

First reported implementation of a German-language progress test in an undergraduate dental curriculum: A prospective study

B. Kirnbauer¹  | A. Avian² | N. Jakse¹ | P. Rugani¹ | D. Ithaler³ | R. Egger⁴

¹Division of Oral surgery and Orthodontics, Medical University of Graz, Graz, Austria

²Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria

³Organizational Unit for Teaching and Studies, Medical University of Graz, Graz, Austria

⁴Institute for Educational Science, Karl-Franzens University Graz, Graz, Austria

Correspondence

Kirnbauer Barbara, Division of Oral surgery and Orthodontics, Medical University of Graz, Graz, Austria.

Email: barbara.kirnbauer@medunigraz.at

Abstract

Introduction: Progress testing is a special form of longitudinal and feedback-oriented assessment. Even though well established in human medical curricula, this is not the case in dental education. The aim was the prospective development and implementation of the first reported German-language Dental Progress Test (DPT) for the undergraduate dental curriculum at the Medical University of Graz, Austria.

Material and methods: Participation in DPT was compulsory for all dental students in terms 7-12 (years 4-6). Three tests, each consisting of 100 items out of a pool of 375, were administered within 3 consecutive terms in 2016 and 2017. Rasch analyses were used to evaluate the questionnaire and identify misfitting items.

Results: In the item responses, 59.7% were "correct," 27.0% were "false" and 13.3% were answered with "don't know," with similar results at all 3 time points. The assumption of parallel ICC was met (T1: $\chi^2 = 51.071$, $df = 74$, $P = .981$; T2: $\chi^2 = 57.044$, $df = 67$, $P = .802$; T3: $\chi^2 = 58.443$, $df = 72$, $P = .876$) and item difficulties for the thematic fields were similarly distributed across the latent dimensions.

Conclusion: The newly introduced DPT is appropriate for testing dental students and is well balanced for the tested target group.

KEYWORDS

assessment, dental progress test, multiple choice items

1 | INTRODUCTION

Progress testing (PT) is a special form of longitudinal and feedback-oriented assessment performed at regular intervals and usually based on multiple choice items. It was first introduced by the University of Missouri Kansas City School of Medicine and Maastricht University in the 1970s as a novel assessment tool in the undergraduate academic world.¹⁻⁴ The basic intention of progress testing is the evaluation of the growth of knowledge during the course of an educational programme, motivating students to break the link between learning and examination.²⁻⁴ Instead of plain repetition of facts they have learned,

examinees are encouraged to call up and apply acquired information from their long-term memory.^{1,3-5} Consequently, the PT is a tool from which students as well as educators can benefit in various ways. It helps to trace the educational development of students, allows detailed feedback and identifies gaps in knowledge.⁶ In addition, PT is not course specific, but comprehensive and suitable for internal and external evaluation over the boundaries of courses and curricula.^{3,7-11}

Single tests mostly consist of multiple-choice items and are given at regular intervals. Each test assumes a comprehensive final-exam level of knowledge for a particular subject. In contrast to the usual multiple-choice answers there is an additional "don't know" (dk)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2018 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd.

