


Are we passing the acceptable? Standard setting of theoretical proficiency tests for foreign-trained dentists

Jesper Dalum¹  | Nikolaos Christidis¹ | Ida Hed Myrberg² | Klas Karlgren^{3,4,5} | Charlotte Leanderson⁶ | Gunilla Sandborgh Englund¹

¹Department of Dental Medicine, Division of Oral Diagnostics and Rehabilitation, Karolinska Institutet, Huddinge, Sweden

²Institute of Environmental Medicine, Division of Biostatistics, Karolinska Institutet, Huddinge, Sweden

³Department Learning, Informatics, Management and Ethics, Karolinska Institutet, Huddinge, Sweden

⁴Department of Research, Education, Development and Innovation, Education Center, Södersjukhuset, Stockholm, Sweden

⁵Faculty of Health and Social Sciences, Western Norway University of Applied Sciences, Bergen, Norway

⁶Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Huddinge, Sweden

Correspondence

Jesper Dalum, Department of Dental Medicine, Karolinska Institutet, Box 4064, SE141 04 Huddinge, Sweden.
Email: jesper.dalum@ki.se

Funding information

National Board of Health and Welfare

Abstract

Introduction: Dentists educated outside the EU, EEA and Switzerland have the option to undergo proficiency testing to practice dentistry in Sweden. A standardised evidence-based procedure for proficiency tests is crucial. This paper explores if the standard setting of the theoretical examination in the Swedish proficiency test for foreign dentists is fit to differentiate candidates who meet the acceptable standard for licensure from those who do not, by analysis of inter-rater reliability and credibility of Angoff panels and the test's ability to differentiate candidates regarding learning outcomes and dental disciplines.

Materials and Methods: In a cohort study design, test results were collected from seven examinations during 2018–2019 ($n = 316$), combined with Angoff judgements and demographic data from The National Board of Health and Welfare. Intraclass correlation, Pearson correlation and linear regression were calculated.

Results: All exams demonstrated substantial judge inter-rater reliability using the Angoff procedure (0.65–0.82). Item difficulty predicted by the expert panels showed significant positive correlations with the item difficulty by candidate results (0.69–0.77). Candidates who did not meet the requirements for pass mark in the theoretical test revealed a significantly reduced performance overall in professional qualifications (i.e., learning outcomes) and reduced performance within dental disciplines.

Conclusions: The results indicate that a modified Angoff method is a feasible, simple, and defensible method for setting the pass mark for the theoretical examinations. The standard setting appears to differentiate candidates who do not meet the requirements of the professional qualifications of the dental degree. Adjustments could be undertaken to further discriminate for competence within single dental disciplines.

KEYWORDS

assessment, dentists, educational measurement, foreign-trained dentists, licensure and certification, standard setting

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *European Journal of Dental Education* published by John Wiley & Sons Ltd.

1 | BACKGROUND

International migration of dentists is not a new phenomenon¹ and many countries have implemented a process of recognition for foreign-trained dentists to be able to work as dentists within the respective country.²⁻⁴ The primary purpose of proficiency testing is to secure patient safety, and the focus is to differentiate between dentists who meet the requirements for licensure from those who do not.⁵ To promote mobility and employability within the European Union, dentists educated within the European Higher Education Area⁶ are recognised⁷ through the Bologna process^{8,9} in all 49 included countries. To further strengthen the homogeneity of the curricula in the European dental schools, the Association of Dental Education in Europe has developed a framework for the graduating European dentist.^{10,11} Thus, to acquire a licence as a registered dentist in Sweden, an equivalent of the Swedish Degree of Master of Science in Dental Surgery¹² or the requirements of recertification¹³ must be met, both harmonising with the domains of ADEE framework.¹⁰

However, dentists educated outside the European Higher Education Area (EHEA) must go through a recertification process in the respective country. In 2017 the Swedish National Board of Health and Welfare (NBHW) redesigned the process of recognition for dental practitioners educated outside EU, EEA and Switzerland, henceforth called foreign-trained dentists, to obtain a licence through proficiency tests.^{13,14} The Swedish proficiency test for foreign-trained dentists comprises several examination methods.

As all aspects of the professional qualifications cannot be tested using a single assessment method,⁵ these are tested with multiple methods with principles of assessment recommended by the ADEE.^{10,15} The route to licensure through proficiency test includes three parts (Figure 1), in which part I and part II are performed at the Department of (Blinded) Sweden. After having passed part I and part II at (Blinded), participants undergo a 6-month clinical training period at a dental clinic,¹⁴ aimed to evaluate professional practice. The Table S1 provides the professional qualifications of the degree of dental surgery mapped against the stages within NBHW's route for licensure. The Swedish proficiency test procedure also encompasses a mandatory course in Swedish laws and regulations and certification of language skills.

The theoretical part of the proficiency test is a computer-based examination. The test examines knowledge and application of knowledge, i.e., cognitive skills, as described by Miller.¹⁶ Question types primarily consist of single best answer questions. In addition, short-essay questions are used to further examine, e.g., reasoning skills and patient management.¹⁶ Since the introduction of the new proficiency test, post-hoc analysis paired with valuable feedback from the test takers has led to revisions regarding the number of questions. The number of questions in the theoretical examination ranged from 83 to 91. The test is constructed by application of a task-specific test blueprint by a committee in collaboration with subject representatives, to preserve authenticity and relevance of test items. The validity and reliability of the theoretical examination depends fundamentally upon a systematical and wide sampling.

