



Four ways to get a grip on making robust decisions from workplace-based assessments

Quatre stratégies pour la prise de décisions solides fondées sur les évaluations en milieu de travail

Tim J Wilkinson

Volume 13, Number 3, 2022

URI: <https://id.erudit.org/iderudit/1091008ar>
DOI: <https://doi.org/10.36834/cmej.73361>

[See table of contents](#)

Publisher(s)

Canadian Medical Education Journal

ISSN

1923-1202 (digital)

[Explore this journal](#)

Cite this document

Wilkinson, T. (2022). Four ways to get a grip on making robust decisions from workplace-based assessments. *Canadian Medical Education Journal / Revue canadienne de l'éducation médicale*, 13(3), 43–46.
<https://doi.org/10.36834/cmej.73361>

Article abstract

Synthesising the results of workplace-based assessments to inform robust decisions is seen as both important and difficult. Concerns about failing to fail the trainee not ready to proceed has drawn disproportionate attention to assessors. This paper proposes a model for a more systems-based view so that the value of the assessor's judgement is incorporated while preserving the value and robustness of collective decision-making. Our experience has shown it can facilitate robust decisions on some of the more difficult areas, such as professionalism.

© Tim J Wilkinson, 2022



This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

<https://apropos.erudit.org/en/users/policy-on-use/>

Four ways to get a grip on making robust decisions from workplace-based assessments

Quatre stratégies pour la prise de décisions solides fondées sur les évaluations en milieu de travail

Tim J Wilkinson¹

¹University of Otago, Christchurch, New Zealand

Correspondence to: Professor Tim Wilkinson, P O Box 4345, Christchurch, New Zealand 8140; email: tim.wilkinson@otago.ac.nz

Published ahead of issue: April 26, 2022; published on July 6, 2022. CMEJ 2022, 13(3). Available at <https://doi.org/10.36834/cmej.73361>

© 2022 Wilkinson; licensee Synergies Partners. This is an Open Journal Systems article distributed under the terms of the Creative Commons Attribution License. (<https://creativecommons.org/licenses/by-nc-nd/4.0>) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited.

Abstract

Synthesising the results of workplace-based assessments to inform robust decisions is seen as both important and difficult. Concerns about failing to fail the trainee not ready to proceed has drawn disproportionate attention to assessors. This paper proposes a model for a more systems-based view so that the value of the assessor's judgement is incorporated while preserving the value and robustness of collective decision-making. Our experience has shown it can facilitate robust decisions on some of the more difficult areas, such as professionalism.

Résumé

Il est aussi important que difficile de synthétiser les résultats d'évaluations en milieu de travail pour fonder des décisions solides. Les préoccupations liées à la renonciation à faire échouer les stagiaires qui ne sont pas prêts à poursuivre leur formation ont focalisé l'attention sur les évaluateurs, et ce de manière disproportionnée. Cet article propose un modèle avec une vision plus systémique permettant d'intégrer la valeur du jugement de l'évaluateur tout en préservant la valeur et la solidité de la prise de décision collective. Notre expérience a montré qu'il peut faciliter la prise de décisions robustes en ce qui concerne des domaines plus difficiles comme le professionnalisme.

Introduction

Combining the results of workplace-based assessments (WBAs) to inform high stakes decisions is a goal for many educational programmes but one that can be difficult to realise.

WBAs fill an important gap in any programme-level assessment blueprint but seem more susceptible to failure to fail than point-in-time exams. A brief observation of a trainee in a workplace will not give a robust picture of that trainee's overall competence; the long case is a well-described example of this with reliability estimates of around 0.3-0.4.¹ This is not to say long cases have no value, but basing a high stakes decision on one episode alone is hazardous. Greater generalisability requires synthesis of several observations.¹ The problem of reliance on single observations also makes sense intuitively. Trainees

recognise this problem and realise that the result they get on such assessments may be as much due to serendipity and circumstance as it is to their own ability—such a perception is well backed up by data.^{1,2} It is this serendipity and circumstance that likely also contributes to trainees' high levels of anxiety because much of the result they obtain may be out of their control. Examiners also recognise this—they see a performance on a single observation but realise that performance may not actually reflect the trainee's overall competence. If the performance is below expectations, assessors often correctly feel some hesitancy in giving a fail.³ If assessors are uncertain about a trainee's performance and are then pushed to make a judgement, they will tend to give the benefit of the doubt in the trainee's favour.³ If such reluctance happens on several WBAs, then a series of substandard or borderline performances can all be

recorded as passes. As a result, trying to make anything other than a pass decision at the end of a year becomes indefensible. High stakes decisions are more trust-worthy if they are made by a group of people and after incorporating several observations.^{3,4}

This susceptibility to failure to fail illustrates the ‘black ice’ and highlights a paradox—the attributes that worry people the most and that get graduates into trouble (such as professionalism) are the attributes that seem to be less well represented on an assessment blueprint that focuses just on point-in-time examinations.⁵ A second paradox is that even if we do assess those important attributes, they are also possibly the things that are most susceptible to ‘failure to fail.’

