Comparison of the evaluation of formative assessment at two medical faculties with different conditions of undergraduate training, assessment and feedback

Abstract

Introduction: Both formative and summative assessments have their place in medical curricula: formative assessment to accompany the learning process and summative assessment to ensure that minimum standards are achieved. Depending on the conditions of undergraduate training, assessment and feedback, students place more or less importance on formative assessment, and thus the fulfilment of its function may be questionable. This study describes how the low-stakes formative Berlin Progress Test (BPT) is embedded at two medical faculties with partially different framework conditions and what effects these have on the students' testing efforts and the evaluation of the test, especially the perception of its benefits and (intangible) costs, such as non-participation in contemporaneous activities and emotional impairments.

Methods: In this study, the proportion of non-serious BPT participants at two medical faculties (total sample: N_{F1}=1,410, N_{F2}=1,176) in winter term 2015/16 was determined both by the number of unanswered questions on the test itself and in a survey using a standardized instrument (N_{F1}=415, N_{F2}=234). Furthermore, open questions were asked in this survey about perceived benefits and perceived costs, which were analyzed with qualitative and quantitative methods.

Results: The BPT is generally better accepted at Faculty 2. This can be seen in the higher proportion of serious test takers, the lower perceived costs and the higher reported benefit, as well as the higher proportion of constructive comments. Faculty 2 students better understood the principle of formative testing and used the results of the BPT as feedback on their own knowledge progress, motivation to learn and reduction of exam fear.

Discussion: When medical faculties integrate formative assessments into the curriculum, they have to provide a framework in which these assessments are perceived as an important part of the curriculum. Otherwise, it is questionable whether they can fulfil their function of accompanying the learning process.

Keywords: formative assessment, medical education, progress test, test effort

Katrin Schüttpelz-Brauns[1]
Yassin Karay[2]
Johann Arias[3]
Kirsten Gehlhar[4]
Michaela Zupanic[5]

1 Medical Faculty Mannheim at Heidelberg University, Mannheim, Germany

2 University of Cologne, Medical Faculty, Cologne, Germany

3 RWTH Aachen University, Medical Faculty, Aachen, Germany

4 Carl von Ossietzky University, School of Medicine and Health Sciences, Oldenburg, Germany

5 University Witten/Herdecke, Faculty of Health, Witten, Germany

Introduction

According to the Medical Licensing Regulations (ÄAppO), §2 subsection 7, successful participation in the pre-clinical phase must be proven with 17 major course assessments (Appendix 2a) and in the clinical phase with 40 major course assessments (Appendix 2b). This proof is either provided by a graded assessment, which tests the learning outcome of a section, such as a subject or module, or by a pass/fail assessment. Therefore, these are assessments of learning or summative assessment [1]. On the other hand, there are assessments that accompany the learning process. These formative assessments [1] promote continuous and in-depth learning [2]. Feedback is a central aspect of continuous learning in that gaps in learning are identified and corrected in a targeted manner. Continuous learning prepares for lifelong learning, which is becoming increasingly important due to the fast pace of knowledge and constantly changing requirements [3]. There are already some studies on formative assessment that investigate the effect on learning. This so-called educational impact is part of the model of utility of assessment methods [4] and can be seen as an indication that the effect on continuous learning is given by the formative assessment or its feedback. Wade et al. developed a questionnaire to compare the perception of progress tests – a type of formative assessment (see below) – as a learning tool at

two different medical schools and found that the learning environment has an impact on the appreciation of progress tests as a learning support [5]. Cobb et al. asked students in semi-qualitative interviews about their perception of DOPS (formative direct observation of procedural skills) compared to MCQs (summative multiple-choice assessment) and found that formative testing promoted deeper learning, but summative testing was more important for students [6]. In a questionnaire study at the Faculty of Health Sciences in Maastricht, students found summative block tests more rewarding and did not use the results of the progress test for self-regulated learning [7]. Embedding feedback through/with progress tests in a comprehensive examination programme increased student use of the progress test feedback tool and integration into learning [8] .

Both the continuous accompaniment of the learning process through formative assessment and the assurance of the achievement of minimum standards in the form of summative assessment are justified in the medical curriculum.

It can be assumed that the two aims – learning for the assessment vs. assessment for learning – are pursued by students with varying degrees of intensity, which can be seen in the different extent of test effort. This can be explained with the Expectancy Value Theory of Wigfield & Eccles [9]. The Expectancy Value Theory states that the motivation to complete a task depends on two components: the expectation of being able to solve a task and the value that this task has for the individual. Wigfield & Eccles [9] distinguish four different components that can make up this value:

- Performance value (to master the task in the best possible way),
- Intrinsic value (the fun or joy in solving the task),
- Utility value (how well the task fits into future plans, i.e. how useful the task is)
- Costs (costs in the strict sense, the extent to which activities compete with each other, but also emotional costs).