DEGREE of DENTAL SURGERY	PROFESSIONAL QUALIFICATIONS (ILOs)											
	1 2 3 6 7 8 9 11 13 14				3 6 7 8 13 14				4 5 10 11 12 14 15			
DOMAIN of COMPETENCE	Knowledge and application of knowledge				Clinical and Communication skills				Professional practice			
MILLER'S PYRAMID	KNOWS		KNOW HOW		SHOW HOW				DOES			
EXAMINATIONS	PART I Theoretical examination				PART II Pre-clinical skills				PART II Workplace based assessment			
ASSESSMENT	Computer-based test (SBA,VSA)				OSCE Dental manikin				Six month clinical practice			

FIGURE 1 Assessment blueprint mapping the professional qualifications of the Degree of Dental surgery to domains of competence, the levels of Miller's pyramid, as well as examination and assessment categories. Abbreviations: ILOs, Intended Learning Outcomes; OSCE, Objective Structured Clinical Examination; SBA, Single Best Answer questions; VSA, Very Short Essay questions

Items in the test are generated by application of a blueprint based on the intended learning outcomes of the Qualifications Ordinance of the Swedish Degree of Master of Science in Dental Surgery¹² and competency domains. As the aim of the degree is intended to set a standard for a profession, intended learning outcomes are specified in the degree as professional qualifications, addressed as professional qualifications in this paper. To secure that all professional qualifications are covered within the proficiency test route a main assessment blueprint was created, presented in Figure 1. The blueprint maps learning outcomes and is used to allocate outcomes of the theoretical examinations to the blueprint. In total 10 intended learning outcomes are assessed within the theoretical test. Blueprint for theoretical exams also contains subject domains of the curriculum of a Swedish dental education. The sampling process and the respective representation of professional qualifications/subject domains in the theoretical test is based on the dimensioning within the curriculum of a Swedish dental education.

1.1 | Standard setting

The purpose of the standard setting is to set a standard criterion (pass mark) that determines whether a participant fulfils the set criteria or not (i.e., the test renders a pass or fail mark). The reference to criteria ensures that only those demonstrating essential knowledge and skills are to succeed in the test and is used to standardise the examination procedure. The standard setting for the theoretical examination is determined by (a) a criterion-referenced, modified method of Angoff and (b) a minimum of 50% of possible points of all questions mapped to each learning outcome.

The Angoff method uses a panel of subject-matter experts to set a criterion or an absolute standard that predicts item and examination difficulty.¹⁷ Hence, it is a commonly used method to set a defensible pass mark in medical education and in high stakes standardised tests.^{18–20} In attempts to improve the agreement amongst panel members several modifications have been made to the original Angoff method.^{19,21,22} The modifications often include a set number of iterative processes, i.e., a discussion amongst panel members to increase agreement and reduce variability.^{21,23,24} The modifications can also include panel members presented with data of the panel judgements on specific items to improve inter-judge agreement.^{23–26} The modifications made in our modified Angoff are rational decisions essential to present a feasible workload on top of other faculty obligations. In the original Angoff method panel members are asked to estimate the proportion of 100 minimally competent examinees who would correctly answer a test item. Although the Angoff method is widely used in medical education many versions exist with reliability and validity of the method relying heavily on the procedure of implementation.

In the modified Angoff method used in this study, panel members are asked to determine the probability of a single (1) minimally competent candidate correctly answering an item. The rationale is that it is easier to think of a single person than one hundred candidates

and to make a simple yes/no decision.²⁷ The two methods produce essentially the same pass mark.²² Further, iterations and presentations of other panel members' estimations are not included within the process.

For each test, 10–15 subject-matter experts are recruited to predict how a minimally qualified candidate would answer the test items. The panels comprised of dentists working as clinical supervisors and teachers at (Blinded). The panel receive oral and written instructions, and the evaluators independently scored each item as yes (1), maybe (0.5), or no (0). The scores of all raters are averaged to a percentage on each item and multiplied with item score to determine the pass mark.

Considerable efforts have been put into setting up examinations within the proficiency route to licensure. Hence, it is of importance to evaluate if the method for setting pass mark is reliable and that candidates with passing grades in fact possess the competence in alignment with requirements for minimal passing level, and to rule out if the tests falsely fail to acknowledge acceptable dentists. Such evaluations have not previously been performed.

This paper aims to explore if the current standard setting of the theoretical examination is fit to differentiate between foreign-trained dentists who meet the acceptable standard for licensure from those who do not. We do so by analysing the degree of inter-rater reliability and credibility of Angoff panels used for setting pass mark. We also assess the test's ability to differentiate acceptable from not acceptable candidates based on their performance related to intended learning outcomes and different dental disciplines.

2 | METHODS

In this cohort study, test results were collected from all (seven) theoretical examinations during 2018–2019, including 316 participants taking the Swedish proficiency test for dentists educated outside EU, EEA and Switzerland. All tests and results were included, with consecutive sampling from all dentists taking their first theoretical test during 2018–2019. Test results were combined with Angoff judgements and background data from public documents of the National Board of Health and Welfare (documentation of the candidates' assessed education). Test data of individual scores, the number of attempts and subject area scores were extracted from the computer-based assessment platform Inpera Assessment® (Inpera AS). The study was approved by the Swedish Ethical Review Authority (Dnr: Blinded).

2.1 | Background variables

The background variables include country of education and length of education collected from public documents of the National Board of Health and Welfare's decisions on eligibility based on previous education. Data regarding age and gender were collected from participants' applications for the theoretical tests. The region of education,

i.e., country, was presented according to the M49 standard for statistical use.²⁸

2.2 | Professional qualifications

To ensure that a pass mark also corresponds to acceptable levels of the respective individual learning outcomes, the participants need to acquire 50% of the points of each learning outcome. The percentages of achieved scores *per* professional qualification were extracted from Inspira Assessment[®] for all individuals in each individual test.

