doctors, who more frequently supervise students' performance of practical skills.

## Authors' information

PJM is an Endocrinology & Diabetes Specialist Registrar in North-West London. MN is a Urology Specialist Registrar in North Central and East London. JF is a Paediatric Specialist Registrar in London with a special interest in paediatric oncology. AR is Associate Professor in the Research Department for Medical Education at University College London Medical School. DH is a senior educationalist at the Royal College of Physicians, and an Honorary Medical Education Research Fellow at University College London Medical School. CAB is a medical education researcher and psychometrician. AHS is Head of Imperial College School of Medicine.

## Abbreviations

GP, General Practice; OSCE, Objective Structured Clinical Examination; SBAQ, Single Best Answer Question; SD, Standard Deviation; WPBA: Work-Place Based Assessment; UK, United Kingdom.

## Data Sharing Statement

Data that support the findings of this study are available from the corresponding author upon request.

## Ethics Approval and Consent to Participate

This study adhered to the Declaration of Helsinki. Informed consent was obtained from all participants in the study. Ethical approval was received from the Imperial College London Education Ethics Review Process (EERP2223-060).

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically

reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

There is no funding to report.

## Disclosure

The authors declare that they have no competing interests.

## References

- Kaufman DM. Teaching and learning in medical education: how theory can inform practice. In: Swanwick T, Forrest K, O'Brien BC, editors. *Understanding Medical Education: Evidence, Theory and Practice*. 3rd ed. Chichester: John Wiley & Sons; 2019:37–69.
- McGill DA, Van der Vleuten CP, Clarke MJ. Supervisor assessment of clinical and professional competence of medical trainees: a reliability study using workplace data and a focused analytical literature review. *Adv Health Sci Educ*. 2011;16:405–425. doi:10.1007/s10459-011-9296-1
- Keshavarzi MH, Azandehi SK, Koohestani HR, Baradaran HR, Hayat AA, Ghorbani AA. Exploration the role of a clinical supervisor to improve the professional skills of medical students: a content analysis study. *BMC Med Educ*. 2022;22(1):399. doi:10.1186/s12909-022-03473-w
- Pangaro L, Ten Cate O. Frameworks for learner assessment in medicine: AMEE Guide No. 78. *Med Teach*. 2013;35(6):e1197–210. doi:10.3109/0142159X.2013.788789
- Dudek NL, Marks MB, Wood TJ, Lee AC. Assessing the quality of supervisors' completed clinical evaluation reports. *Med Educ*. 2008;42(8):816–822. doi:10.1111/j.1365-2923.2008.03105.x
- Tavares W, Kinnear B, Schumacher DJ, Forte M. "Rater training" re-imagined for work-based assessment in medical education. *Adv Health Sci Educ*. 2023;28(5):1697–1709. doi:10.1007/s10459-023-10237-8
- Beigzadeh A, Adibi P, Bahaadinbeigy K, Yamani N. Strategies for teaching in clinical rounds: a systematic review of the literature. *J Res Med Sci*. 2019;24(1):33. doi:10.4103/jrms.JRMS\_460\_18
- Gormley GJ, Johnston J, Thomson C, Mcglade K. Awarding global grades in OSCEs: evaluation of a novel eLearning resource for OSCE examiners. *Med Teach*. 2012;34(7):587–589. doi:10.3109/0142159X.2012.682745
- Yeates P, O'Neill P, Mann K, Eva K. Seeing the same thing differently: mechanisms that contribute to assessor differences in directly-observed performance assessments. *Adv Health Sci Educ*. 2013;18(3):325–341. doi:10.1007/s10459-012-9372-1