With regard to summative and formative assessment, there is a difference in the value given to a task, the assessment. The value that a task or assessment has for each student is also influenced by the general conditions at the medical school. This has a decisive influence on the perceived benefits and costs. For example, it is very likely that the summative assessments are in the focus of the students, if the evidence required by the ÄAppO has to be proven at their medical faculty exclusively by summative assessment. In the worst case, they learn extremely efficiently, i.e. they learn all the required content shortly before the corresponding assessment, the so-called bulimia learning [10]. From the students' point of view, the benefit – passing the assessment – would be maximum at minimum cost. The knowledge acquired in this way runs the risk of being "ticked off" after the exam and soon forgotten [11].

Especially with regard to the benefits and costs of Expectancy Value Theory, faculties can provide framework conditions to increase the motivation to use formative assessment and thus the influence on learning. Formative assessment can be seen as an additional effort, especially if summative assessments and/or work-intensive courses (study load) have to be taken in parallel. If, on the other hand, formative assessment is perceived as a meaningful and valuable component of the overall curriculum and is valued by faculty members, the benefit of formative assessment could be regarded as high – despite contemporaneously graded assessment and high study load.

The formative progress tests in medicine offer an opportunity to investigate under which conditions formative assessment can be successfully implemented despite the competing summative assessments that are perceived as more useful.

Progress tests are multiple-choice tests that regularly test students' medical knowledge during undergraduate training at the level of a new graduate and compare it with the knowledge level of fellow students in the same semester in order to identify gaps in the current level of knowledge and to constructively influence learning behaviour. All types of progress tests provide feedback, but are used differently with regard to their stakes. In the Dutch consortium and in the USA, for example, the results of the progress test are accumulated over several test times for each individual [12], [13]. This means that the progress tests are not graded, but they do have an influence on the progress of study. In Germany and Austria, participation is mandatory, but is not graded (low stakes) [2], [14]. In the German-speaking Progress Test Medicine (BPT) consortium, test preparation and analyses are carried out centrally at the Charité University Medicine in Berlin. All test takers receive detailed feedback from Berlin about 4-6 weeks after the test on their results over the years, but also in comparison with their fellow students, differentiated by organ systems and subjects. The varying degree of test efforts at the individual faculties is reflected in the proportion of serious test takers, which is routinely computed after each test. In the case of the low-stakes BPT, it is shown that there are very different proportions of serious test takers at the various faculties. Proportions of 75-90% were reported by the participating faculties [15].

This study examines how the low-stakes BPT is embedded in two faculties and how this affects the students' testing efforts and the perception of the progress test, especially the perception of costs and benefits as a formative test. The framework conditions for the BPT differ at both faculties, among other things, in their integration into each curriculum: The conditions of undergraduate training, assessment and feedback are shown in detail in table 1. Looking at the conditions of assessment and feedback at both faculties, the proportion of serious test takers and the associated perception of the costs and benefits of the BPT should be comparable, as both faculties have conditions that should have a positive effect on motivation and, accordingly, on testing efforts.

Table 1: Conditions of undergraduate training, assessment and feedback at the two faculties

Characteristic	Medical school 1	Medical school 2
Course of studies	Model study course	Model study course
Number of students	2900	1250
Structure of studies	Model study course with 3 study sections 1. Pre-clinical phase, then M1 2. Clinical phase, then M2 3. Final year, then M3	Model study course with 4 study sections 1. 1st and 2nd sem. 2. 3rd to 6th sem., then M1 3. 7th to 10th sem., then M2 4. Final year, then M3
BPT since	winter term 2003	winter term 2003
Test administration	Computer-based	Paper-based with option of computer-based administration
Participation	Compulsory from the first semester, in pre-clinical phase at least 3 times, in clinical phase at least 8 times (= 8 times per undergraduate training)	Compulsory from the first semester and one BPT per semester is prerequisite for admission to the next study section. In the 3rd study section only one compulsory participation in the 8th or 9th semester (elective semester), planning avoids contemporaneous activities
Consequences	With no participation no admission to further courses	With no participation no admission to further courses / further study sections
Integration of the BPT	In the quality management (evaluation regulations)	In the assessment system (assessment regulations)
Information about BPT	The students are informed about the results by e-mail. Students are informed about the BPT in the introductory courses of the first semester (Power Point presentation), but not afterwards.	Information about dates is sent by e-mail. Benefits and general information about the BPT are communicated each semester in a lecture before the respective implementation.
Receiving the results	Students receive first results immediately after the test. They receive their test result and can compare their answers with the solution key. At the end of the semester students can download the Berlin report of their results.	Paper-based tests offer a solution booklet for direct feedback. PDF evaluation is carried out via an online portal. In addition, the BPT results are published in a feedback tool ("HIP-How I Perform") in which the BPT results are compared with the summative course results.
Dialogue on the results	None	In the next group presentation at the beginning of the semester, the results of the previous semester are briefly analysed. Based on the results, mentoring interviews will take place if necessary.

Note: M1: first part of the state examination, M2: second part of the state examination, M3: third part of the state examination

Students at Faculty 1 have a choice that is not available at Faculty 2. They can choose which 8 out of 10 BPTs they would like to take. According to the Self-Determination Theory of Ryan & Deci [16], this should increase intrinsic motivation and thus increase the proportion of serious test takers.