Despite the necessity for WBAs, they also have problems: (1) they can be seen as a ‘tick box exercise’ that trainees somehow just have to do;⁶ (2) people hardly ever fail them;⁷ and (3) they are too subjective to reliably inform decisions.⁸ Oftentimes the blame for such shortcomings has been levelled at the assessor or supervisor whereas a shift to focussing on the system is needed.⁹

This paper presents reflections on some strategies used in the medical programme at Otago University in New Zealand to achieve robust high-stakes decisions from in-course workplace-based assessments. An unanticipated benefit was that the system works to highlight many aspects of professionalism so that now issues with professionalism carry the highest risk of trainees failing a year (odds ratio of 17.2).¹⁰ There are four key components to success.

1. Make expectations clear

Criterion referenced assessments based on pre-specified standards are now commonplace in health professional education. Nevertheless, having explicit expectations is a precursor to making judgements about whether a trainee has met those standards. This could be in the form of expected learning outcomes or codes of conduct.¹¹ The program must then communicate these to the trainees in multiple ways to ensure understanding.

2. Make space to convey supervisor uncertainty

We need a system whereby it becomes easy for examiners to convey their uncertainty while encouraging them to provide rich data to inform learning (and to inform decision making). At Otago, we tried introducing a ‘borderline’ category to address this but found that unworkable. This was because many (staff and students) still perceived ‘borderline’ as being equivalent to ‘fail’ and it was

therefore not used often. Even when used, a series of borderline results was usually insufficient to help make robust end of year decisions, largely because any accompanying narrative was not sufficiently informative.

As a result, we made a very simple change that had a surprisingly effective result: we introduced the category of ‘conditional pass’ (CP).¹⁰ See Box 1 for a worked example. This was usually determined based on a series of assessment episodes or WBAs. It could be used after single assessments but “unsure” or “insufficient data” could also be used for single observations. Put simply, if an observed performance or series of performances was not a clear pass and not clearly irremediable then we encourage the assessor or supervisor to give CP. However, in doing so, they must also provide the conditions needed for that trainee to pass.

3. Use future-focused language that aligns with assessor views

Invoking a language that is familiar to clinicians has been shown to be helpful in other areas of assessment¹². In cases of uncertainty, and when a CP is given, assessors are asked “what would the trainee need to do in order for you to be confident (or less uncertain) that they’re ready to progress?” Alternatively, “what would reassure you that the trainee is now up to standard?” Questions posed this way are future-focused and seem easier to answer than asking them to explain what the trainee did wrong, where the problem lies or how the deficit might best be filled. The answers the assessors give to these questions are usually easily converted into the conditions of a conditional pass. There are some similarities here with diagnostic uncertainty—if we are not sure about a patient diagnosis, we do not expect doctors to guess. Instead, we suggest they need to gather more information; and they need to be purposeful in the sort of information they should seek. Likewise, if we are uncertain about a trainee’s competence, we should not push the assessor to decide prematurely. Instead, we should ask what additional information they would need to make them less uncertain.

We know that feedback is more effective for learning when presented as a narrative rather than numbers.¹³ Secondly, the wording of uncertainty through ‘conditions to pass’ is like the wording of learning outcomes. This future-directed narrative contrasts with past-focused narratives that emphasize deficits. It also places more of the onus on the learner to show competence rather than a competence committee having to show incompetence. Learners must

show the conditions have been met thereby positively affecting learning. Most CPs are converted to pass in our institution – attesting to learning happening.

4. Facilitate shared decision-making.

Any trainee with a CP is reviewed at a progress meeting (held four times per year) where a group of staff check other indicators of that trainee's performance, determine whether the conditions are sufficiently clear and that the trainee has been informed. This process is like what competency committees do as part of competency-based medical education (CBME). Shared decision making has four benefits. Decisions by groups generally mitigate implicit biases of individuals.¹⁴ Secondly, knowing that these conditions are reviewed by others, and shared with any other concerns about that trainee, means that any high stakes decision no longer rests only on one assessor's shoulders. In turn, that creates a forum that makes it easier for assessors to express their concerns and uncertainties. Thirdly, peer review helps determine if the conditions are specific enough and if they are reasonable for the student's stage of learning, thus adding another component of defensibility. Finally, it contributes to supervisors having a shared mental model of what a satisfactory performance looks like.

Closing remarks

The contention is that 'conditional pass' works synergistically with WBAs to act on concerns, particularly those related to professionalism, and partially addresses the failure to fail problem. The automatic creation of a paper trail, being clear about future expectations, being clear about reasons for decisions and creating joint decision-making have all contributed to assessors being able to convey uncertainty which creates greater confidence, and defensibility in identifying students who are not yet ready to progress.