2.3 | Disciplines of dentistry

Eight dental disciplines were included and analysed: *cariology, oral radiology, prosthodontics, periodontology, endodontology, oral medicine and paediatric dentistry*. The percentages of achieved scores *per* dental discipline were extracted from Inspira Assessment[®] for all individuals in each individual test.

2.4 | Definitions of passing and failing the test

Candidates' results are presented as pass/fail grades or a percentage of the pass mark. The *percentage of pass mark* was extracted from Inspira Assessment[®] for each participant in each individual test. The participants' number of tests taken sequentially is referred to as *Test number*.

2.5 | Inter-rater reliability and credibility

To assess the degree of reliability and consistency amongst Angoff raters, we performed tests of inter-rater reliability. The correlation between the difficulty predicted by the Angoff panel and the actual item score from participants offers a credibility measurement of the accuracy of the panel ratings. A kappa score was determined for each panel and exam. The test gives an index of the consistency of raters, on a scale of 0.0 to 1.0, where 1.0 indicates perfect consistency.²⁹

2.6 | Statistical methods

Descriptive statistics were used to describe baseline data. Inter-rater agreement amongst Angoff panel members was determined using intraclass correlation coefficient, ICC (2, *k*), with multiple raters based on a mean rating, absolute agreement, and 2-way random effects model as defined by Shrout and Fleiss.³⁰ Landis and Koch's rating of kappa reliability²⁹ for standards of strength was used for understanding the calculated kappa coefficients (agreement). The credibility of the modified Angoff procedure standard setting for each examination was evaluated using the Pearson correlation

coefficient. The outcome variables in linear regression were percentage of pass mark, percentage of achieved score per dental discipline, and percentage of achieved score per professional qualification. The corresponding independent variables were the background variables described above. Considering that some participants have done the test more than once, uni- and multivariable linear regression with cluster robust standard errors was used to evaluate associations between demographic factors and outcome variables. *p*-values of <.05 were considered statistically significant. The entire cohort was included, and no sample population was selected, thus calculation of sample size was found non-applicable for the setting. R version 4.0.1 (Team RC, Vienna, Austria)³¹ was used for all statistical analyses.

3 | RESULTS

3.1 | Demographics

Most participants were younger than 40 years of age (median age 34; min-max age 24–62) with a fair gender distribution (47% female, 53% male). Almost all participants had graduated from educations with an educational length of 5 years or more (93%) with a median of 9 years since graduation. Over half of the participants were educated in Western Asia (60%). For a detailed description of distribution of pass and fail, see [Table 1](#). For every additional year of age, the achieved score decreased statistically significantly (-0.29 ; $p = .013$) in the multivariable analysis. Age as a significant factor was also followed by a statistically significant decreased score for every additional year since graduation (-0.24 ; $p = .045$). Male participants scored statistically significantly lower (-4.46 ; $p = .025$) than female participants. Taking the test multiple times was a statistically significant predictor for an increase of test scores (5.83 ; $p < .001$). Educational length had no statistically significant impact. A summary of results from the regression models are shown in [Table 2](#).

3.2 | Reliability and credibility of Angoff panel

All exams demonstrated substantial (to almost perfect) panel member inter-rater reliability, ranging from 0.65 to 0.82, see [Table 3](#). The panel members' predicted item difficulty was statistically significantly correlated ($p < .001$) to the item difficulties observed by the candidates' results, with correlation coefficients ranging from 0.69 to 0.77.

3.3 | Differentiation of candidates

3.3.1 | Professional qualifications

The group of candidates that did not pass the test scored statistically significantly lower in all professional qualifications compared to passing candidates. Lower age and taking the proficiency test

TABLE 1 Number and distribution of pass/fail and participants by category

Variable	Category	All (n = 316)	Pass, n = 98 (31%)	Fail, n = 218 (69%)
Age at first test (median; IQR; min-max)		34 (30–42; 24–62)	33 (29–40; 25–52)	35 (30–43; 24–62)
Gender	Female	149 (47.2)	45 (45.9)	104 (47.7)
	Male	167 (52.8)	53 (54.1)	114 (52.3)
Age group	24–29 y	74 (23.4)	28 (28.6)	46 (21.1)
	30–39 y	146 (46.2)	44 (44.9)	102 (46.8)
	40–49 y	75 (23.7)	24 (24.5)	51 (23.4)
	> = 50 y	21 (6.6)	2 (2.0)	19 (8.7)
Country of education	Central, Eastern, South and South-eastern Asia	36 (11.4)	13 (13.3)	23 (10.6)
	Eastern Europe (including Northern Asia)	37 (11.7)	7 (7.1)	30 (13.8)
	Latin America and the Caribbean	19 (6.0)	7 (7.1)	12 (5.5)
	Northern Africa	30 (9.5)	9 (9.2)	21 (9.6)
	Southern Europe	20 (6.3)	7 (7.1)	13 (6.0)
	Western Asia	147 (46.5)	49 (50.0)	98 (45.0)
	Missing	27 (8.5)	6 (6.1)	21 (9.6)
Education length	< 5 years	18 (5.7)	2 (2.0)	16 (7.3)
	≥ 5 years	250 (79.1)	80 (81.6)	170 (78.0)
	Missing	48 (15.2)	16 (16.3)	32 (14.7)
Years since graduation ^a (median)		9	8	9
Number of tests				
	1	201 (63.6)	60 (61.2)	141 (64.7)
	2	86 (27.2)	26 (26.5)	60 (27.5)
	3	29 (9.2)	12 (12.2)	17 (7.8)

^an = 294.

multiple times showed overall to be important factors in achieving scores linked to the professional qualifications, with eight and nine out of the ten qualifications respectively. Gender and educational length were occasionally statistically significantly associated with professional qualifications. A summary of results from the regression models of the professional qualifications can be seen in Table 4.