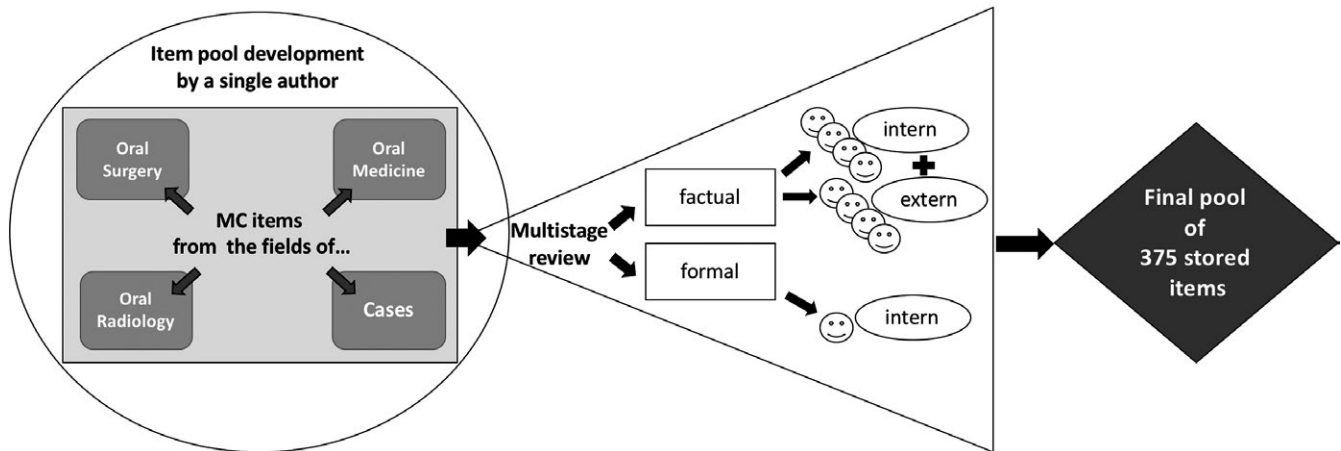


FIGURE 1 Description of item pool development

option to prevent students from guessing and to identify learning objectives they have not yet mastered. Items are usually taken from a pool previously compiled by specially trained educators that reflect predefined learning objectives.^{3,12}

Progress testing is now an established instrument in human medical curricula throughout Europe, including many German-speaking universities, and internationally. The Medical University of Graz, for instance, has applied this test format yearly since 2008 in the human medical curriculum in cooperation with the Charité University Hospital in Berlin.^{8,9,13,14}

This kind of tool is not, however, widespread in dental education. A Pubmed search produced only results for a dental progress test (DPT) for a Bachelor of Dental Surgery Programme and for a Dental Therapy and Hygiene Programme in the Peninsula School of Dentistry in Plymouth, UK.^{12,15,16} Since no German-language DPT is presently available, this prospective study aimed to develop and implement a reliable German-language DPT for the undergraduate dental curriculum at the Dental School of Medical University of Graz (Austria).

2 | MATERIALS AND METHODS

In 2016 and 2017, all 4th–6th-year (7th–12th terms) students at the Dental School of the Medical University of Graz were required to participate in this prospective study. Compulsory attendance was approved by the local Advisory Committee on Dental Study Affairs. Based on enrolment of 12 students per term, up to 72 students were expected, comprising men and women, mainly from Austria, but also from Germany and Southeastern European countries.

The Ethics Committee of Medical University of Graz reported no concerns about performance of this study.

2.1 | DPT development

The DPT project was developed by a senior staff member at the Division of Oral Surgery and Orthodontics with 10 years of experience in dental education and specially trained in the formulation of multiple

choice (MC) questions. A pool of 375 single-best and true/false (K-type) MC items was designed and stored password protected at the IMS² Item Management System (Umbrella Consortium for Assessment Networks, Heidelberg, Germany). Each item contained an explanatory introduction or case vignette, the question text itself and 5–6 possible answers, only 1 of which the correct key answer was. Each question contained a “Don’t know” (dk) option as a special feature of this progress test and depending on the topic, the author chose either 3 or 4 distractors. There were no double negatives. All items were at final exam level for the fields of “oral surgery,” “oral medicine,” “oral radiology” and “cases.” Fields included subcategories and correlated with the local catalogue of learning objectives pursued in the course of clinical dental training. Clinical images and radiographs were also included in 95 items to simulate routine dental practice (Figure 1).

2.2 | MC question review

Each question underwent a multistage review process. The first factual group review was performed in-house by a review committee of 4 senior academics. A second individual review by senior staff members followed. A third further individual review was by senior academics at Dental School of Medical University of Vienna. Before final inclusion in the question pool, there was an additional formal review by the local examination department (Figure 1).

2.3 | Test schedule and content details

In the course of this project, 3 progress tests were administered within 3 consecutive terms in 2016 and 2017 (Figure 2). For each test, 100 items were randomly selected based on a predesigned blueprint containing the 4 categories “oral surgery,” “oral medicine,” “oral radiology” and “cases.” In detail, each test consisted of 30 items from “oral surgery” including diagnostics, indications, surgery techniques, instruments, complication management and implant surgery, 30 from “oral radiology” including X-ray techniques, radiation protection and image interpretation, 20 from “cases” containing realistic everyday case vignettes and 20 from “oral medicine” including 5

			Study year						Students (n)
			4		5		6		
			Term						
Items n=375			7	8	9	10	11	12	
100 out of 375	Test 1	Cohorts	A	B	C	D	E	F	55
				↘	↘	↘	↘	↘	
100 out of 275	Test 2	Cohorts	G	A	B	C	D	E	62
				↘	↘	↘	↘	↘	
100 out of 175	Test 3	Cohorts	H	G	A	B	C	D	56
				↘	↘	↘	↘	↘	
									173