In addition, Faculty 1 provides immediate feedback from the computer-based administration. Immediate feedback is important for completing tasks and being satisfied with the work [17], [18], [19]. Therefore, the condition of computer-based administration should also increase the test effort and thus the proportion of serious test takers. Although the feedback is immediate, there is no dialogue about the results at Faculty 1. The dialogue about the results is integrated into the mentoring programme at Faculty 2. Dialogue is essential for effective feedback and thus for the functioning of formative assessment [20], [21], [22]. This should increase the *perceived benefit* of the BPT at Faculty 2.

Since the BPT is communicated as an assessment at Faculty 2, as opposed to Faculty 1, where it is presented as an evaluation, the BPT at Faculty 2 should be perceived as more useful for another reason. As Heeneman et al. were able to show in their study, students use the feedback system of the moderate-stakes progress test more and have higher test scores when the progress test is integrated into a holistic examination system [8]. The higher test scores were seen as an indirect indicator of test efforts. At the same time, the *perceived costs* are lower when the formative test is part of the assessment system.

Taking into account the conditions at the two faculties and their theoretical influences on the test effort, measured by the proportion of serious vs. non-serious test takers, and on the perceived costs and benefits of the BPT, the following hypotheses can be derived:

1. *The proportion of non-serious test takers at Faculty 1 (F1) is lower than at Faculty 2 (F2).*
2. *The perceived costs of the BPT are higher at Faculty 1 (F1) than at Faculty 2 (F2).*
3. *The perceived benefit of the BPT is lower at Faculty 1 (F1) than at Faculty 2 (F2).*

Methods

The study is conducted as a mixed-method approach, in which the proportions of non-serious test takers are determined quantitatively. In the qualitative part, the themes are identified which are relevant for the students in terms of perceived benefits and costs in relation to the BPT at both faculties.

Sample

In winter semester 2015/16, N=1,410 (F1) and N=1,176 (F2) medical students participated in the BPT. This corresponds to 50% of the enrolled medical students at F1 and 61% at F2. The proportion of female students at the faculties is 62% (F1) and 68% (F2).

Material

The proportion of non-serious test takers was determined in two different ways. On the one hand, those students who chose the "don't know" option for all questions or skipped all questions when filling out the test in winter term 2015/16 were identified as non-serious, since even in the first semester at least two questions can be answered. On the other hand, the test effort was determined by means of the Test-Effort Short Scale (TESS) [23]. TESS consists of three five-stage Likert items with the gradations 1 to 5, which ask for the performance value ("I would like to achieve the best possible result on the BPT"), the utility value ("I find the BPT useful") and the perceived costs ("The BPT is a valuable part of my undergraduate training"). The mean value is calculated from the answers to all three questions. Students who did not agree with these statements and answered all questions with 1 (corresponding to a TESS score of 1) are categorized as non-serious test takers. Both procedures each have a methodological disadvantage that could reduce their validity. The disadvantage of self-response tests is that there is an unknown percentage of students who answer in a socially desirable manner. This means that they could indicate a higher level of testing effort than is actually the case. The disadvantage of identification via the "don't know" option is that there may also be so-called pattern markers. These are test takers who answer all questions but do so without knowing the text of the questions [24]. Due to these disadvantages, both methods have been used in parallel.

In order to make the perceived costs and benefits measurable, we have asked open questions. Both the concept of costs and the concept of benefits are very abstract. Therefore, we asked formally balanced questions that provoke possible answers that can be assigned to these two terms. These are, on the one hand, questions about the disadvantages and advantages of the BPT, but also questions directly about the benefits of the BPT. Students who use the BPT should also talk to other people about their results, such as their mentor, in order to change their own learning behaviour.

The perceived costs were addressed in two open questions:

- Do you feel emotionally impaired by the BPT? (Question 1)
- What disadvantages do you see in the BPT? (Question 2)

The perceived benefit was determined by means of five questions (two closed and three open questions) on different aspects:

- Dialogue with other people about the results of the BPT with the sub-questions:
  - I talk with fellow students about my results on the BPT. (Likert item, with 1 "does not apply" to 5 "applies");
  - I talk to my mentor about my results on the BPT. (Likert item, with 1 "does not apply" to 5 "applies");
  - I talk to other people about my results on the BPT. With ... (open question, Question 3).
- Do you use the results of the BPT for other purposes? (open question, Question 4)
- What advantages do you see in the BPT? (open question, Question 5)

There was no limit to the number of comments that students could make on the open questions.

In addition, the questionnaire asked for gender and semester of study in order to check the comparability of both groups.

Procedures

At Faculties 1 and 2, the BPT took place in the first weeks of the semester on the university premises and under supervision. At least two non-overlapping dates were planned for each cohort, which the students could choose independently. At both faculties the testing was computer-based. At Faculty 2, additional dates for paper-based testing were offered.

The students at both faculties participated regularly in the BPT. At the beginning of the test, students were informed about the overall study in addition to the regular introduction. The overall study examines the motivation on the BPT and its influence on learning on the BPT. Therefore, the questionnaire contained more questions than the ones given here. In the regular introduction, the participants were asked to complete the questionnaire after the test had been completed and were informed that this participation was voluntary and anonymous. The Ethical Review Board of the Medical Faculty Mannheim, Heidelberg University, approved the study (2015-542-N-MA).