The system itself also promotes change. Regular meetings of staff to discuss CPs have been associated with a slow cultural shift in staff viewpoints from using assessments solely to judge students towards using them to help students. Such meetings are also surprisingly well attended attesting to staff perceiving them as a good use of their time.

The process has similarities to performance reviews of underperforming employees—fair process includes making the deficits known, creating clear goals, and then giving an opportunity for the employee to meet those goals. Failure

to improve then becomes a defensible basis to back up high-stakes decisions.

This list should not be regarded as a full programmatic assessment model or as a full CBME model. This is because the trainees who receive the most attention around setting and meeting personal learning goals are only those who are not clearly progressing as expected. When time and resources are short, one could argue that most time and energy should be devoted to the few trainees who give us the most concern, while also setting up the circumstances in which learning can occur.

Box 1: Worked example: abruptness in patient interactions

A trainee is offhand with patients on two occasions. Each episode on its own would not cause concern but a pattern may be emerging. Mitigating factors from the trainee's perspective are noted. The breaches are not judged to be sufficient to fail the trainee but if unchecked, could be seen as serious. The trainee is therefore given a CP, with the condition that they need to show evidence of reflecting the patient's concerns during the consultation. As the year progresses, and at subsequent progress meetings, the information related to the student's CP is noted. If the trainee responds to that feedback, they pass. If they do not, they fail. A fail decision is defensible as there is an automatic paper trail. Trainees must meet all conditions of a conditional pass before they are permitted to sit any end of year examinations. There is one exception—if the conditions that need to be met are likely to be assessed in the examinations, then the trainee is permitted to sit, as passing the exam would provide us with sufficient evidence that the conditions were met. An example here could be having a sufficient knowledge base. However, there are many conditions we are not confident an exam could or would pick up.

Conflicts of Interest: None to report.

Funding: None.

References

1. Wilkinson TJ, Campbell PJ, Judd SJ. Reliability of the long case. *Med Educ* 2008; 42(9): 887-93. <https://doi.org/10.1111/j.1365-2923.2008.03129.x>
2. Wilkinson TJ, D'Orsogna LJ, Nair BR, Judd SJ, Frampton CM. The reliability of long and short cases undertaken as practice for a summative examination. *Intern Med J* 2010; 40(8): 581-6. <https://doi.org/10.1111/j.1445-5994.2009.01970.x>
3. Tweed MJ, Thompson-Fawcett M, Wilkinson TJ. Decision-making bias in assessment: the effect of aggregating objective information and anecdote. *Med Teach*. 2013; 35(10): 832-7. <https://doi.org/10.3109/0142159X.2013.803062>
4. Schultz K, Risk A, Newton L, Snider N. Program foundations and beginning of concerns (part 1 of 3). When residents shouldn't become clinicians: getting a grip on fair and defensible processes for termination of training. *Can Med Ed J*. 2021; 12(4): 116-20. <https://doi.org/10.36834/cmej.70170>

5. Boulet JR, Durning SJ. What we measure ... and what we should measure in medical education. *Med Ed*. 2019; 53(1): 86-94. <https://doi.org/10.1111/medu.13652>
6. Bindal T, Wall D, Goodyear HM. Trainee doctors' views on workplace-based assessments: are they just a tick box exercise? *Med Teach*. 2011; 33(11): 919-27. <https://doi.org/10.3109/0142159X.2011.558140>
7. Massie J, Ali JM. Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Adv Health Sci Educ Theory Pract* 2016; 21(2): 455-73. <https://doi.org/10.1007/s10459-015-9614-0>
8. Hodges B. Assessment in the post-psychometric era: learning to love the subjective and collective. *Med Teach*. 2013; 35(7): 564-8. <https://doi.org/10.3109/0142159X.2013.789134>
9. Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med*. 2005; 80(10 Suppl): S84-7. <https://doi.org/10.1097/00001888-200510001-00023>
10. Wilkinson TJ, Tweed M, Egan T, et al. Joining the dots: Conditional pass and programmatic assessment enhances recognition of problems with professionalism and factors hampering student progress. *BMC Med Educ*. 2011; 11(1): 29. <https://doi.org/10.1186/1472-6920-11-29>
11. Anderson LC, Pickering NJ. The student code: ethical and professional expectations of medical students at the University of Otago. *New Zealand Med J*. 2010; 123(1318): 43-9.
12. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. *Med Educ* 2012; 46(1): 28-37. <https://doi.org/10.1111/j.1365-2923.2011.04166.x>
13. Watling C, Driessen E, van der Vleuten CPM, Lingard L. Learning culture and feedback: an international study of medical athletes and musicians. *Med Educ* 2014; 48(7): 713-23. <https://doi.org/10.1111/medu.12407>
14. Tweed M, Wilkinson T. Student progress decision-making in programmatic assessment: can we extrapolate from clinical decision-making and jury decision-making? *BMC Med Educ* 2019; 19(1): 176. <https://doi.org/10.1186/s12909-019-1583-1>