3.3.2 | Dental disciplines

The group of candidates who did not pass the test demonstrated lower scores in all dental disciplines in comparison with the group of participants who passed the test. However, the differences were only statistically significant in prosthodontics ($p = .024$). The results within dental disciplines in relation to the Angoff pass mark show that single candidates passing the examination failed to reach the acceptable level in all disciplines. Visual representation of candidates' pass/fail results within dental disciplines, in relation to the Angoff pass mark, are presented in Figure 2. For every added year of

age, statistically significantly lower percentages of the scores were obtained in the dental disciplines of *periodontology* ($p = .038$), *prosthodontics* ($p = .04$) and *endodontology* ($p = .034$). Results in nine out of ten dental disciplines improved with the participants taking the test multiple times. However, scores within *endodontology* declined statistically significantly ($p = .019$) when taking multiple tests. Gender and length of education were not statistically significantly associated with the outcome per dental discipline. A summary of the results from the regression models per dental discipline is presented in Table 4.

4 | DISCUSSION

The exams assessed in the study demonstrated substantial to almost perfect panel member inter-rater reliability when determining the pass level by the Angoff procedure. Item difficulty predicted by the expert panel showed significant ($p < .001$) positive correlations with item difficulties observed in candidates' results. Failure to pass the

TABLE 2 Multivariable linear regression with percentage of pass limit as outcome variable

Variables	Multivariable β (95% CI) <i>p</i> -value
Age at first test	-0.29 (-0.51 – -0.06) .013*
Male versus Female	-4.46 (-8.35 – -0.57) .025
Education ≥ 5 years vs < 5 years	5.04 (-2.82–12.91) .21
Years since graduation	-0.24 (-0.47 – -0.01) .045
Test number	5.83 (3.78–7.89) <.001

Note: Estimation (unstandardised) indicates the change in the outcome when a continuous variable increase with one unit (e.g., 1 year), and for categorical variables the difference compared to the reference category. Bold indicates significant values.

* $p < .05$.

TABLE 3 Inter-panel agreement (reliability) and correlation of item difficulty set by the modified Angoff method and item difficulty demonstrated by participants (credibility). Judgement and items derived from the seven examinations included in the study group

Exam date	Inter-panel Agreement	Correlation Angoff and item difficulty
1803	0.82	0.69*
1805	0.77	0.72*
1809	0.69	0.60*
1812	0.65	0.73*
1903	0.71	0.72*
1909	0.69	0.72*
1911	0.67	0.77*

Note: Standards of strength of kappa coefficient¹⁷: .61–.80 = substantial agreement, .81–1.0 = almost perfect agreement. Pearson correlation - Degree of correlation: ± 0.50 – ± 1 High degree of correlation. * $p < .001$.

proficiency test was significantly associated with poorer results in professional qualifications and in dental disciplines. However, candidates who passed the test could still fail to achieve an acceptable level set by the Angoff panel within one or more dental disciplines. The multivariable analysis showed that taking the test multiple times was associated with an increase in total score, whilst being older and male were factors associated with a decrease in total score.

4.1 | Reliability and credibility of the theoretical examinations

The reliability of Angoff panels has implications for validity as they offer guidance in the interpretation of candidates' test scores to pass or fail the acceptable candidate.³² The number of panel members to include in the panel has been debated in previous studies of standard setting with the Angoff methodology, with recommendations ranging from five to fifteen.^{33,34} In our panels 10–15 panellists were

used, a decision based on evidence-based practice and available resources. Which kinds of subject experts to use in Angoff panels is also under debate,^{35–38} especially as how panel members conceptualise a minimally qualified candidate^{36,39} may be affected by experience and training in the Angoff methodology and the assessment format. In the modified Angoff method used in the current study, the training period of panellists was relatively short, the raters were relatively young, and they rated individually in accordance with the protocol without conferring with each other. The simplified version of method used here is less time-consuming than the original Angoff procedure, requiring joint meetings to reach agreement. In addition, in the original version each question is rated from 0–100 percentage predicted correctness for the just qualified testtaker,¹⁷ whilst the simplified version used here applies a 3-graded rating scale only. The modifications of the method potentially impact the reliability and validity but are rational decisions essential to ensure a high quality as well as a feasible workload on top of other faculty obligations. Still, the simplified version of the Angoff method showed a high level of inter-rater reliability. The relationship between empirical and judged item difficulties is perceived as important, as greater agreement is associated with lower error in the pass mark⁴⁸ as well as a measure of internal consistency.^{20,40,41} The high level of agreement amongst raters and the strong agreement between judgement and candidates' results indicate a feasible and defensible method to set pass mark for the theoretical part of the Swedish proficiency test for dentists educated outside EU, EEA and Switzerland.

Other methods for setting the pass mark, e.g., Ebel,⁴² have been discussed to reach further comprehension of agreement and item implication, by including scales of item difficulty and relevance. Such ratings could give test providers important information that is currently unavailable. However, the Ebel method is a complex standard-setting process, imposes a burden on the standard-setting panel and needs to be further explored and tested to fit current and future arrangements for standard setting of the theoretical examinations. The pass mark set for the theoretical test should reflect what is acceptable for a Swedish dentist. However, the current Angoff panel is recruited from one, out of four, Swedish universities providing dental education, which could affect the generalisability of the consensus reached within the panel. To conclude whether this method is reliable across institutions, future studies with inclusion of Angoff panel members from all Swedish universities providing dental education in Sweden are required.