FIGURE 2 Test performance, Tests 1-3, each containing 100 different questions from the pool of 375 items. Student cohorts A-H (separated according to academic year 4-5, or, respectively, terms 7-12) reaching the next higher term (marked with →) or graduating in the course of the 3 tests

items from the subfield of “local anaesthesia” and 5 from the subfield “acute pain management.” No items were repeated. There was a 3-hour time limit for the computer-based test. Correct answers were scored +1, false answers -1 and 0 points were given for dk options. Participation was mandatory, but the results did not affect the students’ grades; accordingly, DPT performance did not influence pass/fail decisions. The students did, however, receive feedback on their scores concerning number of correct, incorrect and don’t know responses, their rank in class and their rank in the total cohort. The best performers were rewarded, eg with free congress registration and items from the university shop to increase motivation. For low achievers, their results helped identify knowledge gaps that individually assigned tutors could help to fill in.

2.4 | Post-review

The post-review committee analysed each test. Exclusion criteria were applied before the final evaluation. Reasons for exclusion from analysis were technical problems during test administration, eg projection of images or items with the wrong answer options.

2.5 | Statistical analysis

Data were analysed anonymously and blinded after the third test. They are presented as median and inter-quartile range (IQR) or absolute and relative numbers. After an extensive descriptive analysis, Rasch analysis was used to evaluate the test and identify misfitting items. For Rasch analysis, the response categories “don’t know” (0) and “false” (-1) were collapsed into “false.” Any items with only correct answers or only false answers were to be excluded from analysis. Item parameters and person parameters were estimated using response patterns and were expressed on a common log-odds scale. To identify items that did not fit into a unidimensional model or had different item parameters ($\hat{\beta}$) in subsamples of respondents (sample independence), infit and outfit measures (mean square statistics) and the Wald test were applied. To evaluate the assumption of parallel item characteristic curves (ICC), Andersen’s likelihood-ratio tests for goodness-of-fit

with mean split criterion were calculated. Person-item maps were created. As a measure of internal consistency, the person separation reliability was calculated. Since item dependency can inflate reliability, a second analysis was made grouping the items of the 4 thematic fields into 4 polytomous items. For these datasets of 4 polytomous items the reliability was also calculated. All analyses were done separately for the 3 time points using the R-package eRm (Version 0.15-7)¹⁷ and mirt (Version 1.27.1).¹⁸

3 | RESULTS

3.1 | Overall results

Overall 173 students including of men and women at a ratio of 1:0.7 sat for 3 tests. Three hundred MC items (100 per test) were primarily included, 6 of which were excluded by the test administrator after the post-review process because of mistakes in the answer options or technical problems with clinical photographs and radiographs. Overall, 59.7% of the item responses were “correct,” 27.0% were “false” and 13.3% were answered with “don’t know,” with similar results for all 3 time points (Table 1).

3.2 | Test results

In Test 2 the number of correct answers increased from term 7/8 (4th year) to 9/10 (5th year) ($P < .001$) and 7/8 to 11/12 (6th year) ($P = .002$) (term 7/8: median number correct answers: 48, IQR: 42-55; term 9/10: 61, 55-69; term 11/12: 60, 51-67). In the other 2 tests, the number of correct answers did not increase. In Test 1 and Test 2 the number of “don’t know” decreased from term 7/8 to 9/10 (Test 1: $P = .003$; Test 2: $P < .001$) and 7/8 to 11/12 (Test 1: $P < .001$, Test 2: $P < .001$) (Test 1: term 7/8: median number “don’t know” answers: 19, IQR: 10-29; term 9/10: 6, 2-13; term 11/12: 2, 0-5; Test 2: term 7/8: median number “don’t know” answers: 24, IQR: 18-32; term 9/10: 8, 5-15; term 11/12: 7, 2-12). The number of false answers increased in Test 1 from term 7/8 to 9/10 ($P = .009$) and 7/8 to 11/12 ($P = .022$) (term 7/8: median number false answers: 19, IQR: 14-25; term 9/10: 26, 22-38; term 11/12: 25, 20-34).

