






RESEARCH

Open Access



Refining competency benchmarks: a scoping review of Angoff standard-setting in dental education

Galvin Sim Siang Lin^{1*} , Abdul Rauf Badrul Hisham^{2,3} , Muhammad Nazmi Abdul Majid⁴, Chan Choong Foong⁵ , Ting Khee Ho^{6,7}  and Lara T. Friedlander⁸ 

Abstract

Background Summative assessments in dental education are important for evaluating competencies and ensuring that graduates meet the required standards for independent practice and regulatory expectations through standard-setting. This scoping review aimed to appraise the current evidence on the use of the Angoff standard-setting method in dental education.

Methods This scoping review followed the framework developed by Arksey and O'Malley, incorporating five key stages: formulating research questions, identifying and selecting relevant studies, charting the data, and collating and summarising results. The review involved a systematic search of six databases (Google Scholar, Scopus, Web of Science, PubMed, Cochrane Library, and EMBASE) and two major dental education journals up to May 2024 using specific search terms, with duplicates removed via EndNote X9 software. Two investigators (GSSL & ARBH) independently applied the inclusion and exclusion criteria to select relevant studies published in English language, with a third investigator (MNAM) resolving any disagreements. Data extraction was performed independently by four investigators (GSSL, ARBH, MNAM, CCF) using a Google Spreadsheet, and findings were analysed through quantitative frequency and qualitative thematic analysis.

Results Out of the 1412 records, ten articles were selected after screening. The Angoff and Modified Angoff methods showed good reliability and pass rates, although the results varied between studies. Factors such as exam difficulty, panel-item interaction, age, and repeated test attempts might have influenced score variation, while inter-rater reliability remained strong across studies using Angoff methods.

Conclusion Although the Angoff and Modified Angoff methods provide flexible, expert-driven approaches to standard-setting, their effectiveness depends on contextual factors such as assessment design, panel training, and educational objectives. Dental educators should therefore tailor implementation to programme and assessment needs rather than adopting a one-size-fits-all approach.

Keywords Assessment, Dental education, Educational measurement, Health profession

*Correspondence:
Galvin Sim Siang Lin
galvin@iiium.edu.my

Full list of author information is available at the end of the article



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Summative assessment and evaluation of competence within dental education are essential to not only gauge the progress of aspiring oral health professionals but also play a fundamental role in shaping their future trajectories [1]. These high-stakes assessments serve as gatekeepers, ensuring that individuals who graduate from dental programmes meet the necessary standards to provide safe and effective oral healthcare. Unlike formative assessments, summative assessments are designed to evaluate whether candidates have acquired the competencies needed for independent dental practice and may also inform their suitability for advanced training [2, 3]. In this context, it is essential that the evaluation process is robust and transparent, with the methodologies employed for standard-setting being key to ensuring precision [4].

Standard-setting is an important step in summative assessments, involving the systematic establishment of a minimum acceptable performance level, known as the cut-off score [5, 6]. Most commonly, this is carried out by a panel of subject-matter-experts [7], and defines the boundary between success and failure. Importantly, at graduation, this process serves as a benchmark to ascertain whether candidates possess the necessary competencies for independent dental practice [4, 8]. Institutions and accrediting bodies set standards to maintain the quality and integrity of dental education, ensuring that professionals entering the field are adequately prepared for the challenges of modern dentistry and the delivery of safe care [9].

The calculation of the cut-off score depends on whether candidates are being evaluated relative to their peers or against the test content [6]. Two main categories of standard-setting methods exist: the norm-referenced or relative method, where cut-off scores are determined based on assessment results; and the criterion or absolute method, where the cut-off score is set independently of assessment results [10]. Norm-referenced methods compare test-takers or candidates to their peers, assessing a specific portion required to pass. However, this method focuses on a relative comparison of performance, and it is not widely used for high-stakes examinations [11]. On the other hand, criterion-referenced methods, such as the Angoff or Ebel methods, evaluate candidates based on their performance relative to the test content, considering item difficulty [12]. This type of assessment necessitates a standard-setting process to establish a valid cut-off score for the assessment [10]. Another category of standard-setting is the compromise method, including Hofstee and Hostee-Angoff methods. This method combines both norm- and criterion-referenced in standard-setting [13, 14].

Among the various methods available, the Angoff standard-setting method stands out as a widely utilised and respected technique in criterion-referenced assessments [15–17]. The Angoff method, developed by William H. Angoff in the mid-20th century [18], offers a systematic approach to determining the passing threshold in educational assessments. This method engages a panel of subject-matter experts who, through a thoughtful and rigorous process, estimate the probability that a minimally competent candidate would answer each test item correctly [16]. This collective judgment is then aggregated to establish the cut-off score, delineating the boundary between success and failure in the assessment. The strength of the Angoff method lies in its reliance on expert judgment [19], which considers the intricate and specific requirements of dental education.

Given the complexity and uniqueness of dental education [20], exploring how the Angoff method is applied within this specific domain becomes essential for assessing its suitability, challenges, and potential enhancements. Dental education is characterised by its multidimensional nature, encompassing theoretical knowledge, clinical skills, communication, ethics and professionalism [21]. The diverse nature of these competencies necessitates a thorough examination of the standard-setting methodologies to ensure that they align with the expectations of regulatory bodies and the profession [22]. While there is an existing body of research on standard-setting in general, the specific nuances and intricacies of the Angoff method in dental education warrant focused attention. This scoping review aims to appraise the current evidence on the use of the Angoff standard-setting method in dental education and contribute to the ongoing dialogue on refining assessment practices to better meet the needs of dental education programmes.

Methods

The review protocol was prospectively registered on the Open Science Framework (OSF) Preregistration platform (<https://osf.io/pwbnz>). The study design was guided by the scoping review framework developed by Arksey and O'Malley [23], which includes the following stages (Fig. 1): (1). formulating research questions; (2). identifying potentially relevant studies; (3). selecting relevant studies to be included; (4). charting the data; and (5) collating, summarising, and reporting results. The present scoping review also adhered to the reporting guidelines outlined in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Scoping Review extension (PRISMA-ScR) [24].

