Scopus

Back

Standard setting for dental knowledge tests: reproducibility of the modified Angoff and Ebel method across judges

Ho, Ting Khee [a, b] ✉ ; Abu Kassim, Noor Lide [c] ; O'Malley, Lucy [a] ; Roudsari, Reza Vahid [a] ✉

[a] Division of Dentistry, School of Medical Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

Show all information

0

Citations 🔔

View PDF   Full text ⌄   Export ⌄   🔖 Save to list

Document   Impact   Cited by (0)   References (89)   Similar documents

Abstract

Introduction: Criterion-referenced standard setting methods establish passing scores based on predefined competency levels. The credibility of these scores must be supported by validity evidence. This study evaluated the reproducibility of modified Angoff and Ebel standards across different test formats and panels in dental assessments. Inter-rater reliability for each method was also assessed. Methods: Twelve judges, selected via purposive sampling, were divided into two equal groups representing various specialisms. Each panel applied modified Angoff and Ebel methods to set standards for one-best answer (OBA) and short answer question (SAQ) items. Method replicability across panels was assessed using the Mann–Whitney U-test to compare passing scores between Groups A and B. The Wilcoxon signed-rank test compared passing scores between modified Angoff and Ebel within groups. Inter-rater reliability was estimated using the intraclass correlation coefficient for modified Angoff and Fleiss' kappa for Ebel. Statistical analysis was conducted using IBM SPSS, with significance set at $p < 0.05$. Results: The median (IQR) years of teaching experience were 14.0 (17.0) for Group A judges and 21.5 (18.0) for Group B judges. In Group A, median (IQR) passing scores using modified Angoff were 49.75 (3.31) for OBA and 51.75 (6.13) for SAQ, with statistical no significant differences ($p > 0.05$) from Ebel OBA 47.38 (2.02), SAQ 49.50 (5.38). In Group B, modified Angoff passing scores were significantly higher than Ebel ($p < 0.05$): modified Angoff OBA 66.12 (3.31), SAQ 58.00 (7.50); Ebel OBA 55.92 (2.73), SAQ 49.50 (8.25). Passing scores were consistent across panels for SAQ but not for OBA. Inter-rater agreement, intraclass correlation coefficients (ICC) and Fleiss' kappa were higher in Group A across both methods. Conclusion: Reproducibility of modified Angoff and Ebel standards across panels was mixed. Passing scores were consistent across judges for SAQ but varied for OBA in both methods. Group A showed consistency between modified Angoff and Ebel standards, whereas Group B had differing passing scores between both standards. These findings should be carefully considered when establishing defensible and reliable passing standards for dental knowledge assessments. © The Author(s) 2025.

Author keywords

Angoff; Dental education; Ebel; Education measurement; Malaysia; Passing mark; Passing score; Reproducibility of results; Standard setting

Indexed keywords

MeSH

Clinical Competence; Education, Dental; Educational Measurement; Humans; Observer Variation; Reproducibility of Results