TABLE 1 Overview of responses, analysed items and final item pool

	Overall	Test 1	Test 2	Test 3
Correct answer (%)	59.7	61.6	56.0	62.1
False answer (%)	27.0	26.6	27.8	26.5
Don't know (%)	13.3	11.8	16.3	11.4
Items with all responses correct	0	0	0	0
Items with all responses false or "don't know"	0	0	0	0
Analysed Items	294	98	99	97
Items excluded because of				
... sample dependency	75	23	30	22
... outfit MS statistics	3	0	1	2
... infit MS statistics	0	0	0	0
... inappropriate response pattern	1	0	0	2
Final number of items				
All items	215	75	68	71
Oral surgery items	62	23	20	19
Oral medicine items	47	17	15	15
Oral radiology items	62	21	19	22
Cases items	43	14	14	15
Andersen's LR-Test (χ^2 ; <i>df</i> ; <i>P</i> -value)		51.071; 74; 0.981	57.044; 67; 0.802	54.559; 70; 0.913

3.3 | Item analysis

No item had to be excluded because all respondents had answered correctly or all had answered "false"/"don't know." A quarter (25.5%, $n = 75$) of the items had to be excluded due to the sample dependency, 1.0% ($n = 3$) because of too high or too low MSQ outfit statistics, none because of too high or too low MSQ infit statistics and 0.7% ($n = 2$) due to inappropriate response patterns within subgroups. These analyses resulted in 75 (Test 1), 68 (Test 2) and 72 (Test 3) items. The assumption of parallel ICC was met in all 3 tests (test 1: $\chi^2 = 51.071$, $df = 74$, $P = .981$; test 2: $\chi^2 = 57.044$, $df = 67$, $P = .802$; test 3: $\chi^2 = 58.443$, $df = 72$, $P = .876$). Person separation reliability was 0.88 for test 1, 0.86 for test 2 and 0.82 for test 3. Using the 4 polytomous score for the thematic field instead of all items, reliability decreased to 0.87 for test 1, 0.83 for test 2 and 0.77 for test 3. In all 3 tests, item difficulties for the thematic fields

"oral medicine," "oral radiology" and "cases" were similarly distributed across the latent dimensions. Whilst this is also true for "oral surgery" for Test 1, there were fewer difficult items for "oral surgery" in Test 2 and fewer easy items for "oral surgery" in Test 3. Using trait estimations, which were calculated using Rasch models, there was no significant increase in the measured latent trait (term 7/8: 0.23 ± 0.60 ; term 9/10: 0.40 ± 0.81 ; term 11/12: 0.64 ± 0.92) in Test 1. There was a significant increase from term 7/8 to 9/10 ($P < .001$) and 7/8 to 11/12 ($P = .001$) (term 7/8: -0.12 ± 0.77 ; term 9/10: 0.70 ± 0.62 ; term 11/12: 0.64 ± 0.62) in Test 2 and a significant increase from term 7/8 to 9/10 ($P = .043$) in Test 3 (term 7/8: 0.40 ± 0.57 ; term 9/10: 0.84 ± 0.64 ; term 11/12: 0.63 ± 0.69) (Figures 3-5).

4 | DISCUSSION

This new DPT is appropriate for testing dental students and is well balanced for the target group (Table 1; Figures 3-5).

To the best of our knowledge this is the first study to report the implementation of a German-language progress test in an undergraduate dental curriculum. Whilst well established in human medical curricula in Europe and beyond since the 1970s, the situation in dental education is different. To date, only the Dental School at the University in Plymouth (UK) is known to have established a DPT in different educational programmes.^{4,15,16} Wider use of PT in dental education should be pursued, as it is an essential source of information for dental educators as advocates for high-quality patient care, and for students as a meaningful feedback instrument.^{15,19,20}

In our DPT in its present form, overall numbers of "correct," "false" and "don't know" answers are similar to other PTs used in human and dental medical curricula (Table 1).^{4,6,7,9,15,19-21} Separate evaluation of all 3 tests administered showed similar ratings for each (Table 1), even though cohort composition and items varied. The reliability of our tests ranged from 0.82 to 0.88 (0.77-0.87), indicating good internal consistency. Compared to other PTs these values are similar or slightly higher.²²⁻²⁴

The congruity of our DPT question pool is supported by a similar unidimensional distribution of the 215 items ultimately selected for the test, satisfying the assumptions for Rasch analysis, and a pleasingly consistent distribution between the categories "oral surgery," "oral medicine," "oral radiology" and "cases" (Table 1). After linking these 3 tests with anchor items, an appropriate item bank for further testing can be provided. The application of IRT methods will also allow inclusion of new items and their calibration within the existing test.