Stage 1: formulating research questions

The present scoping review aimed to answer the following research questions: (1). What is the currently

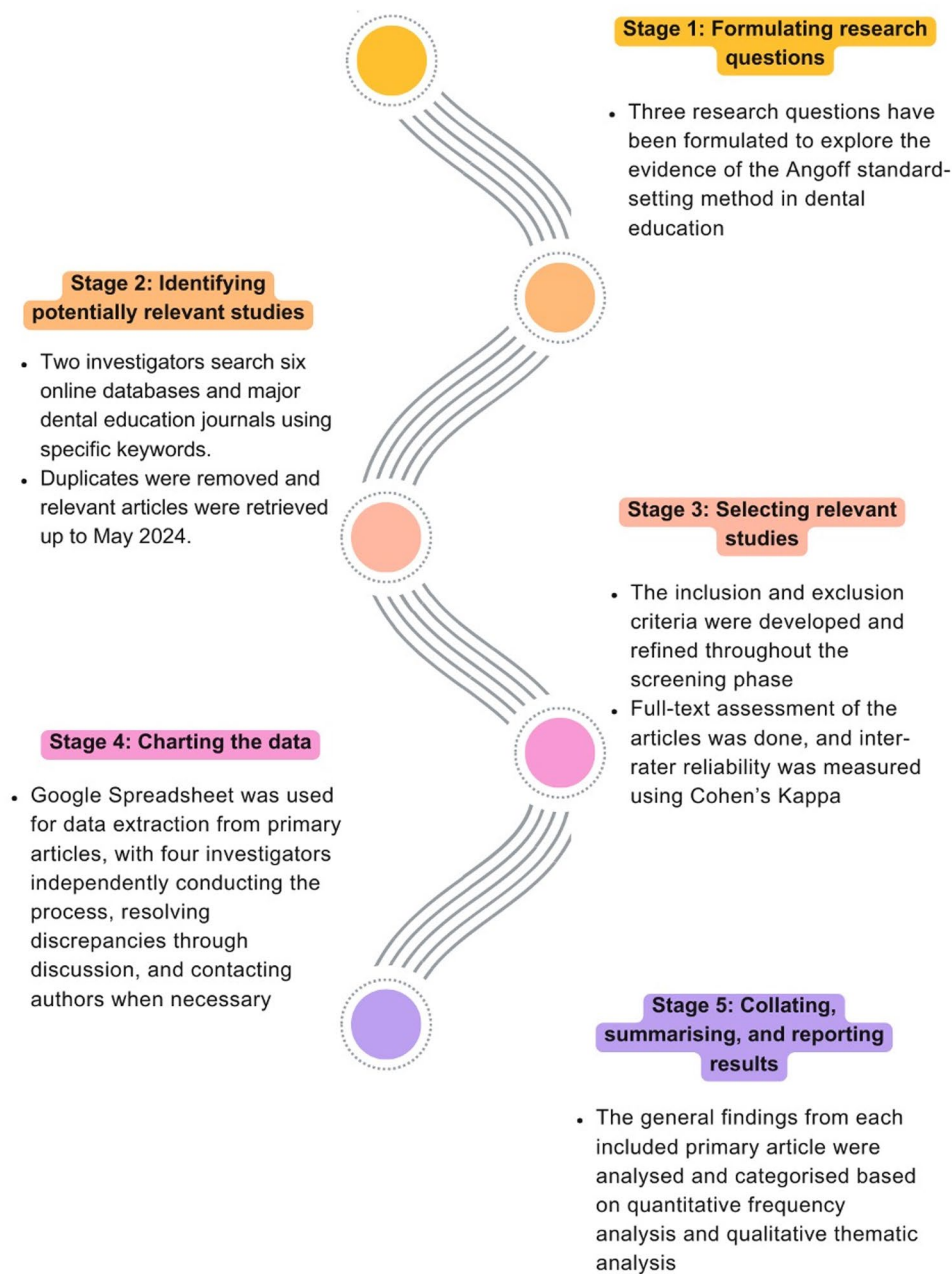


Fig. 1 The present scoping review framework

available evidence on the application of the Angoff standard-setting method in dental education? (2). How has the Angoff method been used to refine assessment practices in dental education? (3). Are there any challenges and potential improvements required in its implementation across dental educational programmes?

Stage 2: identifying potentially relevant studies

Two independent investigators (GSSL & ARBH) initially searched six online databases (Google Scholar, Scopus, Web of Science, PubMed, Cochrane Library,

and EMBASE) to identify relevant primary articles published up to May 2024 based on the titles and abstracts (Appendix 1). The search strategy also included the two major dental education journals, the European Journal of Dental Education and the Journal of Dental Education. The search strategy employed in each database included the following terms: (“Angoff” OR “Modified Angoff”) AND (“dental” OR “dentistry” OR “oral health” OR “oral care”). EndNote X9 software (Clarivate Analytics, Philadelphia, PA, USA) was used for reference management. Duplicated articles were also removed at this stage using

the EndNote X9 software. This approach ensured that articles relevant to Angoff or Modified Angoff methods in the context of dental education were systematically retrieved and reviewed.

Stage 3: selecting relevant studies

The inclusion and exclusion criteria were initially developed according to the research questions. These criteria were further refined and adjusted as necessary throughout the screening phase, ensuring that they effectively captured the relevant aspects of the primary articles under review. The inclusion criteria were: (1). Articles related to the use of Angoff or Modified Angoff method of standard-setting in dental education; (2). Involving either undergraduate or postgraduate dental students; (3). Published in the English language; (4). Randomised or non-randomised study designs, which include cross-sectional, cohort or case-control designs. Meanwhile, the exclusion criteria were: (1). Articles pertaining to standard-setting other than the Angoff or Modified Angoff method; (2). Involving students from health professional programmes outside dentistry; (3). Articles published in languages other than English; (4). Reviews, case reports, letters to the editor, commentaries, editorials, perspectives, and short communications; (5). Full-text unavailable. Two investigators (GSSL and ARBH) conducted independent evaluations of the full text of all primary articles based on the inclusion and exclusion criteria. Moreover, reference lists of the included primary articles were scrutinised to enhance the comprehensiveness of the search strategy. In cases of disagreement, a third investigator (MNAM) facilitated consensus. Inter-rater reliability was quantified using Cohen's Kappa coefficient to ensure the robustness of the selection process.