- Lockspeiser TM, O'Sullivan P, Teherani A, Muller J. Understanding the experience of being taught by peers: the value of social and cognitive congruence. *Adv Health Sci Educ*. 2008;13(3):361–372. doi:10.1007/s10459-006-9049-8
- Holmboe ES, Ward DS, Reznick RK, et al. Faculty development in assessment: the missing link in competency-based medical education. *Acad Med*. 2011;86(4):460–467. doi:10.1097/ACM.0b013e31820cb2a7
- Reid CM, Kim DY, Mandel J, Smith A, Bansal V. Correlating surgical clerkship evaluations with performance on the National Board of Medical Examiners examination. *J Surg Res*. 2014;190(1):29–35. doi:10.1016/j.jss.2014.02.031
- Rassie K. The apprenticeship model of clinical medical education: time for structural change. *NZ Med J*. 2017;130(1461):66.
- McNair R, Griffiths L, Reid K, Sloan H. Medical students developing confidence and patient centredness in diverse clinical settings: a longitudinal survey study. *BMC Med Educ*. 2016;16:1–8. doi:10.1186/s12909-016-0689-y
- Watling CJ, Lingard L. Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process. *Adv Health Sci Educ*. 2012;17:183–194. doi:10.1007/s10459-010-9223-x
- Luu K, Sidhu R, Chadha NK, Eva KW. An exploration of "real time" assessments as a means to better understand preceptors' judgments of student performance. *Adv Health Sci Educ*. 2023;28(3):793–809. doi:10.1007/s10459-022-10189-5
- Bandiera GW, Morrison LJ, Regehr G. Predictive validity of the global assessment form used in a final-year undergraduate rotation in emergency medicine. *Acad Emerg Med*. 2002;9(9):889–895. doi:10.1111/j.1553-2712.2002.tb02189.x
- McGown PJ, Nichols MM, Forshaw JA, et al. What is the predictive validity of clinical placement sign-off forms for medical students? *BMC Med Educ*. 2025;25(1):840. doi:10.1186/s12909-025-07237-0
- Yu TC, Wheeler BR, Hill AG. Clinical supervisor evaluations during general surgery clerkships. *Med Teach*. 2011;33(9):e479–84. doi:10.3109/0142159X.2011.590558
- Goldstein SD, Lindeman B, Colbert-Getz J, et al. Faculty and resident evaluations of medical students on a surgery clerkship correlate poorly with standardized exam scores. *Am J Surg*. 2014;207(2):231–235. doi:10.1016/j.amjsurg.2013.10.008
- Kogan JR, Hatala R, Hauer KE, Holmboe E. Guidelines: the do's, don'ts and don't knows of direct observation of clinical skills in medical education. *Perspect Med Educ*. 2017;6(5):286–305. doi:10.1007/S40037-017-0376-7
- Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372.
- DaRosa DA, Niehaus AH, Markwell SJ. Assessment of a surgery clerkship's performance evaluation system. *Am J Surg*. 2000;179(2):145–149. doi:10.1016/S0002-9610(00)00249-X
- Malau-Aduli BS, Hays RB, D'Souza K, et al. Examiners' decision-making processes in observation-based clinical examinations. *Med Educ*. 2021;55(3):344–353. doi:10.1111/medu.14357
- Schuwirth LW, van der Vleuten CP. A history of assessment in medical education. *Adv Health Sci Educ*. 2020;25(5):1045–1056. doi:10.1007/s10459-020-10003-0
- Mazotti L, O'Brien B, Tong L, Hauer KE. Perceptions of evaluation in longitudinal versus traditional clerkships. *Med Educ*. 2011;45(5):464–470. doi:10.1111/j.1365-2923.2010.03904.x
- Krosnick JA, Presser S. Question and questionnaire design. In: Marsden PV, Wright JD, editors. *Handbook of Survey Research*. Bingley, UK: Emerald Group Publishing Ltd; 2010:263–314.