4.2 | Differentiation of candidates

4.2.1 | Professional qualifications

The group that did not pass the test scored significantly lower in all areas of professional qualifications. This indicates that the standard setting, combining the Angoff pass mark and the 50% cut score of professional qualifications, is well set to identify candidates who do not meet the requirements of the professional qualifications of the dental degree in Sweden.

TABLE 4 Univariable linear regression with percentage of professional qualifications and dental disciplines as outcome variable

Variables	Age at first test	Male versus Female	Education ≥ 5 years vs < 5 years ^a	Years since graduation ^b	Test number	Grade fail
Professional Qualification (ILO's) β (95% CI) p-value						
PQ 1	-0.26 (-0.45 - -0.07) 0.0061*	-2.26 (-5.12- 0.60)0.12	3.42 (-2.09-8.93) 0.22	-0.11 (-0.28-0.06) 0.19	4.41 (2.87-5.94) <0.001*	-20.95 (-22.62 - -19.29) <0.001*
PQ 2	-0.22 (-0.43 - -0.01) 0.042*	-2.97 (-6.12-0.18) 0.064	3.38 (-2.86-9.62) 0.29	-0.11 (-0.32-0.10) 0.29	5.13 (3.41-6.86) <0.001*	-20.65 (-22.87 - -18.44) <0.001*
PQ 3	-0.23 (-0.41 - -0.05) 0.012	-2.50 (-5.43- 0.42)0.093	2.53 (-3.95-9.02) 0.44	-0.09 (-0.26-0.08) 0.3	3.36 (1.59-5.13) <0.001*	-19.57 (-21.60 - -17.55) <0.001*
PQ 6	-0.27 (-0.45 - -0.09) 0.0033*	-2.10 (-4.97-0.77) 0.15	3.45 (-2.55-9.44) 0.26	-0.17 (-0.33-0.00) 0.05	3.48 (2.02-4.95) <0.001*	-20.22 (-22.06 - -18.38) <0.001*
PQ 7	-0.15 (-0.43-0.12) 0.28	-1.67 (-5.60- 2.27)0.41	0.22 (-8.05-8.48) 0.96	-0.02 (-0.29-0.25) 0.88	2.05 (-0.11-4.20) 0.063	-18.06 (-21.47 - -14.65) <0.001*
PQ 8	-0.20 (-0.45-0.04) 0.1	-1.75 (-5.80- 2.30)0.4	7.21 (0.89-13.54) 0.026*	-0.07 (-0.34-0.20) 0.59	7.13 (4.60-9.65) <0.001*	-22.90 (-26.67 - -19.13) <0.001*
PQ 9	-0.28 (-0.50 - -0.06) 0.013*	-2.48 (-5.85- 0.89)0.15	3.12 (-3.01-9.24) 0.32	-0.16 (-0.36-0.04) 0.12	2.71 (0.69-4.72) 0.0085*	-18.86 (-21.73 - -16.00) <0.001*
PQ 11	-0.35 (-0.55 - -0.16) <0.001*	-2.30 (-5.23 - 0.63)0.12	3.51 (-2.87-9.90) 0.28	-0.26 (-0.44 - -0.09) 0.0032*	3.42 (1.78-5.06) <0.001*	-19.75 (-21.69 - -17.81) <0.001*
PQ 13	-0.28 (-0.47 - -0.09) 0.0042*	-2.30 (-5.27- 0.67)0.13	4.01 (-1.74-9.76) 0.17	-0.14 (-0.32-0.03) 0.11	3.72 (1.90-5.53) <0.001*	-19.88 (-21.76 - -18.00) <0.001*
PQ 14	-0.29 (-0.51 - -0.06) 0.013*	-3.99 (-7.38 - -0.59) 0.022*	4.22 (-3.53-11.97) 0.28	-0.19 (-0.41-0.03) 0.097	4.64 (2.72-6.55) <0.001*	-20.39 (-22.88 - -17.89) <0.001*
Dental Disciplines β (95% CI) p-value						
Cariology	0.02 (-0.17-0.21) 0.85	-1.43 (-4.13-1.28) 0.3	0.89 (-5.64-7.42) 0.79	0.03 (-0.17-0.24) 0.75	1.80 (0.07-3.54) 0.042*	-1.97 (-5.10-1.16) 0.22
Endodontology	-0.23 (-0.45 - -0.02) 0.034*	2.41 (-1.14-5.96) 0.18	0.85 (-7.86-9.56) 0.85	-0.21 (-0.43-0.01) 0.063	-3.15 (-5.79 - -0.52) 0.019*	-2.54 (-6.44-1.35) 0.2
Oral Medicine	0.04 (-0.19-0.26) 0.76	-2.57 (-6.32-1.18) 0.18	2.09 (-4.59-8.78) 0.54	0.05 (-0.20-0.30) 0.7	2.18 (-0.96-5.32) 0.17	-3.68 (-8.07-0.71) 0.1
Oral Radiology	-0.16 (-0.33-0.01) 0.063	-0.75 (-3.63- 2.13)0.61	1.19 (-7.23-9.60) 0.78	-0.15 (-0.32-0.03) 0.11	3.25 (1.18-5.32) 0.0021*	-2.94 (-5.91-0.03) 0.052*
Periodontology	-0.21 (-0.42 - -0.01) 0.038*	-0.56 (-3.69- 2.57)0.73	0.88 (-5.45-7.20) 0.79	-0.15 (-0.37-0.07) 0.18	3.89 (1.23-6.55) 0.0043*	-2.27 (-6.06-1.52) 0.24
Paediatric Dentistry	-0.06 (-0.25-0.13) 0.53	-2.44 (-5.89-1.01) 0.16	2.17 (-5.66-10.00) 0.59	-0.06 (-0.27-0.15) 0.57	4.25 (1.94-6.56) <0.001*	-2.77 (-6.84 - -1.30) 0.18
Prosthodontics	-0.21 (-0.42 - -0.01) 0.04*	-2.05 (-5.56-1.46) 0.25	3.31 (-6.84-13.47) 0.52	-0.21 (-0.42-0.01) 0.061	5.06 (2.52-7.61) <0.001*	-4.79 (-8.94 - -0.63) 0.024*