Generally, DPT items varied in difficulty within each of the 3 different tests; they fit to the trait distribution of the tested students and lay within a preferable range (Figures 3-5).^{25,26} Distribution of item difficulties was good for 3 out of 4 categories, with deviations only in "oral surgery," with missing easy items in 1 test and lack of difficult items in another (Figures 4 and 5). The decision to split the item pool in groups of hundreds and administer 3 tests was made

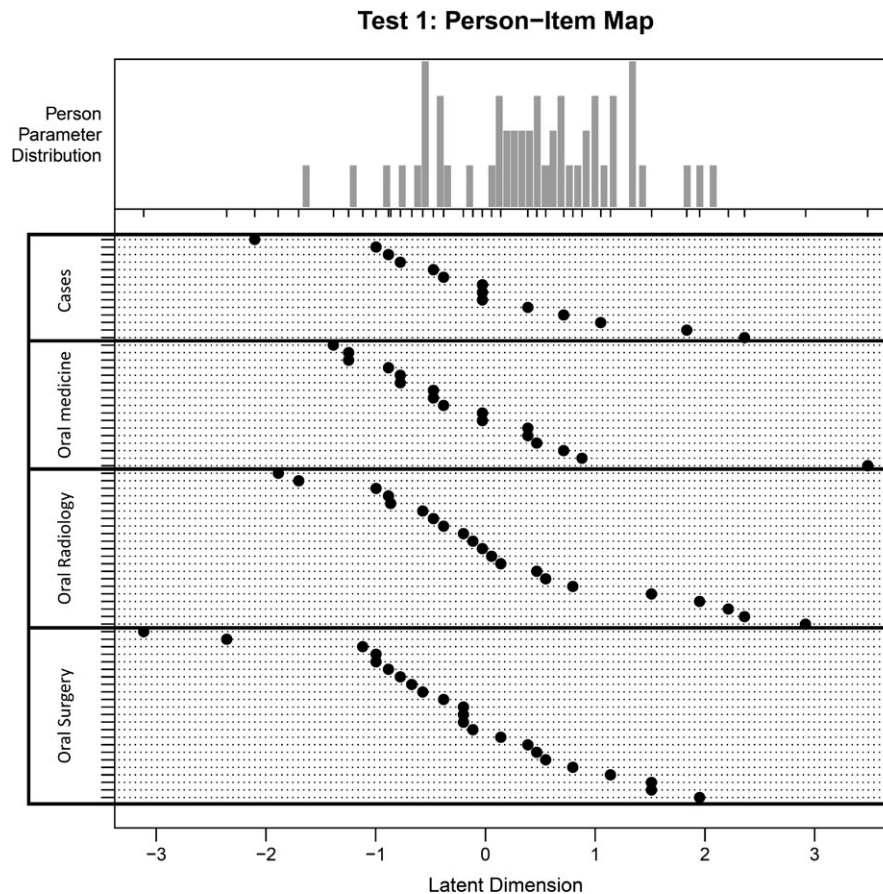


FIGURE 3 Person Item Map Test 1; Distribution of item difficulties within test 1, separated in the 4 subcategories Oral Surgery, Oral Radiology, Oral Medicine and Cases. Each point is showing the difficulty level of an item from difficult to easy (left to right), while the bars are showing the number of students with a certain ability level

to achieve an acceptable test length, considering the small number of different fields and the long time needed to complete the test. Therefore, and in contrast to established PTs, the difficulty parameters and the trait estimations between tests are not comparable.^{7,9,15,21,26} To get comparable difficulty parameters and therefore trait estimations on the same scale, items will be chosen from all 3 tests representing a wide difficulty range within each test and will be analysed together.

Based on the evaluation of the test results, there was a significant increase in “correct” answers according to length of training, indicating growth of knowledge, though only in 1 of 3 tests. Additionally, there was a significant decrease of “dk” answers as well as fewer “false” responses in 2 tests. These results basically conform to reported characteristics of other PTs.^{4,7,9,15,21}

We chose to give our DPT to more advanced students doing their clinical work, assuming that their growth of knowledge would be more transparent. First-year students cannot be expected to have good results, but with more senior students, our DPT could be expected to motivate rather than demotivate.^{7,9,12} Furthermore, PT’s formative character (non-relevance for grades), as chosen for this DPT, does in fact tend to prevent interference with the curriculum, in that it is not an extra burden for students, does not influence formal evaluation of the student’s progress and provides an ad hoc picture

of spontaneously recalled knowledge.^{14,27} However, according to Albano et al.,²⁸ growth curves can be more irregular, with more ups and downs at a formative design, as seen with this DPT. To counter the potentially negative aspect of high variance of motivation, best performers were rewarded. Moreover, the changes in cohort composition from term to term also could have influenced results.^{14,27,28}

Some further limitations of this study have to be mentioned. The Item pool was limited to 375 questions for 2 main reasons. First, the items were written by a single author whilst the project was being developed, and second the presented DPT only concerns the surgical part of the undergraduate dental curriculum at the Medical University of Graz. Furthermore, the study period, defined as 3 consecutive terms and representing a pilot phase, produced only a small number of evaluable tests, items and cohorts. Despite these limitations the presented evaluation will be important for future development of our DPT. Concerning the number of participating students, the limited group capacity of 12 students per term, or 24 students per year, has to be kept in mind. This makes it all the more important to continue data acquisition with DPT.