28. British Education Research Association: ethical guidance for education researchers; 2018. Available from: <https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018>. Accessed July 3, 2023.
29. QSR International Pty Ltd: NVivo (Version 12.7.0); 2018. Available from: <https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>. Accessed December 1, 2023.
30. Cohen L, Manion L, Morrison K. *Research Methods in Education*. Routledge; 2002.
31. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. *PLoS One*. 2020;15(5):e0232076. doi:10.1371/journal.pone.0232076
32. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? *Assess Eval Higher Educ*. 2008;33(3):301–314. doi:10.1080/02602930701293231
33. Fu G, Saunders G, Stevens J. Holm multiple correction for large-scale gene-shape association mapping. *BMC Genet*. 2014;1:1–8.
34. Torre DM, Schuwirth LW, Van der Vleuten CP. Theoretical considerations on programmatic assessment. *Med Teach*. 2020;42(2):213–220. doi:10.1080/0142159X.2019.1672863
35. Lee AS, Ross S. Continuity of supervision: does it mean what we think it means? *Med Educ*. 2021;55(4):448–454. doi:10.1111/medu.14378
36. Archer JC. State of the science in health professional education: effective feedback. *Med Educ*. 2010;44(1):101–108. doi:10.1111/j.1365-2923.2009.03546.x
37. Hoellein AR, Feddock CA, Wilson JF, Griffith CH III, Rudy DW, Caudill TS. Student involvement on teaching rounds. *Acad Med*. 2007;82(10):S19–21. doi:10.1097/ACM.0b013e31814004d7
38. Norcini J, Zaidi Z. Workplace Assessment. In: Swanwick T, Forrest K, O'Brien BC, editors. *Understanding Medical Education. Understanding Medical Education: Evidence, Theory, and Practice*. 3rd ed. Chichester: Wiley-Blackwell; 2019:1–6.
39. Schmidt HG, Moust JH. What makes a tutor effective? A structural equations modelling approach to learning in problem-based curricula. *Acad Med*. 1995;70(8):708–714. doi:10.1097/00001888-199508000-00015
40. Williams RG, Klamen DA, McGaghie WC. Cognitive, social and environmental sources of bias in clinical performance ratings. *Teach Learn Med*. 2003;15(4):270–292. doi:10.1207/S15328015TLM1504\_11
41. Viney R, Rich A, Needleman S, Griffin A, Woolf K. The validity of the annual review of competence progression: a qualitative interview study of the perceptions of junior doctors and their trainers. *J R Soc Med*. 2017;110(3):110–117. doi:10.1177/0141076817690713
42. Thistlethwaite JE, Jordan JJ. Patient-centred consultations: a comparison of student experience and understanding in two clinical environments. *Med Educ*. 1999;33(9):678–685. doi:10.1046/j.1365-2923.1999.00427.x
43. Farrell TM, Kohn GP, Owen SM, Meyers MO, Stewart RA, Meyer AA. Low correlation between subjective and objective measures of knowledge on surgery clerkships. *J Am Coll Surg*. 2010;210(5):680–683. doi:10.1016/j.jamcollsurg.2009.12.020
44. Prentice S, Benson J, Kirkpatrick E, Schuwirth L. Workplace-based assessments in postgraduate medical education: a hermeneutic review. *Med Educ*. 2020;54(11):981–992. doi:10.1111/medu.14221
45. Van Der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ*. 2005;39(3):309–317. doi:10.1111/j.1365-2929.2005.02094.x
46. 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. doi:10.1080/0142159X.2016.1215414
47. Heeneman S, Oudkerk Pool A, Schuwirth LW, van der Vleuten CP, Driessen EW. The impact of programmatic assessment on student learning: theory versus practice. *Med Educ*. 2015;49(5):487–498. doi:10.1111/medu.12645
48. Norcini J, McKinley DW. Concepts in assessment including standard setting. In: Dent JA, Harden RM, Hunt D, editors. *A Practical Guide for Medical Teachers*. 6th ed. Poland: Elsevier; 2021:271–278.
49. Tracy BM, Hazen BJ, Ward CR, Winer JH, Pettitt BJ. Sustained clinical performance during surgical rotations predicts NBME shelf exam outcomes. *J Surg Educ*. 2020;77(6):e116–20. doi:10.1016/j.jsurg.2020.06.033
50. Daniels VJ, Pugh D. Twelve tips for developing an OSCE that measures what you want. *Med Teach*. 2018;40(12):1208–1213. doi:10.1080/0142159X.2017.1390214
51. White CB, Fantone JC. Pass–fail grading: laying the foundation for self-regulated learning. *Adv Health Sci Educ*. 2010;15:469–477. doi:10.1007/s10459-009-9211-1
52. Halman S, Rekmán J, Wood T, Baird A, Gofton W, Dudek N. Avoid reinventing the wheel: implementation of the Ottawa Clinic Assessment Tool (OCAT) in Internal Medicine. *BMC Med Educ*. 2018;18:1–8. doi:10.1186/s12909-018-1327-7
53. Bell J, Waters S. *Doing Your Research Project: A Guide for First-Time Researchers*. 7 ed. London: Open University Press; 2014.
54. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S. PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Serv Res*. 2014;14(1):1. doi:10.1186/s12913-014-0579-0

## Advances in Medical Education and Practice

### Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>

**Dovepress**  
Taylor & Francis Group