Analyses

The proportion of non-serious test takers per faculty was checked for independence in each case using a χ^2 test. Since the sample is very large and therefore even small differences can become significant, the effect size was

measured with Cohen's w for contingency tables and Cohen's d for metric data (see below) in order to assess the relevance of differences [25]. The effect size w is categorized as no effect with w<0.1, small effect with w<0.3, moderate effect with w<0.5 and large effect with w≥0.5 [25].

To compare the TESS scores between the two faculties, a t-test for independent samples with unequal variances was calculated, and the effect size d according to Cohen [25], with pooled standard deviations according to Leonhart (2004) was calculated [26]. The categorization of d is as follows: d<0.2 no effect, d<0.5 small effect, d<0.8 moderate effect and d≥0.8 large effect.

The analyses of the two Likert items (*"I talk with fellow students"* and *"I talk with my mentor about my results on the BPT"*) were recoded so that statements of 4 or 5 were considered as agreement.

Qualitative and quantitative methods were used to evaluate the open questions about the costs and benefits of the BPT. The data from the evaluation questionnaire were analysed in three steps: First, two authors (KG, MZ) examined all comments on the open questions and coded them independently of each other using the thematic content analysis [27]. In a second step, after joint discussion of discrepancies and new perspectives, these codes were again independently grouped into categories and a category list was created. In the third step, this category list was checked for inter-coder reliability with perfect matches (100% each) for the open questions 1 (8 categories), 3 (7 categories) and 4 (4 categories). Very good matches were found for the open questions 2 (94%, 9 categories) and question 5 (97%, 12 categories), so that this category list was used in the further analyses. The number of entries per category is given in the results section. The corresponding percentages refer to the total number of mentions for the given question.

Results

Descriptives

415 students at F1 and 453 students at F2 took part in the survey. 234 students at F1 answered the questions included in the analysis (57% female, respondents=56% of the sample, 234/415). At F2, 248 students answered these questions (71% female; respondents=55% of the sample, 248/453). An overview can be found in table 2.

Table 2: Participants in the BPT survey

Characteristic	Gradation	Medical school 1	Medical school 2
Number of participants	Total	415 (100%)	453 (100%)
	of which responded*	234 (56%)	248 (55%)
Sex	Female	133 (57%)	176 (71%)
	Male	101 (43%)	72 (29%)

Note: *respective research question

The two universities differed in a statistically significant manner in the distribution of the sexes (χ^2=10.52, df=1, p<.001) with a higher proportion of women at F2, but not in the distribution of students in the pre-clinical and clinical phase of their undergraduate training (n. s.). No statistically significant effects were found in preliminary analyses, so that the variable sex was not included as a covariate in the evaluations.

Proportion of non-serious test takers on the BPT

Regardless of the approach to operationalisation, it is shown that at Faculty 1 the proportion of non-serious test takers is significantly higher than at Faculty 2.

At F1 there are N_{F1}=173/1,410 (12%) students who answered all questions on the BPT with "don't know" or not at all, at F2 there are N_{F2}=5/1,191 (<1%). This is a significant difference with $\chi^2(1)=142.20$; p<0.001 and a small effect of w=0.23.

On the questionnaire, the following average TESS values, which reflect the self-evaluated test effort, were calculated at F1 for 291/415 (70%) students: M_{F1}=2.51; SD_{F1}=1.08 and at F2 of 409/453 (90%) students M_{F2}=3.63, SD_{F2}=0.88. This difference is also significant with T(543.80)=14.68; p<0.001 and has a large effect of d=-1.19. The testing effort of the students at F2 was therefore significantly greater than at F1. If the test takers are categorized as serious vs. not serious, there are N_{F1}=52/415 (13%) and N_{F2}=3/453 (<1%) non-serious test takers. This difference is also significant with $\chi^2(1)=68.96$; p<0.001 and a moderate effect (w=0.31).

Perceived costs at faculties with different examination and feedback conditions

Overall, the students from F1 reported more frequently on perceived costs of the BPT. F2 received more positive, constructive comments than F1. Multiple answers were possible when answering Question 1: *"Do you feel emotionally impaired by the BPT?"* At F1 there were 55 responses (24% of the 234 respondents) to this question, of which 53% (29/55) were constructive. Of the 19 mentions (8% of the 248 respondents) at F2 who answered this question, 15/19 (79%) were constructive. Table 3 shows the allocation of responses to the individual categories per faculty.

Question 2 *"What disadvantages do you see in the BPT?"* resulted in 241 responses at F1 (103% of the 234 respondents; viz. some multiple responses), 43% of which (104/241) were constructive responses. At F2, 65/105 (62%) of 105 responses (42% of the 248 responders) were constructive. The allocation of responses to the individual categories per faculty is shown in table 4.

Perceived benefit at faculties with different examination and feedback conditions

163 (39%) of the students from F1 and 309 (68%) of the students from F2 talk to other people about their BPT results. 84 (20%) of the students from F1 agreed with

Table 3: Categorisation of the responses to the open question "Do you feel emotionally restricted by the BPT? In what way?"