Stage 4: charting the data

A Google Spreadsheet data collection form was used to extract various details from the primary articles. These included publication information (authors, publication year, and country), study characteristics (study design, standard-setting methodology used), participant details (panels and student involvement), specific interventions and evaluation metrics (test items and overall findings). Four investigators (GSSL, ARBH, MNAM, CCF) independently performed data extraction from all included primary articles and any discrepancies were resolved through discussion among all investigators. The data extraction process was iterative, where investigators continuously added and edited columns on the spreadsheet as needed throughout the process. In the event that any primary study lacked essential information or if the information was not supplied, 'not available' or 'N/A' will be noted in the extraction form.

Stage 5: collating, summarising, and reporting results

The general findings from each included primary article were analysed and categorised based on quantitative frequency analysis and qualitative thematic analysis using NVivo 12 software. First, (GSSL and ARBH) carefully read through the data multiple times to familiarise themselves with it. Initial codes were identified and grouped into coherent axial codes, consolidating similar aspects of Angoff or Modified Angoff utilisation in dental education. Next, they categorised these axial codes into broader themes that encapsulated the main ideas and findings from the data analysis. Cross-checking agreements was conducted with the third investigator (CCF), and any disagreements were discussed with all investigators.

Results

Selected studies

An initial search of the online databases revealed a total of 1412 records. After duplicates were removed, 781 articles remained. 760 records were then excluded following titles and abstract screenings. The remaining 21 articles were assessed in full text, and only ten articles were selected for the present scoping review. During screening, the agreement between raters was substantial, with a Kappa score of 0.78 [25]. Fig. 2 illustrates the PRISMA-ScR flow diagram for identifying articles and reasons for article exclusion.

Study characteristics

The characteristics of each included primary study are summarised in Table 1. The present scoping review included ten primary studies conducted between 2009 and 2023. Most studies used retrospective designs, with only one being prospective [8]. Three studies originated from Saudi Arabia [26–28] and two from the United Kingdom [29, 30]. Meanwhile, there was one published study from Pakistan [31], Sweden [32], Malaysia [33], South Korea [34], and the Netherlands [8], respectively. The findings from these studies represent a wide spectrum of dental students, various dental programmes across the world, different dental disciplines, and types of assessments. The candidates included in the primary studies ranged from undergraduate dental students at different academic levels, from first year to final year [8, 26–31, 33, 34]. One study examined foreign-trained dentists taking the Swedish proficiency test for licensure [32].

Several studies gathered data over multiple years. Notably, Kim J et al. [34]. analysed data from two years of final competency examinations, while Yousef MK et al. [27]. examined preclinical operative dentistry across three years of different cohorts. The findings also provided insight into assessment in different dental disciplines, including operative dentistry, endodontics, prosthodontics, and paediatric dentistry [26–28]. The types of

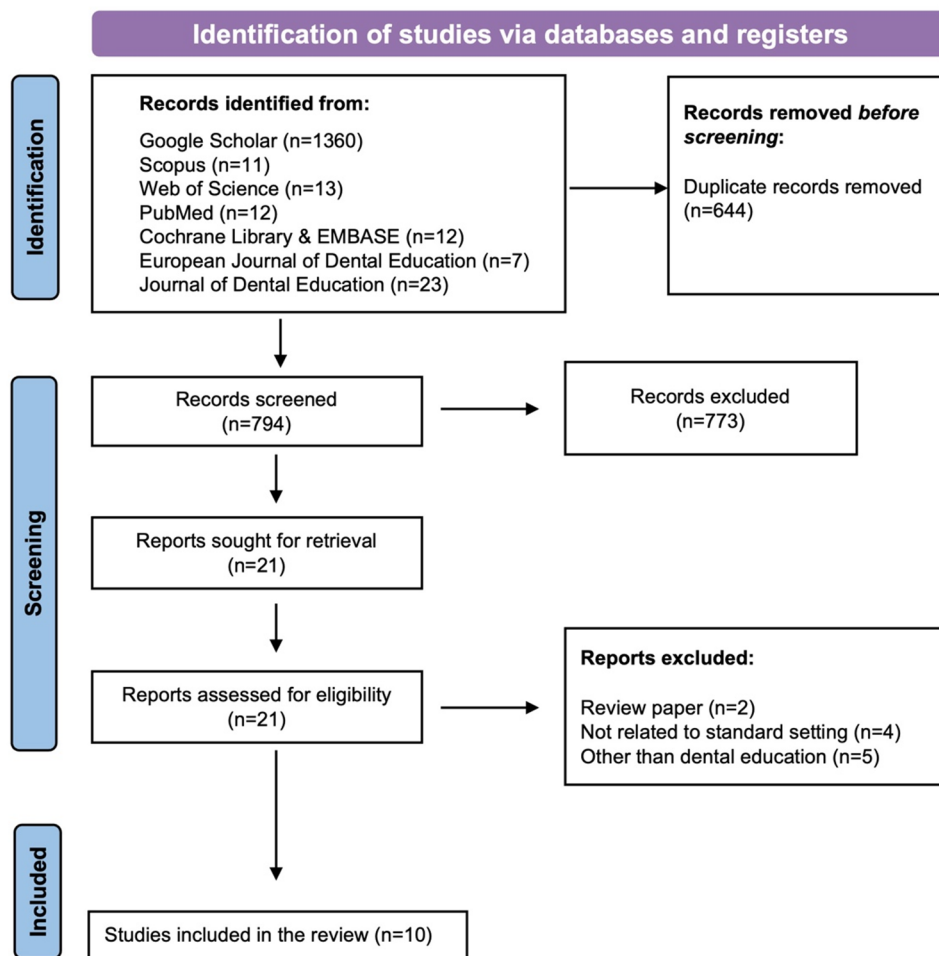


Fig. 2 PRISMA-ScR flow diagram of the present scoping review

assessments that were evaluated included clinical examinations (e.g., Objective Structured Clinical Examination or OSCE) [8, 29], and theoretical examinations (e.g., Multiple Choice Question or MCQ, Short Answer Question or SAQ) [26–28, 30–32]. There were also studies which presented findings from combinations of clinical and theoretical examinations [33], with MCQ assessment practices most common [26–28, 30, 33].