Note: Estimation (unstandardised) indicates the change in the outcome when a continuous variable increase with one unit (1 year), and for categorical variables the difference compared to the reference category ($n = 316$). * $p < 0.05$. ^a $n = 268$, ^b $n = 294$.

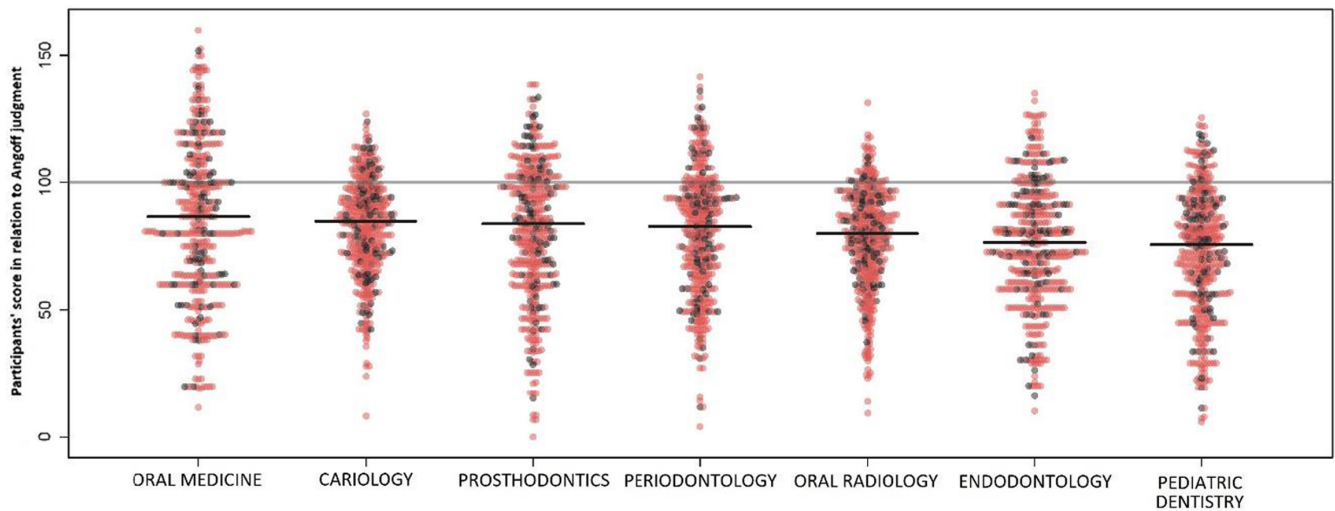


FIGURE 2 Bee swarm plots of participants' scores in relation to Angoff judgements within dental disciplines. Proportions and medians of achieved points in dental disciplines of all participants ($n = 316$) across seven theoretical examinations. Passing participants in black and failed participants in red. Grey line showing Angoff pass mark within dental discipline for reference.

4.2.2 | Dental disciplines

Participants not passing the test scored less in all dental disciplines throughout, compared to participants that passed the test. However, some passing participants had low scores within single disciplines, in relation to the acceptable standard set by the Angoff panellists. In this aspect, the standard setting of the theoretical examination has its limitations, as no preset standards are defined for the various dental disciplines. A suggestion for future improvements of the standard setting is to require that candidates should also reach an acceptable standard within each dental discipline. The risk of individuals passing the test despite knowledge gaps in certain disciplines is not exclusive to the Swedish proficiency test. Similar risks have been reported by the General Dental Council, UK, concerning the framework of the Overseas Registration Examination.⁴³

Can awareness of the identified potential knowledge gaps amongst participants who passed the theoretical proficiency test be used to further strengthen the current framework? One possible method could be to provide a detailed competency-based targeted feedback of candidate strengths and limitations, as a foundation for preparation and remediation for the next test application. A summary of results from both theoretical and practical examinations could also help identify specific areas for attention during the 6-month workplace-based assessment.

4.3 | Background variables

When analysing the results from background variables, proficiency in Swedish could be an undefined variable affecting test results, thus introducing noise to the interpretations of results. As participants are educated outside of Sweden one can expect that language skills vary and thereby affect performance on testing. Since the test

is evaluating proficiency in a Swedish dental context, participants need to be proficient in Swedish and recognise Swedish dental terminology and definitions. The strong association between lower test scores and increase in age could implicate that participants' prerequisites for being able to adapt to a new language is interwoven with age. A higher age may also implicate having a family to provide for and a need to take survival jobs to manage whilst preparing for the test. Survival jobs are considered an important factor, as working presumably affects the time needed for preparation and thus may extend the process.⁴⁴ However, working within the dental care area could contribute to the added experience of the dental profession in a Swedish context, and the development of overall language skills and dental terminology. Lower scores associated with higher age and language proficiency in this study are also comparable to physicians taking the Swedish proficiency test for physicians.⁴⁵ Other studies confirm the importance of verbal skills and language skills needed to understand questions.^{46,47}

The significant difference between genders was somewhat surprising, as both men and women are similarly educated with a dental degree from a higher education. To some extent gender might be interwoven with language proficiency, expectations to support their family financially with survival jobs, and ability to get a job within dentistry. This is not explored in the study and needs to be addressed in further studies.