	F1	F2	Sample comment
Number of comments[A]	55 (24%)	19 (8%)	
Of which more negative comments about BPT[B]	26 (47%)	4 (21%)	
- waste of time	13	0	*"It stresses me out wasting an afternoon with the BPT right before exams."*
- BPT gets on my nerves	10	4	*"You're annoyed/frustrated because you know so little at first."*
- circumstances	3	0	*"I could do something more meaningful with my one day off."*
Of which more positive or constructive comments about BPT[B]	29 (53%)	15 (79%)	
- result	11	7	*"The BPT shows how many deficits you have in your knowledge, and the closer you get to the exam, the more it hurts you if you do badly."*
- none at all	12	2	
- acceptance	4	1	*"The BPT is good for tracking knowledge, I don't have any extra learning sessions for the BPT."*
- comparison	2	5	*"A clear comparison of different universities and courses of study (traditional vs. model course of study)."*

Note: The number of responses to this question is given, [A]percentages refer to the number of total responses (Responder N_{F1}=234 or N_{F2}=248), [B]percentages refer to the number of comments for this one question.

Table 4: Categorisation of the responses to the open question "What disadvantages do you see in the BPT?"

	F1	F2	Sample comments
Number of comments[A]	241 (103%)[C]	105 (42%)	
Of which more negative comments about the BPT[B]	137 (57%)	40 (38%)	
- waste of time	50	13	*"superfluous; useless; pointless; a waste of time; causes confusion."*
- frustration	46	17	*"Frustrating if/when hardly any questions can be answered; excessive demands; frustration for low semesters or regular courses of study, because mainly clinical questions; demotivating; too little can be answered."*
- without consequences	41	10	*"Compulsory event; not taken seriously, therefore no statement; no comparability."*
Of which more constructive comments about the BPT[B]	104 (43%)	65 (62%)	
- test administration	59	26	*"Test format: too long, too many questions, too little time; question areas mixed up; questions of varying difficulty; questions repeated; points deducted for incorrectly answered questions (one no longer dares to tick); evaluation method; no preparation possible; no indication of how many questions one should be able to solve."*
- organisation	19	10	*"University planning: poor placement (e.g. weekend, too early in the semester); not integrated into the curriculum; poor implementation (noise, exchange of information)."*
- exerts pressure	9	12	*"Apply pressure, encourage competitive thinking and frustration."*
- no disadvantage	8	10	*"From home; no debriefing; results without explanation (no learning effect)."*
- not comparable to M2	7	6	*"Not comparable with 2nd state examamination (no "don't know option" in state examination, no case questions or picture questions on BPT), leads to misinterpretation of own knowledge".*
- no viewing of results	2	1	*"No good viewing of results possible; test results too late, too complicated, not at all; no viewing."*

Note: Indicated is the number of mentions for this question, [A]percentages refer to the number of total responses (Responder N_{F1}=234 or N_{F2}=248), [B]percentages refer to the number of comments for this one question, [C]percentages > 100% result from multiple responses per person.

the statement that they talked with their fellow students about their BPT results. At F2 this number was 147 (32%). The statement that they talked with their mentor about their own results on the BPT was agreed with by 4 (1%) of the F1 students and 16 (4%) of the F2 students. A total of 75 (18%) of the participating F1 medical students and 146 (32%) of the F2 students talked with others about their BPT results. The frequency of agreement on the two closed questions, as well as the allocation of mentions to the individual categories per faculty for the other persons (open Question 3), are listed in table 5.

Question 4 *"Do you use the results of the BPT for other purposes? If so, how?"* There were 72/234 responses from F1 (31% of respondents) and 33/248 responses from F2 (13% of respondents). Although there were more responses at F1 than at F2, a high percentage of the responses from F1 were more likely to be in categories with negative connotations (70/72 responses, 97%), compared to only 22/33 responses (67%) from F2 with more negative connotations, as documented in table 6.

In response to Question 5 *"What advantages do you see in the BPT?"* there were just over 200 responses from both faculties (F1 with 207/234, 88% and F2 with 202/248, 81% of the respondents). At F1 163/234 (79%) of the responses could be assigned to positive categories, at F2 198/248 (98%), as shown in table 7.

Table 5: Dialogue with other people about the results of the BPT

	F1	F2
Fellow students[A]	84 (20%)	147 (32%)
Mentor[B]	4 (1%)	16 (4%)
Number of comments[C]	75 (18%)	146 (32%)
- Friends	18	39
- Family	16	35
- Parents	13	47
- Partner	13	23
- Nobody	12	2
- other	3	0

Note: Indicated is the number of agreements to the statements
- A "I talk to fellow students about my BPT results."
- B "I talk to my mentor about my BPT results."
- C "I talk to other people about my BPT results. With..."

Percentages refer to the number of students at Medical School 1 and 2.

Discussion

Formative assessment is important as an essential part of the assessment for learning. If formative assessment is not graded, it may be perceived by students as having high costs and/or lower benefits compared to summative assessment. In these cases, the proportion of non-serious test takers may be high. The present study investigated

Table 6: Categorisation of the responses to the open question "Do you use the results of the BPT for other purposes? If so, how?"