Angoff standard-setting comparisons

Two studies compared the Angoff method with other standard-setting methods. Khalid MN et al. [31], contrasted Angoff with relative/norm-referenced Hofstee and Cohen methods to determine cut-off scores for dental students completing written assessments. Another study applied both Angoff and Hofstee methods to establish cut-off scores for dental summative examinations, comparing these to the traditional fixed passing score of 60% [27]. Meanwhile, Ali K et al. [30], discussed using Angoff and Hofstee methods to evaluate students' level

of learning and progress within an undergraduate dental programme.

Four studies focused on the implementation and refinement of the Angoff method in dental education. For instance, Dalum J et al. [32], evaluated its effectiveness in the Swedish proficiency test for foreign dentists. Bahaman LA et al. [28], determined the cut-off scores of three final-year endodontic course examinations using the Angoff rating method. Kim J et al. [34], sought to improve the reliability of cut-off scores in dental education by comparing the Yes/No and percentage rating approaches used in Angoff's method. Similarly, Yousef MK et al. [26], evaluated Angoff's standard-setting method across five dentistry programme courses in the fifth academic year, using data from two consecutive student cohorts to determine appropriate cut-off scores.

The remaining three articles explored the use of a Modified Angoff method. Moreno-Lopez R et al. [29], compared the reliability of the Borderline Regression method and the Modified Angoff method for standard-setting in OSCEs, particularly in small sample sizes. Furthermore,

Table 1 The characteristics of each included primary study

Author (Year)	Type of study	Country	Standard-setting Method	Panels	Students	Test Items	general findings
Khalid MN et al. (2021)	Non-randomised experimental	Pakistan	Absolute (Angoff) vs. compromise (Hofstee and Cohen) vs. relative (Mean, SD) methods	Experienced panels received 2 days of training	UG: Year 1 (81); Year 3 (81); Year 5 (75)	SAQs	The Angoff method resulted in lower failure rates and higher pass rates in early and middle years. Cohen and Modified Cohen methods led to higher cut-off scores and lower pass rates. In exit exams, Angoff and relative methods had similar pass rates (81%–86.8%), while Cohen and Modified Cohen methods approached 100%. The Hofstee method had higher pass rates than Angoff but lower than Cohen, with Angoff and relative methods showing higher failure rates overall.
Dalum J et al. (2022)	Retrospective Cohort	Sweden	Absolute method (Modified Angoff) vs. compromise method (Hofstee and Cohen) vs. relative method (Mean, SD)	Dentists working as clinical supervisors and teachers. 10–15 experts for each test.	foreign-trained dentists taking the Swedish licensure exam	Theoretical examinations (Swedish proficiency test)	The study confirmed the reliability of the Modified Angoff method for setting pass marks, with strong agreement among panel members on item difficulty. Those who failed performed significantly worse in professional qualifications and dental subjects, demonstrating the test's effectiveness in distinguishing qualified candidates. Higher scores were associated with younger age and repeated attempts, while male participants scored slightly lower than females.
Moreno-López et al. (2022)	Retrospective	United Kingdom	Borderline Regression method vs. Modified Angoff method	N/A	dental students in years 2, 3, and 4	OSCEs	Borderline Regression method proved valid for standard-setting OSCEs when station quality was high, though a significant gap between it and Modified Angoff was noted. The Borderline Regression method is considered an acceptable alternative for small cohorts.
Abd-Rahman ANA et al. (2021)	Retrospective	Malaysia	conventional vs. norm-referenced vs. Modified Angoff	2 professors, 4 specialists from different disciplines, 2 clinical senior lecturers, 2 preclinical lecturers (mix of race, gender, and seniority)	40 dental students from the final year (5th year)	MCQs, OSCEs, SAQs	Passing rates and passing scores varied by method: the Modified Angoff method had a passing rate of 80% with a passing score of 54.6%, the norm-referenced method had a 62.5% passing rate with a score of 59.3%, and the absolute standard had a 100% passing rate at a 50% score. The Modified Angoff method demonstrated strong reliability.
Kim J & Yang JS (2020)	Retrospective	Korea	Yes/No vs. Percentage methods within the Modified Angoff	11–13 experts	68 and 65 students in 2017 and 2018, respectively	Final Competency Exam (PCE)	Increasing the number of items enhances the quality of cut-off scores. Meanwhile, the percentage method offers lower error and higher reliability than the Yes/No method, making it preferable. Consistent panel training across rounds and years enhances cut-off score reliability in any context.
Yousef MK et al. (2020)	Retrospective	Saudi Arabia	Yes/No Angoff method	4 panels for each dental discipline	5th-year students (two consecutive years)	MCQ	The Angoff method yielded significantly different cut-off scores compared to the fixed 60% passing mark used by the dental school, with a pass-fail rate differential of up to 41%.

Table 1 (continued)

Author (Year)	Type of study	Country	Standard-setting Method	Panels	Students	Test Items	general findings
Yousef MK et al. (2017)	Retrospective	Saudi Arabia	Angoff vs. Hofstee methods compared with fixed passing score of 60%	Four panels	3rd year students from 2012 to 2014	MCQ	The cut-off scores from both methods were consistently higher than the fixed 60% passing score. The Angoff method is preferred over the Hofstee method, and the Angoff method yielded a higher score (66.75 vs. 62.25 for Hofstee) in the third year.
Bahammam LA et al. (2017)	Retrospective	Saudi Arabia	Angoff method	6 qualified panels who have attended training for 1 year	5th-year UG students	MCQ	Judgment is key in setting standards. Though the Angoff method is easy to implement, it can be time-consuming due to frequent meetings. All cut-off scores must be validated as small deviations can greatly affect student outcomes.
Ali K et al. (2016)	Retrospective	United Kingdom	Angoff and Hofstee methods	N/A	All year 1–4 students	MCQ	Angoff method was used to assess question difficulty for borderline students and the Hofstee method to set pass/fail marks. Despite being resource-intensive, using these methods progress testing is a valid and reliable way to assess knowledge growth in dental education.
Schoonheim-Klein M et al. (2009)	Prospective	Netherlands	Angoff I vs. Modified Angoff II vs. Borderline Regression method	N/A	119 third year students	OSCE	The Borderline Regression method produced more favourable results than the two Angoff methods, with higher pass rates. Angoff I and II had pass rates of 86.6%, while the Borderline Regression method achieved 97.5%.

