Author/s:
Reid, K;Pearce, J;Chiavaroli, N

Title:
Standard setting using programmatic thinking for small cohort performance-based exams

Date:
2021-05-01

Citation:
Reid, K., Pearce, J. & Chiavaroli, N. (2021). Standard setting using programmatic thinking for small cohort performance-based exams. Medical Education, 55 (5), pp.639-. https://doi.org/10.1111/medu.14516.

Persistent Link:
https://hdl.handle.net/11343/298324

DR. KATHARINE REID (Orcid ID : 0000-0002-4600-8847)

DR. JACOB PEARCE (Orcid ID : 0000-0003-3081-9132)

MR. NEVILLE CHIAVAROLI (Orcid ID : 0000-0003-1488-9747)

Article type : Really Good Stuff

Standard setting using programmatic thinking for small cohort performance-based exams

Reid, K.,[1] Pearce, J.[2], Chiavaroli, N.[2]

[1] The University of Melbourne and the Australian Council for Educational Research

[2] Australian Council for Educational Research

Corresponding author:

Dr Jacob Pearce

Senior Research Fellow

Tertiary Education (Assessment)

Australian Council for Educational Research

19 Prospect Hill Rd, Camberwell, Vic, Australia

Jacob.Pearce@acer.org

Standard setting using programmatic thinking for small cohort performance-based exams

What problem was addressed?

Setting standards for determining competent performance on examinations is well-established in medical education. Establishing a conceptual boundary that differentiates competent from non-competent test-takers typically involves either expert judgement based on test item characteristics (e.g. Angoff), or an empirical process based on examinee performance (e.g. Borderline regression).[1] Standard-setting

procedures are a mandatory component of assessment processes in Australian specialist medical training programs. However, small specialist medical Colleges cannot implement existing standard-setting methods for performance-based examinations. Angoff-type methods are often not suited for performance examinations and empirical methods are unsuitable for small cohorts. A reform of assessment processes undertaken by the Royal Australian College of Dental Surgeons (RACDS) Oral and Maxillofacial Surgery (OMS) final examinations required implementation of appropriate standard-setting processes; however, the chosen method had to work with small candidate numbers and small numbers of examiners to support the process.

What was tried?

Our standard-setting approach captured examiner judgements during the examinations, with standards embedded in a holistic rating scale. Every item in each examination format (short answer questions, clinical examinations, surgical anatomy vivas and OMS vivas) was blueprinted to two OMS curriculum domains (modules [15] and proficiencies [7]) creating 272 data points describing candidate performance. Borderline performance judgements were embedded in a six-point global rating scale (i.e. Significant Concerns, Clear Fail, Borderline Fail, Borderline Pass, Clear Pass, Outstanding). To support examiner judgements and enhance rater consistency, we developed comprehensive assessment rubrics for every item describing expected performance at all scale levels. This process also generated extensive data that could be visually displayed to inform candidate review using a programmatic decision-making approach. Data were represented in different ways to inform judgements, including candidate performance by module and proficiency, and frequency distributions of judgements across all items.

What lessons were learned?

Our standard-setting process was defensible, rigorous, and did not require additional resources. Clearly-defined standards were embedded in the rating scale supported by clear performance descriptors in the rubrics, with judgements applied during the assessment process. Examiners were easily able to apply the scale and were not required to meet independently to set standards. This was a valuable aspect of the standard-setting process when other methods require additional standard-setting panels or must apply statistical techniques to set standards following examinations. Embedding the standard-setting process within performance judgements was well-suited to the small numbers of examination candidates, where empirical methods for performance examinations cannot be applied. Our standard-setting process generated a rich data set to support the examiner committee to review candidate performance, providing valuable data to aggregate to inform pass/fail determinations: a key principle of programmatic assessment approaches. The committee could also undertake detailed review of candidate performance across multiple judgements against the standards. Examiners could consider each candidate's performance across all assessment items and separately for every curriculum domain and proficiency, to determine whether candidates met the standards against the curriculum. Such an approach has significant benefits in

supporting specialist medical Colleges to implement standard-setting processes that meet accreditation requirements and also promote more rigorous use of data to inform progress decisions.

[1] Cizek G.J. & Bunch, M. B. Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests. Thousand Oaks, CA: Sage Publications Ltd, 2007.