	F1	F2	Sample comment
Number of comments[A]	72 (31%)	33 (13%)	
Of which more negative comments about the BPT[B]	70 (97%)	22 (67%)	
- no	67	21	
- resistance	3	1	*"I find it ridiculous to compare myself to other fellow students. What is this pseudo pressure build-up doing in our studies! This has nothing to do with a reasonable error culture!"*
Of which more positive or constructive comments about the BPT[B]	2 (3%)	11 (33%)	
- learning objectives	2	8	*"I look at the subjects I've been bad at and plan to change it."*
- applications	0	3	*"Submission to the German Academic Scholarship Foundation."*

Note: The number of responses for this question is given, [A]percentages refer to the number of total responses (Responder N_{F1}=234 or N_{F2}=248), [B]percentages refer to the number of comments for this one question.

Table 7: Categorisation of the responses to the open question "What advantages do you see in the BPT?"

	F1	F2	Sample comment
Number of comments[A]	207 (88%)	202 (81%)	
Of which more negative comments about the BPT[B]	44 (21%)	4 (2%)	
- no advantage	20	1	*"No advantage!"*
- no advantage in pre-clinical phase	12	1	*"No pre-clinical advantage, rather frustrating; only useful for clinical semesters."*
- results applicable	5	1	*"No advantage, because questions not (completely) visible, no checking/reworking at home; no good evaluation available."*
- no incentive	4	0	*"Just a university requirement; no incentive to get serious about the test."*
- no IMPP questions	2	0	*"No IMPP questions."*
Of which more positive or constructive comments about the BPT[B]	163 (79%)	198 (98%)	
- feedback	109	128	*"(Regular) feedback/assessment of own knowledge level or progress; see what has already been learned, feedback on the status in individual subjects, on the learning strategy; see what still needs to be learned; presentation of exam knowledge; already existing knowledge is queried."*
- comparison with fellow students	21	32	*"Comparison with other students (fellow students, other universities, other countries), conclusions on the quality of teaching at one's own university; comparison of universities/programme models."*
- reduce exam fears	14	23	*"Preparation/exercise (for the 2nd state examination); clinical questions; getting to know questions and examination duration; reducing exam fears."*
- learning motivation	14	7	*"Motivation to learn, sense of achievement, learning effect."*
- assessment without pressure	5	6	*"Exam without test pressure/negative consequences."*
- other assessment	0	2	*"Other, non-university examination; objective, anonymous comparison, other question writers."*

Note: The number of responses for this question is given, [A]percentages refer to the number of total responses (Responder N_{F1}=234 or N_{F2}=248), [B]percentages refer to the number of comments for this one question.

whether different framework conditions at two faculties have an influence on the test effort and the perceived costs and benefits of a formative assessment – the Berlin Progress Test (BPT). The different framework conditions can be found in the required number of participations in the BPT during undergraduate training, the presentation of the BPT, the feedback on the results, as well as the university's implementation. Although both medical faculties are implementing measures to increase the acceptance of the BPT in order to increase test effort, the BPT is better accepted by students at Faculty 2 than at Faculty 1, as evidenced by the higher proportion of serious test takers, the lower perceived costs and higher reported

benefits, and the greater proportion of constructive comments.

Serious test taking

The hypothesis *"The proportion of non-serious test takers at Faculty 1 is lower than at Faculty 2"* could not be confirmed. Contrary to this hypothesis, the proportion of serious test takers at Faculty 1 is lower than at Faculty 2, despite more choices and immediate feedback. Although it has been shown elsewhere that the proportion of serious respondents is higher in computer-based administration than in paper-based administration [28], several studies have already shown that several factors influence test effort. Therefore, unicentric studies can only make a marginal contribution to the explanation of the multifactorial conditions for the test effort on formative tests.

Costs

The present study was able to confirm the hypothesis "The *perceived costs of the BPT are higher at Faculty 1 than at Faculty 2."*. The comments of the participants reflect findings from the literature that the costs of the BPT are perceived as high if the students estimate that they cannot simultaneously perform higher rated alternatives, such as learning for "real" assessment or if they feel emotional stress when filling out the test [29].

Benefits

The results for testing the hypothesis "*The perceived benefit of the BPT is lower at Faculty 1 than at Faculty 2"* must be considered in a more differentiated way. Although more students at Faculty 2 talk about their BPT results, half of those people are from outside the faculty. This is surprising because the BPT should be part of the undergraduate training and therefore students would be expected to talk mainly with their fellow students and mentors about their results. However, a mentor was rarely mentioned in answering this question, although a mentoring programme is available at F2. When asked whether students use the BPT results for other purposes, the proportion of comments made by students at Faculty 1 was higher than that of students at Faculty 2. However, this is a very high proportion of comments with negative connotations or comments that show that the results are not used for other purposes. Faculty 2 students have a better understanding of the principle behind formative testing and use the BPT results as feedback on their own knowledge progress, motivation to learn and reduction of exam fear. Although the attitude towards the BPT is more positive at Faculty 2, students at both faculties rarely mentioned that they use the BPT as a learning tool (10 mentions out of a total of 482 students who completed the questionnaire). The effect on learning is therefore questionable. However, this would be a quality criterion for the utility of an assessment method [4], especially in formative assessment where the function of the assessment is to stimulate and provide feedback on learning. The learning effect must be investigated more closely in further studies, especially since the effect on learning is questionable even in the case of moderate-stakes progress tests. Only a moderate role of the progress test in identifying strengths and weaknesses could be identified [30]. Aarts et al. showed that a majority of students used the results of the moderate-stakes progress test to monitor their knowledge, but it was not clear whether this also had a direct influence on learning [31]. This was also shown by Given et al. They found in semi-structured interviews that, although the students felt informed about their strengths and weaknesses, the feedback had no influence on future learning [32]. Yielder et al. also found with focus groups that in younger students, future learning is influenced by the progress test, but not by the feedback, rather by the content of the test [33]. Students in advanced semesters are more likely to use the progress test as a reminder that they need to learn at all. The proportion of comments on the benefits of the BPT in the present study is roughly comparable at both faculties, but it is also apparent that students at Faculty 2 are more positive about the BPT. It can therefore be concluded that the hypothesis on the perceived benefit of the BPT can be confirmed, but to limit this, its effect as a learning instrument is also questionable at Faculty 2.