Overall, PT is a good technique for longitudinal assessment that lets students prepare more continuously. It is not restricted to any particular form of curriculum; however, differing results are sensitive information and have to be handled with care.²⁹ Several reports,

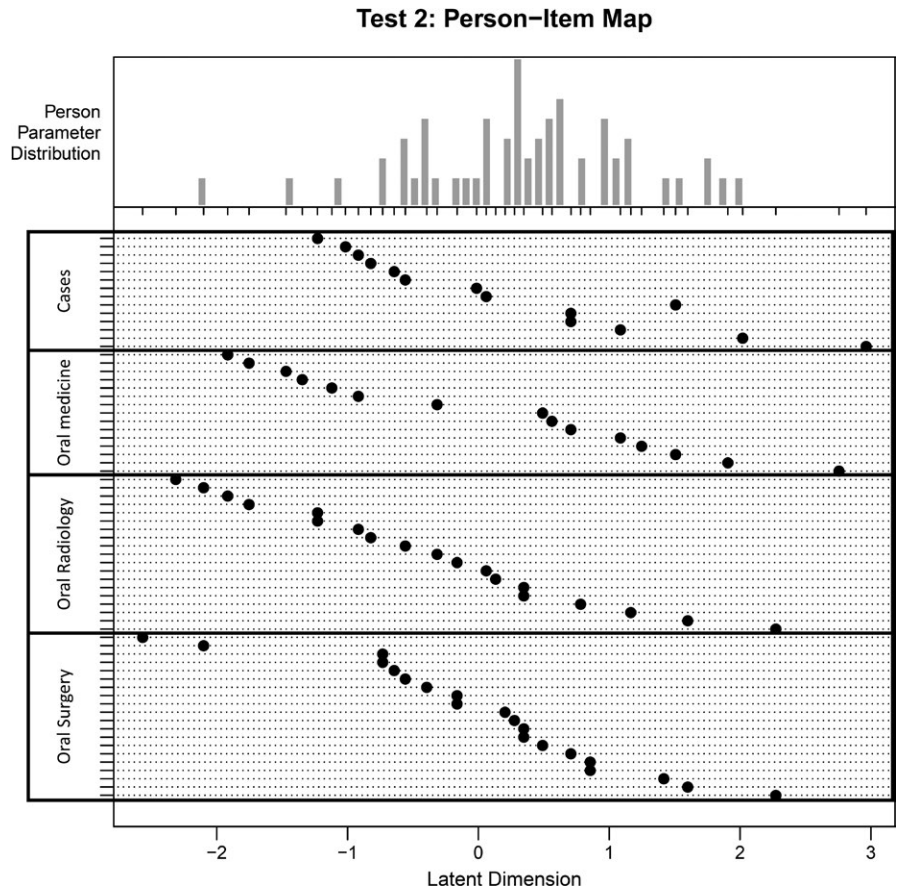


FIGURE 4 Person Item Map Test 2; Distribution of item difficulties within test 2, separated in the 4 subcategories Oral Surgery, Oral Radiology, Oral Medicine and Cases. Each point is showing the difficulty level of an item from difficult to easy (left to right), while the bars are showing the number of students with a certain ability level