Taking the proficiency test multiple times was a significant factor for passing, which could be an indication and that language skills might improve over time as participants become more proficient in Swedish in preparation for their second and third try. However, this could also be an effect of re-testing or some degree of test-wiseness.^{48,49}

The National Board of Health and Welfare include proof of certification of language skills for licensure within the licensing process. Yet in contrast to comparable proficiency tests for

foreign-trained dentists in other countries^{50,51} proficiency in Swedish language is not required before taking part in the proficiency test. This creates an “informal requisite” of language proficiency that could lead to poor results, regardless of the actual level of competence in dentistry. Proficiency in Swedish is an important aspect that needs to be continually addressed to minimise construct-irrelevant variance. To reduce the effect of language skills, a language review of each theoretical test is performed, informal language is used in items on the tests and in communication with participants.

The harmonisation of dental education in Europe is a work in progress and the ADEE framework of the graduating dentists has played an important part in quality assurance and alignment of the dental education within the EHEA. The harmonisation, indirectly, also plays an important part for the foundations of frameworks of recertification and expectations on foreign-trained dentists recertifying within the EHEA. The Swedish recertification process, through examinations and the assessments, is to a large extent harmonised with the ADEE even though it is independent of it. Considering the trends in foreign-trained dentists' migration to Europe there is a need for further harmonisation in recertification within the EHEA countries. ADEE is in a good position to further harmonise not only dental programmes, but also recertification processes within the EHEA.

Proficiency testing of foreign-trained dentists is highly important to validate the competence and skills of dentists who want to practice dentistry in Sweden. The present study is a part of a project aiming to explore the effect of the assessment framework applied in the Swedish proficiency test of dentists educated outside EU, EEA and Switzerland. Further research is needed to ensure the validity of proficiency tests, including standard setting in the examinations and on how the recertified dentists' and employers perceive the tests. These studies will bring a deeper understanding of this topic.

5 | CONCLUSIONS

Angoff panel inter-rater reliability and correlation of difficulty (credibility) propose that the currently used method of modified Angoff method is a simple and defensible method of setting the pass mark in the explored theoretical examinations. The method also significantly enables differentiation of foreign-trained dentists who meet the requirements of the professional qualifications from those who do not. However, Angoff is a method for setting the pass mark after the standard is set, and further exploration of item rigidity and content sampling is needed to ensure that a reasonable assessment of dentists is performed. Young age, being female and taking the test multiple times were associated with a better score.

ACKNOWLEDGEMENTS

We thank the candidates taking part in the tests, the Angoff panelists and the National Board of Health and Welfare for supplying the background information.

CONFLICT OF INTEREST

The authors have no conflict of interests that may have influenced the design, execution, or presentation of the scholarly work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Jesper Dalum  <https://orcid.org/0000-0002-2555-0135>

REFERENCES

- Balasubramanian M, Brennan DS, Spencer AJ, Short SD. The international migration of dentists: directions for research and policy. *Community Dent Oral Epidemiol*. 2016;44(4):301-312.
- General Dental Council. Overseas registration exam. At: <https://www.gdc-uk.org/registration/overseas-registration-exam>. Accessed: December, 2021.
- The National Dental Examining Board of Canada. How to become a licensed dentist in Canada. At: <https://ndeb-bned.ca/>. Accessed: December, 2021.
- The Dental Board of Australia. Overseas qualified practitioners. At: <https://www.dentalboard.gov.au/Registration/Overseas-Practitioner-Registration.aspx>. Accessed: December, 2021.
- Brennan RL. *E. National Council on Measurement in, and E. American Council on, Educational measurement*. Praeger Publishers; 2006.
- The European Higher Education Area. At: <http://www.ehea.info/page-how-does-the-bologna-process-work>. Accessed: February, 2021.
- The European Parliament and of the Council. Directive 2005/36/EC - the recognition of professional qualifications. At: https://ec.europa.eu/growth/single-market/services/free-movement-professionals/qualifications-recognition_en. Accessed: February, 2021.
- The Bologna Follow-up Group, Secretariat, European higher education area and Bologna process - The Sorbonne declaration. 1998.
- The Bologna Follow-up Group, Secretariat, Bologna Process. European Higher Education Area - Bologna Declaration. 1999.
- Field JC, Cowpe JG, Walmsley AD. The graduating European dentist: a new undergraduate curriculum framework. *Eur J Dent Educ*. 2017;21(S1):2-10.
- Field JC, Walmsley AD, Paganelli C, et al. The graduating European dentist: contemporaneous methods of teaching, learning and assessment in dental undergraduate education. *Eur J Dent Educ*. 2017;21(S1):28-35.
- Swedish Council of Higher Education. Higher Education Act 1993:100, Annex 2, System of Qualifications. 1993.
- National Board of Health and Welfare. HSLF-FS 2017:80 Socialstyrelsens föreskrifter och allmänna råd om legitimation för yrke inom hälso- och sjukvården vid utbildning från tredjeländ. 2017. p. 2.
- National Board of Health and Welfare. Dental practitioner educated outside EU and EEA. At: <https://legitimation.socialstyrelsen.se/en/licence-application/outside-eu-eea/dental-practitioner-educated-outside-eu-eea/>. Accessed: December, 2021.
- Williams JC, Baillie S, Rhind SM, Warman S, Sandy J, Ireland A. *A Guide to Assessment in Dental Education*. University of Bristol; 2015.
- Miller EG. The assessment of clinical skills/competence/performance. *Acad Med*. 1990;65(9):S63-S67.
- Angoff WH. *Scales, Norms, and Equivalent Scores*. Educational Testing Service; 1984.