Strengths and weaknesses

In the present study it could be shown that different conditions of assessment and feedback can be associated with different proportions of serious test takers and thus with an increased variance in test efforts. It also showed that the costs and benefits of the progress test are perceived differently at the two faculties. Faculty 2 not only had more serious test takers, but the BPT was also perceived more positively in terms of costs and benefits than at Faculty 1.

The advantage of the present study is the direct comparison of two medical faculties where the BPT was introduced at the same time more than 15 years ago. The conditions at both faculties are comparable in many respects: both have a model study programme and three licensing state examinations, which can have an influence on the BPT results [34]. Both faculties have comparable implementation conditions for the BPT, such as the same test, mandatory participation and no admission to further courses if the BPT is not taken. On the other hand, the two conditions for the implementation of the BPT differ in their different integration into the quality management system vs. into the assessment system and in the feedback (immediate feedback of results in the case of computer-based testing vs. comparison with the solution booklet on request). In addition to the comparable conditions at the two faculties, the present study offers the mixed-method approach as a further methodological advantage, which allows both quantitative and qualitative analyses. Thus a better insight into the perception of the BPT at the two faculties was gained and it could also be

shown quantitatively that the percentage of serious test takers differs greatly between the two faculties.
The methods used to determine the proportion of serious and non-serious test takers each have limitations in their validity, such as an unknown degree of sensitivity/specificity ("objective criteria") and the questionable significance of the self-reports (TESS score). In order to increase the validity of the results, triangulation was used to measure the test effort with different methods. Since both methods lead to the same conclusion, it can be assumed that the test effort is higher at Faculty 2 than at Faculty 1. Furthermore, the answers from the open questions also allow this conclusion to be drawn, since more constructive answers were given at Faculty 2 and also a higher benefit and lower costs were reported. According to the Expectancy Value Theory, the motivation to complete this task, meaning the test effort on the BPT, should therefore be higher at Faculty 2 than at Faculty 1.

Conclusion

The formative BPT as an assessment for learning is intended to give students feedback on the amount of their own medical knowledge, compared to the level at which they will graduate and compared to fellow students of the same level of undergraduate training, in order to accompany and modulate the learning process in the context of continuous learning. It is intended to be an antithesis to bulimic learning, which can occur more frequently due to too many summative assessments [2]. As with other low-stakes tests, there are large variances in test effort on the BPT and thus a questionable effect on learning. It can be assumed that measures to reduce the perceived costs and increase the perceived benefit can positively influence test effort and, in the long term, the effect on learning. Even if there is presumably no problem with the test effort on moderate-stakes progress tests, studies show the limited impact on learning. Therefore, framework conditions should be identified which positively influence the perceived costs and benefits of formative assessment and thus have a long-term effect on the learning process. Since the BPT provides data for feedback on the student's knowledge status as well as the learning progress, but the use of the BPT as a learning tool is up to the students, the BPT and the use of the results for their own learning should be embedded in the curriculum. This can be done by embedding the BPT in the assessment system, both as part of the assessment regulations and in the presentation of information and results, as at Faculty 2. Further possibilities for influencing perceived costs and benefits at a faculty would be to avoid contemporaneous summative assessment during formative assessment phases [6], [8], integration in the mentoring system for all students and not only as identification for the necessary support of underachieving students [13], [30], [31], [35], [36], [37]. It would also be conceivable to use formative assessment to develop and follow up learning plans together with the mentor [38]. If formative assessment is used to provide continuous feedback on knowledge, discussed with the mentor and serves to orient future learning, as envisaged in the programmatic assessment [8], [39], then it will serve its purpose. And only then students will see the value of formative assessment.
Although formative assessment is becoming increasingly important, it is not enough to introduce it as an add-on to the curriculum. Rather, new assessment formats also require the appropriate framework conditions to achieve the desired effect. In formative assessment, therefore, conditions must be created in which the results have a value, both as a guide through the undergraduate training and as guidance for learning behaviour. Only if equal importance is attached to formative and summative assessment will the perceived costs and benefits be comparable along with the test effort. Thus, the focus of students can be shifted to continuous learning, away from bulimic learning, because it can be assumed that students who focus their actions on merely passing MC exams will not be able to recognize the value of formative assessment at all.
Low-stakes assessment is a good way to learn under what conditions assessment for learning works and how it can be effectively embedded in existing curricula. Therefore, further studies should investigate the extent of the individual measures and their interaction. This is a great challenge since the investigation of real conditions in medical education is made difficult by many, often uncontrollable conditions [40].