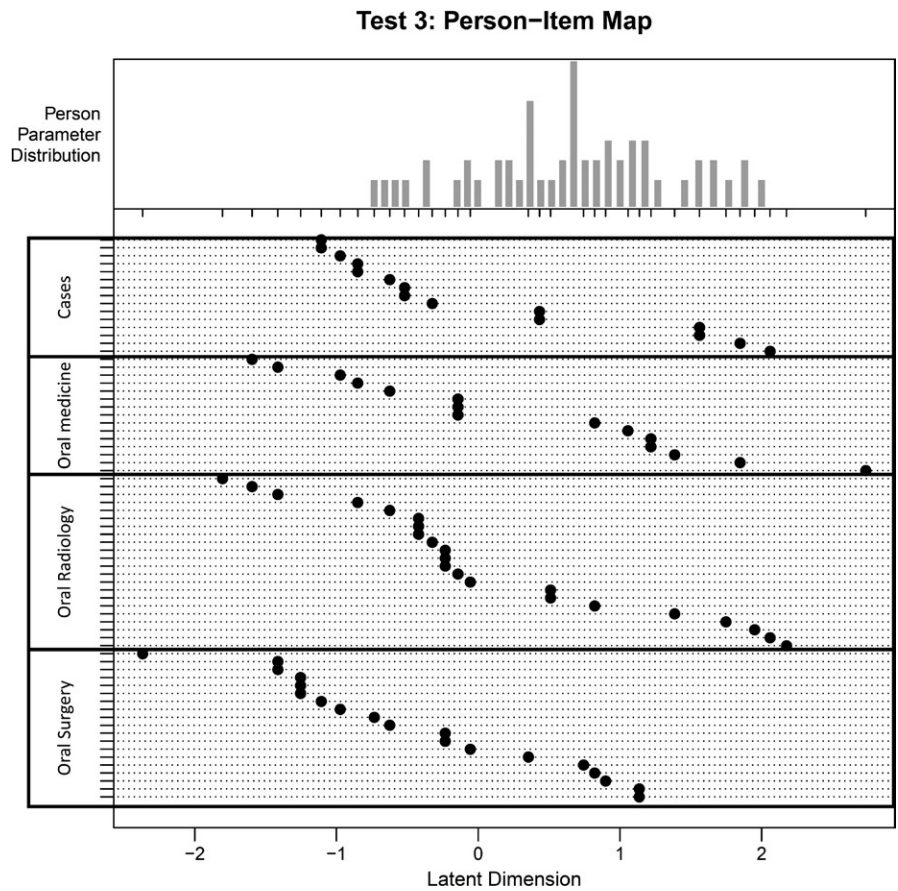


FIGURE 5 Person Item Map Test 3; Distribution of item difficulties within test 2, separated in the 4 subcategories Oral Surgery, Oral Radiology, Oral Medicine and Cases. Each point is showing the difficulty level of an item from difficult to easy (left to right), while the bars are showing the number of students with a certain ability level

including a recently published position paper, "The Graduating European Dentist",¹⁹ show clear advantages of quality control in educational programmes. Results of PT can be divided into many different sub-scores, so providing a rich source of information. Early detection of high and low achievers may identify a need for individualised support. PT can be offered additionally to other exams or can be implemented as an important grade-relevant periodical assessment. No matter how PT is implemented, longitudinal data collection allows more prediction of future competence and/or performance than 1-time assessments and could permit comparison of graduating dentists with the wider educational community, national or international. However, beside many advantages, disadvantages like high workload and costs, especially for development and item bank administration, can be major hurdles in implementing PT.^{4,13,19,20,29-31}

5 | CONCLUSION

Targeting the aim of the presented study, our DPT resulted in a homogeneous distribution of response behaviour, a consistent spread of included and excluded items within the separate tests and fields and a satisfactory range of difficulty of the questions. Growth of knowledge during the clinical educational programme was also documented. Consequently, this introduction of our German-language DPT can be deemed a success.

Our DPT offered an innovative, comprehensive survey of students' knowledge and provides fresh impulses for the educational programme at our dental school.

To obtain a common metric across the items in our 3 tests, we plan a combined test with selected items.

Finally, national or international collaboration to extend the item pool, share resources and compare results would be a desirable perspective.