SAQ Single answer question, MCQ Multiple choice question, OSCE Objective structured clinical examination, PCE Final Competency Exam, UG undergraduate, N/A Not available

Schoonheim-Klein M et al. [8]. compared the Angoff, Modified Angoff with reality check, and Borderline Regression methods in an OSCE assessment to determine the optimal standard-setting method and differentiate between competent and incompetent dental students. On the other hand, Abd-Rahman ANA et al. [33]. compared conventional norm-referenced methods and the Modified Angoff standard-setting method by analysing the grades and pass/fail rates of final-year dental students across different assessment formats.

Panels

Most studies involved panels of 2 to 15 subject-matter experts from various specialities, including oral radiology, prosthodontics, periodontology, endodontology, oral medicine, and paediatric dentistry [26, 27, 30–34]. Some panels received training to familiarise themselves with the standard-setting process [28, 31]. For instance, Bahammam LA et al. [28]. involved four panels, two professors and two associate professors, who were trained in the Angoff method for a year. They found the method to be feasible and easy to implement, supporting its use for a fair assessment. However, the process was reported as time-consuming due to frequent meetings [28]. Most included studies mentioned that the panels were

experienced educators familiar with dental curricula, with one study highlighting the panels' diversity in race, gender, and seniority [33].

Comparison of Angoff standard-setting methods

The Angoff method resulted in lower cut-off scores, leading to higher pass rates and lower failure rates, especially in early and middle clinical years [31]. In contrast, the Cohen and Modified Cohen methods produced higher cut-off scores, while the Hofstee method showed variable results, with pass rates higher than Angoff but lower than the Cohen methods. In addition, the relative method produced similar results to Angoff, with comparable pass/fail rates across all years. One study reported that the Modified Angoff method led to a passing rate of 80%, with a cut-off score of 54.6%, allowing more students to pass [33]. Although it resulted in lower overall grades compared to the absolute standard method, the Modified Angoff method still showed higher overall grades than the norm-referenced method. Furthermore, the Angoff and Modified Angoff methods demonstrated similar pass/fail standards with no statistically significant differences, while the Borderline Regression method generally produced lower pass/fail standards [8].

Nonetheless, when evaluating the effectiveness of the Angoff method, there were contrasting findings, with one study revealing that when the Angoff method was used across different academic years, the cut-off scores were consistently higher than the commonly used arbitrary 60% benchmark [27]. Indeed, the Angoff method significantly affected student outcomes; for instance, a cut-off score of 74.31% would have led to a 41% failure rate in an endodontic examination [26]. Conversely, the Hofstee method encountered challenges in setting cut-off scores. This was due to the malalignment between the cumulative distribution curve and rater score boundaries, resulting in the use of the average minimal passing score as the cut-off [27]. Overall, when combined with the Hofstee method, the Angoff method proved effective in setting cut-off scores for progress testing in dental education [30].

Bahamman LA et al. [28]. noted the effective use of the Angoff method in setting cut-off scores for MCQ-based exams. Kim J et al. [34]. found that different response formats in the Angoff method (Yes/No vs. Percentage) significantly impacted pass rates, which ranged from 40.3% to 89.6%. On the other hand, Moreno-Lopez R et al. [29] observed discrepancies between the Angoff and Borderline Regression methods, suggesting potential concerns with the application of Angoff, particularly in small OSCE cohorts, possibly due to poor use of grading scales.

Reliability and validity of the Angoff method

The evaluation of the Angoff method showed good reliability between the panel of experts. Indeed, the interrater reliability for the Angoff standard-setting process varied from 0.65 to 0.88, indicating good agreement among the panels [32, 33]. Abd-Rahman ANA et al. [33]. reported excellent consistency, with an Intraclass Correlation Coefficient (ICC) and Cronbach's alpha of 0.94. Meanwhile, Dalum J et al. [32]. found a correlation coefficient ranging from 0.69 to 0.77. The high agreement between raters and the strong correlation with candidates' performance support the practicality and validity of the Modified Angoff method. However, two of the ten primary studies lacked specific details about the panels' contributions [8, 29]. Overall, discussions and consensus-building were essential to ensuring validity throughout the process [26].

The Angoff method also demonstrated strong correlations between predicted item difficulties and the expected and actual student performance [32]. Its reliability was enhanced with an increased number of panels [8]. Moreover, there was a significant correlation between the panels' estimates and the difficulty index of the items [28]. A previous study showed that the Angoff method consistently provided reliable cut-off scores reflective of expected performance levels [27]. However,

using percentage ratings in the Modified Angoff method produced higher reliability, with the ICC ranging from 0.48 to 0.65, whereas the 'Yes/No' option in the Angoff method had a maximum ICC of 0.41 [34].

Factors associated with score variation

Age and repeated test attempts have been identified as key factors affecting performance, while gender and the length of education had minimal influence [32]. Score variation across examinations was influenced by the difficulty level of each assessment [28]. Notable differences in cut-off scores were observed between subjects and academic years, with the complexity of courses shaping raters' perceptions of minimal competence, contributing to this variability [26]. Furthermore, panel-item interaction and item effect contributed significantly to the score variance [34].

Discussion

The present scoping review provides insights into the use of the Angoff standard-setting method in dental education. In health professional programmes, incorrect decisions regarding the pass or fail of borderline students can have detrimental effects on both the community and the candidate [35]. Hence, all assessment items need to be carefully reviewed and validated before administration to ensure that the cut-off score is reliable and justifiable for deciding the passing marks. The Angoff method's flexibility is one of the key factors contributing to its popularity. This method uses rigorous standards based on expert consensus on what qualifies as a minimum competent candidate, in contrast to norm-referenced methods, which compare students' results with one another [31, 33]. The Angoff method works especially well in educational settings where competency-based assessments are essential [36], which is often the case in dental schools.

The currently available literature shows mixed results regarding the effectiveness of the Angoff and Modified Angoff methods when compared to other standard-setting methods. While both the Angoff and Modified Angoff methods have been shown to produce lower cut-off scores, which in turn lead to higher pass rates [31, 33], this does not universally translate to their superiority over other methods. The traditional Angoff method, which involves panels estimating the probability that a minimally competent candidate would answer each test item correctly, has been widely accepted for its structured approach [37, 38]. However, its effectiveness varies depending on how it is applied and the specific exam conditions. For example, in early and middle clinical years [31], the Angoff method's lower cut-off scores can be beneficial in formative assessments, as the emphasis is on promoting student progression rather than imposing strict gatekeeping. On the other hand, the Modified

Angoff method, which incorporates adjustments like percentage scoring or reality checks among panels to modify their estimates [19, 39], offers more flexibility. It has also been reported that this modified version resulted in a high inter-panel agreement [33]. Nevertheless, when comparing the Angoff method with other methods such as Hofstee, Cohen, Modified Cohen, and Borderline Regression, the differences become even more apparent.