18. Jelovsek JEMD, Walters MD, Korn A, et al. Establishing cutoff scores on assessments of surgical skills to determine surgical competence. *Am J Obstet Gynecol*. 2010;203(1):81.e1-81.e6.
19. Berk RA. A Consumer's guide to setting performance standards on criterion-referenced tests. *Rev Educ Res*. 1986;56(1):137-172.
20. Plake BS. Setting performance standards for professional licensure and certification. *Appl Meas Educ*. 1998;11(1):65-80.
21. Hambleton RK, Plake BS. Using an extended Angoff procedure to set standards on complex performance assessments. *Appl Meas Educ*. 1995;8(1):41-55.
22. Impara JC, Plake BS. Standard setting: an alternative approach. *J Educ Meas*. 1997;34(4):353-366.
23. Busch JC, Jaeger RM. Influence of type of judge, normative information, and discussion on standards recommended for the National Teacher Examinations. *J Educ Meas*. 1990;27(2):145-163.
24. Woehr DJ, Arthur W, Fehrmann ML. An empirical comparison of cutoff score methods for content-related and criterion-related validity settings. *Educ Psychol Meas*. 1991;51(4):1029-1039.
25. Taube KT. The incorporation of empirical item difficulty data into the Angoff standard-setting procedure. *Eval Health Prof*. 1997;20(4):479-498.
26. 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.
27. Chinn RN, Hertz NR. Alternative approaches to standard setting for licensing and certification examinations. *Appl Meas Educ*. 2002;15(1):1-14.
28. International Organization on Standardization. International Standard ISO 3166-1, Codes for the representation of names of countries and their subdivisions--Part 1: Country codes, ISO 3166-1: 2006 (E/F). 2006.
29. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
30. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.
31. Team, R.C. R: A language and environment for statistical computing. At: <https://www.R-project.org/>. Accessed: December, 2021.
32. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. American Educational Research Association; 2014.
33. 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):885-897.
34. Verhoeven BH, van der Steeg A, Scherpbier AJ, Muijtjens AM, Verwijnen GM, van der Vleuten C. Reliability and credibility of an Angoff standard setting procedure in progress testing using recent graduates as judges. *Med Educ*. 1999;33(11):832-837.
35. Verhoeven BH, Verwijnen GM, Muijtjens AM, Scherpbier AJ, van der Vleuten C. Panel expertise for an Angoff standard setting procedure in progress testing: item writers compared to recently graduated students. *Med Educ*. 2002;36(9):860-867.
36. Verheggen MM, Muijtjens AM, van Os J, Schuwirth LW. Is an Angoff standard an indication of minimal competence of examinees or of judges? *Adv Health Sci Educ Theory Pract*. 2006;13(2):203-211.
37. Wayne DB, Cohen E, Makoul G, McGaghie W. The impact of judge selection on standard setting for a patient survey of physician communication skills. *Acad Med*. 2008;83(10 Suppl):S17-S20.
38. Clauser JC, Hambleton RK, Baldwin P. The effect of rating unfamiliar items on Angoff passing scores. *Educ Psychol Meas*. 2017;77(6):901-916.
39. Boursicot K. Setting standards in a professional higher education course: defining the concept of the minimally competent student in performance-based assessment at the level of graduation from medical school. *High Educ Q*. 2006;60(1):74-90.
40. Goodwin LD. Relations between observed item difficulty levels and Angoff minimum passing levels for a Group of Borderline Examinees. *Appl Meas Educ*. 1999;12(1):13-28.
41. Kane M. Validating the performance standards associated with passing scores. *Rev Educ Res*. 1994;64(3):425-461.
42. Ebel RL. *Essentials of Educational Measurement*. Prentice-Hall; 1972.
43. Bissell V, Chamberlain S, Davenport E, Dawson L, Jenkins S, Murphy R. The overseas registration examination of the general dental council. *Br Dent J*. 2016;221(5):257-261.
44. Östgren C-J, Krook-Brandt M, Carlborg A. Internship-test reveals increased knowledge gaps. *Lakartidningen*. 2016;113.
45. Stureson L, Heiding A, Olsson D, Stenfors T. Did I pass the licensing exam? aspects influencing migrant physicians' results: a mixed methods study. *BMJ Open*. 2020;10(7):e038670.
46. Neiterman E, Bourgeault IL, Covell CL. What do we know and not know about the professional integration of international medical graduates (IMGs) in Canada? *Healthcare Policy/Politiques de Santé*. 2017;12(4):18-32.
47. Pattinson J et al. Exploring reasons for differences in performance between UK and international medical graduates in the membership of the Royal College of general practitioners applied knowledge test: a cognitive interview study. *BMJ Open*. 2019;9(5):e030341.
48. Foster D, Miller H. A new format for multiple-choice testing: discrete-option multiple-choice. Results from early studies. *Psychol Sci Q*. 2009;51:355-369.
49. Millman J, Bishop CH, Ebel R. An analysis of test-wiseness. *Educ Psychol Meas*. 1965;25(3):707-726.
50. General Dental Council. Evidence of English language competence: Guidance for applicants 2016.
51. Dental Board of Australia, Standard for registered dental practitioners: registration standard: English language skills. 2015.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: 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*. 2022;00:1-10. doi: [10.1111/eje.12851](https://doi.org/10.1111/eje.12851)