ORCID

B. Kirnbauer  <http://orcid.org/0000-0002-8944-3160>

REFERENCES

- Arnold L, Willoughby TL. The quarterly profile examination. *Acad Med*. 1990;65:515-516.
- McHarg J, Bradley P, Chamberlain S, Ricketts C, Searle J, McLachlan JC. Assessment of progress tests. *Med Educ*. 2005;39:221-227.
- van der Vleuten CP, Schuwirth LW, Muijtjens AM, Thoben AJ, Cohen-Schotanus J, van Boven CP. Cross institutional collaboration in assessment: a case on progress testing. *Med Teach*. 2004;26:719-725.
- Van der Vleuten CP, Verwijnen G, Wijnen W. Fifteen years of experience with progress testing in a problem-based learning curriculum. *Med Teach*. 1996;18:103-109.
- Freeman A, Nicholls A, Ricketts C, Coombes L. Can we share questions? Performance of questions from different question banks in a single medical school. *Med Teach*. 2010;32:464-466.
- Coombes L, Ricketts C, Freeman A, Stratford J. Beyond assessment: feedback for individuals and institutions based on the progress test. *Med Teach*. 2010;32:486-490.
- Blake JM, Norman GR, Keane DR, Mueller CB, Cunningham J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: psychometric properties and effect on learning. *Acad Med*. 1996;71:1002-1007.
- Freeman A, Van Der Vleuten C, Nouns Z, Ricketts C. Progress testing internationally. *Med Teach*. 2010;32:451-455.
- Osterberg K, Kölbl S, Brauns K. Der Progress Test Medizin: Erfahrungen an der Charite. *GMS Z Med Ausbild*. 2006;23:Doc46. <http://www.egms.de/en/journals/zma/2006-23/zma000265.shtml>. Assessed on 2018/05/28.
- Swanson DB, Holtzman KZ, Butler A, et al. Collaboration across the pond: the multi-school progress testing project. *Med Teach*. 2010;32:480-485.
- Verhoeven BH, Snellen-Balendong HA, Hay IT, et al. The versatility of progress testing assessed in an international context: a start for benchmarking global standardization? *Med Teach*. 2005;27:514-520.
- Bennett J, Freeman A, Coombes L, Kay L, Ricketts C. Adaptation of medical progress testing to a dental setting. *Med Teach*. 2010;32:500-502.
- Aarts R, Steidel K, Manuel BA, Driessen EW. Progress testing in resource-poor countries: a case from Mozambique. *Med Teach*. 2010;32:461-463.
- Nouns ZM, Georg W. Progress testing in German speaking countries. *Med Teach*. 2010;32:467-470.
- Ali K, Coombes L, Kay E, et al. Progress testing in undergraduate dental education: the Peninsula experience and future opportunities. *Eur J Dent Educ*. 2016;20:129-134.
- Ali K, Zahra D, Tredwin C, McIlwaine C, Jones G. Use of progress testing in a UK dental therapy and hygiene educational program. *J Dent Educ*. 2018;82:130-136.
- Mair P, Hatzinger R. Extended Rasch Modeling: the eRm package for the application of IRT models in R. *J Stat Softw*. 2007;20:1-20.
- Chalmers R. mirt: a multidimensional item response theory package for the R environment. *J Stat Softw*. 2012;48:1-29.
- Field JC, Cowpe JG, Walmsley AD. The graduating European Dentist: a new undergraduate curriculum framework. *Eur J Dent Educ*. 2017;21:2-10.
- Field JC, DeLap E, Manzanares Cespedes MC. The graduating European dentist? Domain II: safe and effective clinical practice. *Eur J Dent Educ*. 2017;21:14-17.
- De Champlain AF, Cuddy MM, Scoles PV, et al. Progress testing in clinical science education: results of a pilot project between the National Board of Medical Examiners and a US Medical School. *Med Teach*. 2010;32:503-508.
- Pugh D, Regehr G. Taking the sting out of assessment: is there a role for progress testing? *Med Educ*. 2016;50:721-729.
- Rademakers J, Ten Cate TJ, Bar PR. Progress testing with short answer questions. *Med Teach*. 2005;27:578-582.
- Williams RG, Klamen DL, White CB, et al. Tracking development of clinical reasoning ability across five medical schools using a progress test. *Acad Med*. 2011;86:1148-1154.
- Hendriks J, Fyfe S, Styles I, Skinner SR, Merriman G. Scale construction utilising the Rasch unidimensional measurement model: a measurement of adolescent attitudes towards abortion. *Australas Med J*. 2012;5:251-261.
- Downing SM. Item response theory: applications of modern test theory in medical education. *Med Educ*. 2003;37:739-745.
- Schauber S, Nouns ZM. Using the cumulative deviation method for cross-institutional benchmarking in the Berlin progress test. *Med Teach*. 2010;32:471-475.

28. Albano MG, Cavallo F, Hoogenboom R, et al. An international comparison of knowledge levels of medical students: the Maastricht Progress Test. *Med Educ*. 1996;30:239-245.
29. Schuwirth L, Van der Vleuten CP. The use of progress testing. *Perspect Med Educ*. 2012;1:24-30.
30. Norman G, Neville A, Blake JM, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. *Med Teach*. 2010;32:496-499.
31. Tio RA, Schutte B, Meiboom AA, et al. The progress test of medicine: the Dutch experience. *Perspect Med Educ*. 2016;5:51-55.

How to cite this article: Kirnbauer B, Avian A, Jakse N, Rugani P, Ithaler D, Egger R. First reported implementation of a German-language progress test in an undergraduate dental curriculum: A prospective study. *Eur J Dent Educ*. 2018;22:e698–e705. <https://doi.org/10.1111/eje.12381>