Beyond the methods identified in the included studies, other established criterion-referenced approaches include the Ebel, Nedelsky, and Bookmark methods. The Ebel method classifies items according to difficulty and relevance [40], while the Nedelsky method estimates the probability of success based on distractor elimination in MCQs [41]. The Bookmark method uses ordered test items to identify the transition point representing minimally competent performance [42]. These approaches illustrate that the choice of standard-setting method should be guided by assessment design, available psychometric data, and educational objectives rather than reliance on a single preferred approach.

Judgment is the most critical factor in establishing standards, as it relies on the standard-setting process as well as the panels' backgrounds and subject-matter expertise [43]. This allows them to determine more precise and convincing cut-off scores, as they possess a thorough comprehension of the knowledge and skills needed to demonstrate minimal proficiency [44]. Most included studies indicated that panels were experienced educators familiar with dental curricula, but the extent of their training in various standard-setting methods remains unclear. Nevertheless, panels' perceptions of question difficulty during standard setting are subjective, influenced by factors such as familiarity with the subject, the importance of the knowledge to candidate competency, and question complexity [14, 43, 44]. These perceptions can vary significantly among panels, depending on their training and expertise, the structure of the questions, and the assessment's context or purpose, leading to different interpretations of minimally acceptable performance or minimally competent (borderline) candidate.

The accuracy of the passing score depends on how effectively panels are trained to conceptualise a borderline candidate and define the characteristics of minimal competency [31]. Undeniably, panels frequently struggled to apply the concept of borderline competence due to its lack of a clear, practical definition [14, 28]. In addition, the standard-setting process can be time-consuming and is often complicated by concerns about confidentiality. Thus, through discussions and consensus-building, panels can mitigate biases, ensuring that the final cut-off score reflects a collective understanding of competency, with training further enhancing the method's effectiveness. It is also highly recommended that a

multi-disciplinary panel be used to reduce subjectivity in the standard-setting process [14], as it was discovered that panellists with varying specialisations established minimal variations in cut-off scores [33]. Meanwhile, the literature still offers an elusive answer regarding how many panellists are needed for standard-setting, with a suggested range of five to thirty [45, 46].

One of the key strengths of the Angoff method lies in its reliability, as highlighted by numerous studies reporting high inter-rater consistency [32, 33]. This could be explained through its structured rating process, as panels independently assess each item by estimating the probability that a minimally competent student would answer correctly. This method, coupled with consensus-building, ensures that the cut-off scores are based on informed [15], justifiable and transparent decisions rather than arbitrary choices. Angoff or Modified Angoff methods were found to result in strong inter-rater reliability, particularly when a sufficient number of panels are involved to account for rating differences [14, 32, 33]. Furthermore, the Angoff method's validity is supported by the strong correlation between the panels' estimated item difficulties and the actual student performance [28], indicating that the cut-off scores accurately represent students' real-world competencies. The reliability of Angoff panels plays a key role in ensuring valid interpretations of test scores, especially when making pass-or-fail decisions for students [19, 47]. While reliability measures the quality of the standard-setting process, it does not guarantee that the cut-off score is appropriate for its intended use, nor does it clarify the meaning of the score. To mitigate potential bias from individual panels, consensus on item judgments was achieved through regular meetings, during which extreme scores were reviewed and adjusted [28].

Several factors can introduce variability in cut-off scores using the Angoff or Modified Angoff method. One of them is the complexity of the assessment [28], which is particularly relevant in dental education, where the range of knowledge and practical skills assessed can differ. Another contributor to score variability is the response format used in the Angoff method. Research comparing formats, such as Yes/No versus Percentage rating, indicates that the latter improves reliability [34]. The percentage format allows panels to give more detailed assessments, resulting in more accurate cut-off scores, whereas the Yes/No format may oversimplify the evaluation, increasing variability [48]. This highlights the need to choose the appropriate format to provide consistent results. In addition, factors such as age and repeated assessment attempts can affect performance [32]. Younger students and those with multiple attempts often perform better, likely due to greater familiarity with the material or improved knowledge retention. This suggests

that cut-off scores may need to be adjusted to reflect the characteristics of the student group being assessed, as different populations may require varied competency thresholds. Furthermore, the inclusion of selected anchor questions on repeated assessments can help provide a baseline for equating analysis between cohorts [49], and for further informing the challenges associated with identifying the level of borderline candidates.

The contradictory findings reported in the present scoping review imply that neither the Angoff method nor the Modified Angoff method can be generally regarded as being more effective than other methods. Rather, the specific assessment context, such as the form of testing (written, clinical, or practical) [34], the size of the student cohort [29], and the goals of the assessment (e.g., formative vs. summative), ought to influence the choice of standard-setting method. Even while the Angoff method is well-established and flexible, it might not always be the best option. In addition, the present scoping review has notable implications for dental education, particularly in terms of assessment practices. Dental educators should use and carefully reflect on the use of standard-setting methods to establish valid and reliable benchmarks for student performance across a range of assessment formats.

The significance of consensus-building, training, and judging competence cannot be overstated because these elements are key to the method's reliability [27, 32, 38]. It is also important to meticulously evaluate both the type of assessment being evaluated and the response format selection to ensure that the cut-off scores are suitable for the intended use [14]. If the Angoff method is to be chosen, it should be used with flexibility owing to the variation in cut-off scores throughout different assessment types and student characteristics [15, 19]. Dental educators should customise their approach to the unique requirements of their programmes, assessments, and students' learning preferences as opposed to using a one-size-fits-all strategy. Depending on the assessment's complexity or the characteristics of the student cohort being evaluated, this might include altering the cut-off scores.

The present scoping review has several limitations. First, only articles published in English were included, which limits the representation of relevant studies in other languages. Second, the quality of the included articles was not assessed. Although only ten studies were reviewed, the authors chose to provide an overview of how the Angoff method has been applied in dental education, rather than focusing on the research design quality. Third, none of the included articles involved postgraduate students, meaning that the findings and discussion are limited to undergraduate dental education. Consequently, this review highlights that the application

of the Angoff method in postgraduate dental education remains underexplored.

Conclusion

Establishing standards and determining cut-off scores in summative assessment is crucial to distinguish competent candidates from incompetent ones. Both Angoff and Modified Angoff methods provide valuable tools for standard-setting in dental education, offering flexibility and expert-driven judgments to determine competency thresholds. However, the findings of this scoping review indicate that the effectiveness of these methods depends on several factors, including assessment format, panel composition and training, and characteristics of the student cohort. The diversity of dental disciplines and the complexity of knowledge and skills required in dental practice highlight the need for contextually aligned approaches to standard-setting. Overall, the available evidence demonstrates variability in the implementation and application of Angoff-based methods across dental education settings.

Abbreviations

ICC	Intraclass Correlation Coefficient
MCQ	Multiple-Choice Questions
OSCE	Objective Structured Clinical Examination
OSF	Open Science Framework
PRISMA-ScR	Preferred Reporting Items for Systematic Reviews and Meta-Analyses Scoping Review
SAQ	Short Answer Question

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12903-026-08195-9>.

Supplementary Material 1.

Acknowledgements

Not applicable.

Authors' contributions

GSSL contributed to study design, data collection, data analysis, research administration, and drafting of the article; both ARBH and MNAM contributed to data collection, data analysis and drafting of the article; both CCF and TKH contributed to data analysis, and article preparation; LTF contributed to data validation and review the manuscript. All authors read and approved the final version of the manuscript.

Funding

The current study received no funding.

Data availability

All data generated or analysed during this study are included in this published article.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Present address: Department of Restorative Dentistry, Kulliyah of Dentistry, International Islamic University Malaysia (IIUM), Kuantan Campus, Kuantan, Pahang 25200, Malaysia

²Centre of Paediatric Dentistry and Orthodontics Studies, Faculty of Dentistry, Universiti Teknologi MARA (UiTM), Sungai Buloh Campus, Sungai Buloh, Selangor 47000, Malaysia

³Department of Paediatric Dentistry & Dental Public Health, Kulliyah of Dentistry, International Islamic University Malaysia (IIUM), Kuantan Campus, Kuantan, Pahang 25200, Malaysia

⁴Dental Public Health Unit, School of Dental Sciences, Universiti Sains Malaysia, Health Campus, Kubang Kerian, Kelantan 16150, Malaysia

⁵Present address: Medical Education and Research Development Unit (MERDU), Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia

⁶Present address: Department of Restorative Dentistry, Faculty of Dentistry, Universiti Kebangsaan Malaysia, Kuala Lumpur 50300, Malaysia

⁷Division 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, UK

⁸Faculty of Dentistry, Sir John Walsh Research Institute, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand

Received: 14 October 2024 / Accepted: 20 March 2026

Published online: 27 March 2026

References

1. Tonni I, Gadbury-Amyot CC, Govaerts M, Ten Cate O, Davis J, Garcia LT, Valachovic RW. ADEA-ADEE Shaping the Future of Dental Education III: Assessment in competency-based dental education: Ways forward. *J Dent Educ*. 2020;84(1):97–104.
2. Terry R, Hing W, Orr R, Milne N. Do coursework summative assessments predict clinical performance? A systematic review. *BMC Med Educ*. 2017;17(1):40.
3. Khalaf KA, Moore C, McKenna G, Da Mata C, Lynch CD. Undergraduate teaching and assessment methods in prosthodontics curriculum: An international Delphi survey. *J Dent*. 2022;123:104207. <https://doi.org/10.1016/j.jdent.2022.104207>.
4. Puryer J, O'Sullivan D. An introduction to standard setting methods in dentistry. *Br Dent J*. 2015;219(7):355–8.
5. McLachlan JC, Robertson KA, Weller B, Sawdon M. An inexpensive retrospective standard setting method based on item facilities. *BMC Med Educ*. 2021;21(1):7.
6. De Champlain AF. Standard Setting Methods in Medical Education. In: *Understanding Medical Education*. edn.; 2018. p. 347–59.
7. Bertens LC, Broekhuizen BD, Naaktgeboren CA, Rutten FH, Hoes AW, van Mourik Y, Moons KG, Reitsma JB. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med*. 2013;10(10):e1001531.
8. Schoonheim-Klein M, Muijtjens A, Habets L, Manogue M, van der Vleuten C, van der Velden U. Who will pass the dental OSCE? Comparison of the Angoff and the borderline regression standard setting methods. *Eur J Dent Educ*. 2009;13(3):162–71.
9. Hobson R, Rolland S, Rotgans J, Schoonheim-Klein M, Best H, Chomyszyn-Gajewska M, Dymock D, Essop R, Hupp J, Kundzina R, et al. Quality assurance, benchmarking, assessment and mutual international recognition of qualifications. *Eur J Dent Educ*. 2008;12(Suppl 1):92–100.
10. Wadi MM. Standard Setting in Written Assessment. In: *Written Assessment in Medical Education*. edn.; 2023. p. 137–46.
11. Park SY, Lee SH, Kim MJ, Ji KH, Ryu JH. Comparing the cut score for the borderline group method and borderline regression method with norm-referenced standard setting in an objective structured clinical examination in medical school in Korea. *J Educ Eval Health Prof*. 2021;18:25.
12. Burkett T. Norm-Referenced Testing and Criterion-Referenced Testing. In: *The TESOL Encyclopedia of English Language Teaching*. edn.; 2018. p. 1–5.
13. Cohen-Schotanus J, van der Vleuten CPM. A standard setting method with the best performing students as point of reference: Practical and affordable. *Med Teach*. 2010;32(2):154–60.
14. Kamal D, ElAraby S, Kamel MH, Hosny S. Evaluation of Two Applied Methods for Standard Setting in Undergraduate Medical Programme at the Faculty of Medicine, Suez Canal University. *Educ Med J*. 2018;10(2):15–25.
15. Mubuuke AG, Mwisigwa C, Kiguli S. Implementing the Angoff method of standard setting using postgraduate students: Practical and affordable in resource-limited settings. *Afr J Health Prof Educ*. 2017;9(4):171–5.
16. Yousefi Afrashteh M. Comparison of the validity of bookmark and Angoff standard setting methods in medical performance tests. *BMC Med Educ*. 2021;21(1):1.
17. Ward H, Chiavaroli N, Fraser J, Mansfield K, Starmer D, Surmon L, Veysey M, O'Mara D. Standard setting in Australian medical schools. *BMC Med Educ*. 2018;18(1):80.
18. Angoff WH. Scales, norms, and equivalent scores. *Educational Testing Service*; 1984.
19. Shulruf B, Wilkinson T, Weller J, Jones P, Poole P. Insights into the Angoff method: results from a simulation study. *BMC Med Educ*. 2016;16:134.
20. Lin GSS, Lim TW, Braga MM. Embracing competency-based education for modern dental practice. *Asia Pac Scholar*. 2025;10(2):8–12.
21. Chuenjitwongsa S, Oliver R, Bullock AD. Competence, competency-based education, and undergraduate dental education: a discussion paper. *Eur J Dent Educ*. 2018;22(1):1–8.
22. Patel US, Tonni I, Gadbury-Amyot C, Van der Vleuten CPM, Escudier M. Assessment in a global context: An international perspective on dental education. *Eur J Dent Educ*. 2018;22(Suppl 1):21–7.
23. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19–32.
24. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med*. 2018;169(7):467–73.
25. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–82.
26. Yousef MK, Alshawwa LA, Farsi JM, Tekian A, Yoon Soo P. Determining defensible cut-off scores for dental courses. *Eur J Dent Educ*. 2020;24(2):186–92.
27. Yousef MK, Alshawwa L, Tekian A, Park YS. Challenging the arbitrary cutoff score of 60%: Standard setting evidence from preclinical Operative Dentistry course. *Med Teach*. 2017;39(sup1):S75–9.
28. Bahammam LA. Cutoff score evaluation of undergraduate dental curriculum. *Med Teach*. 2017;39(sup1):S33–6.
29. Moreno-Lopez R, Hope D. Can borderline regression method be used to standard set OSCEs in small cohorts? *Eur J Dent Educ*. 2022;26(4):686–91.
30. Ali K, Coombes L, Kay E, Tredwin C, Jones G, Ricketts C, Bennett J. Progress testing in undergraduate dental education: the Peninsula experience and future opportunities. *Eur J Dent Educ*. 2016;20(3):129–34.
31. Khalid MN, Shafiq F, Ahmed S. A Comparison of Standard Setting Methods for Setting Cut-Scores for Assessments with Constructed Response Questions. *Pakistan J Educational Res Evaluation (PJERE)*. 2022;9(2):65–75.
32. Dalum J, Christidis N, Myrberg IH, Karlgren K, Leanderson C, Englund GS. Are we passing the acceptable? Standard setting of theoretical proficiency tests for foreign-trained dentists. *Eur J Dent Educ*. 2023;27(3):640–9.
33. Abd-Rahman ANA, Baharuddin IH, Abu-Hassan MI, Davies SJ. A comparison of different standard-setting methods for professional qualifying dental examination. *J Dent Educ*. 2021;85(7):1210–6.
34. Kim J, Yang JS. How to improve reliability of cut-off scores in dental competency exam: A comparison of rating methods in standard setting. *Eur J Dent Educ*. 2020;24(4):734–40.
35. Lane AS, Roberts C, Khanna P. Do We Know Who the Person With the Borderline Score is, in Standard-Setting and Decision-Making. *Health Professions Educ*. 2020;6(4):617–25.
36. Dwyer T, Wright S, Kulasegaram KM, Theodoropoulos J, Chahal J, Wasserstein D, Ringsted C, Hodges B, Ogilvie-Harris D. How to set the bar in competency-based medical education: standard setting after an Objective Structured Clinical Examination (OSCE). *BMC Med Educ*. 2016;16:1.
37. Rezigalla A. Angoff's method: The impact of raters' selection. *Saudi J Med Med Sci*. 2015;3(3):220.
38. Hurtz GM, Auerbach MA. A Meta-Analysis of the Effects of Modifications to the Angoff Method on Cutoff Scores and Judgment Consensus. *Educ Psychol Meas*. 2003;63(4):584–601.
39. Yim M. Comparison of results between modified-Angoff and bookmark methods for estimating cut score of the Korean medical licensing examination. *Korean J Med Educ*. 2018;30(4):347–57.

40. Bourque J, Skinner H, Dupre J, Bacchus M, Ainslie M, Ma IWY, Cole G. Performance of the Ebel standard-setting method in spring 2019 Royal College of Physicians and Surgeons of Canada internal medicine certification examination consisted of multiple-choice questions. *J Educ Eval Health Prof.* 2020;17:12.
41. Melican GJ, Mills CN, Plake BS. Accuracy of Item Performance Predictions Based on the Nedelsky Standard Setting Method. *Educ Psychol Meas.* 1989;49(2):467–78.
42. Karantonis A, Sireci SG. The Bookmark Standard-Setting Method: A Literature Review. *Educational Measurement: Issues Pract.* 2006;25(1):4–12.
43. Norcini JJ, Shea JA, Kanya DT. The Effect of Various Factors on Standard Setting. *J Educ Meas.* 2005;25(1):57–65.
44. Burr SA, Zahra D, Cookson J, Salih VM, Gabe-Thomas E, Robinson IM. Angoff anchor statements: setting a flawed gold standard? *MedEdPublish* (2016). 2017;6:167.
45. Norcini JJ, Shea JA. The Credibility and Comparability of Standards. *Appl Measur Educ.* 1997;10(1):39–59.
46. Hurtz GM, Hertz NR. How Many Raters Should be Used for Establishing Cutoff Scores with the Angoff Method? a Generalizability Theory Study. *Educ Psychol Meas.* 1999;59(6):885–97.
47. Kane M. Validating the Performance Standards Associated With Passing Scores. *Rev Educ Res.* 1994;64(3):425–61.
48. Park J. Possibility of using the yes/no Angoff method as a substitute for the percent Angoff method for estimating the cutoff score of the Korean Medical Licensing Examination: a simulation study. *J Educ Eval Health Prof.* 2022;19:23.
49. Combrinck C, Scherman V, Maree D. Evaluating anchor items and reframing assessment results through a practical application of the Rasch Measurement Model. *South Afr J Psychol.* 2016;47(3):316–29.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.