Competing interests

The authors declare that they have no competing interests.

References

1. Schuwirth LW, van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478-485. DOI: 10.3109/0142159X.2011.565828
2. Schuwirth LW, van der Vleuten CP. The use of progress testing. Perspect Med Educ. 2012;1(1):24-30. DOI: 10.1007/s40037-012-0007-2
3. Berkhout JJ, Helmich E, Teunissen PW, van der Vleuten CP, Jaarsma AD. Context matters when striving to promote active and lifelong learning in medical education. Med Educ. 2018;52(1):34-44. DOI: 10.1111/medu.13463
4. van der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ Theory Pract. 1996;1(1):41-67. DOI: 10.1007/BF00596229
5. Wade L, Harrison C, Hollands J. Student perceptions of the progress test in two settings and the implications for test deployment. Adv Health Sci Educ Theory Pract. 2012;17(4):573-583. DOI: 10.1007/s10459-011-9334-z

6. Cobb KA, Brown G, Jaarsma DA, Hammond RA. The educational impact of assessment: a comparison of DOPS and MCQs. Med Teach. 2013;35(11):e1598-e1607. DOI: 10.3109/0142159X.2013.803061
7. van Berkel HJ, Nuy HJ, Geerlings T. The influence of progress tests and block tests on study behaviour. Instruct Sci. 1995;22(4):317-333. DOI: 10.1007/BF00891784
8. Heeneman S, Schut S, Donkers J, van der Vleuten CP, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. Med Teach. 2017;39(1):44-52. DOI: 10.1080/0142159X.2016.1230183
9. Wigfield A, Eccles JS. Expectancy-value theory of achievement motivation. Contemp Educ Psychol. 2000;25(1):68-81. DOI: 10.1006/ceps.1999.1015
10. Gast L. "Kein Ort. Nirgends?" Das Subjekt der Erkenntnis und die Idee der Universität. Einige Gedanken aus psychoanalytischer Perspektive. Psychol Gesellschaftskritik. 2010;33/34(4/1):153-171.
11. Zeigarnik BV. Das Behalten erledigter und unerledigter Handlungen. Psychol Forsch. 1927;9:1-85.
12. Albano MG, Cavallo F, Hoogenboom R, Magni F, Majoor G, Manenti F, Schuwirth L, Stiegler I, van der Vleuten C. An international comparison of knowledge levels of medical students: the Maastricht Progress Test. Med Educ. 1996;30(4):239-45. DOI: 10.1111/j.1365-2923.1996.tb00824.x
13. van der Vleuten CP, Verwijnen GM, Wijnen WH. Fifteen years of experience with progress testing in a problem-based learning curriculum. Med Teach. 1996;18(2):103-109. DOI: 10.3109/01421599609034142
14. Nouns ZM, Georg W. Progress testing in German-speaking countries. Med Teach. 2010;32(6):467-470. DOI: 10.3109/0142159X.2010.485656
15. Osterberg K, Kölbel S, Brauns, K. Der Progress Test Medizin: Erfahrungen an der Charité Berlin. GMS Z Med Ausbild. 2006;23(3):Doc46. Zugänglich unter/available from: https://www.egms.de/static/de/journals/zma/2006-23/zma000265.shtml
16. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: Classic definitions and new directions. Cont Educ Psychol. 2000;25:54-67. DOI: 10.1006/ceps.1999.1020
17. Hackman JR, Oldham GR. Motivation through the design of work: Test of a theory. Organ Behav Hum Perform. 1976;16(2):250-279. DOI: 10.1016/0030-5073(76)90016-7
18. Kulik JA, Kulik CLC. Timing of feedback and verbal learning. Rev Educ Res. 1988;58(1):79-97. DOI: 10.3102/00346543058001079
19. Tuten TL, Galesic M, Bosnjak M. Effects of immediate versus delayed notification of prize draw results and announced survey duration on response behavior in web surveys: An experiment. Soc Sci Comput Rev. 2004;22(3):377-384. DOI: 10.1177/0894439304265640
20. Irons A. Enhancing learning through formative assessment and feedback. London: Routledge Taylor & Francis Group; 2008. DOI: 10.4324/9780203934333
21. Nicol D, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006;31(2):199-218. DOI: 10.1080/03075070600572090
22. Smyth K. The benefits of students learning about critical evaluation rather than being summatively judged. Ass Eval High Educ. 2004;29(3):369-377. DOI: 10.1080/0260293042000197609
23. Schüttpelz-Brauns K, Kadmon M, Kiessling C, Karay Y, Gestmann M, Kämmer JE. Identifying low test-taking effort during low-stakes tests with the new Test-taking Effort Short Scale (TESS) - Development and Psychometrics. BMC Med Educ. 2018;18(1):101. DOI: 10.1186/s12909-018-1196-0
24. Brauns K. Identifikation von Musterkreuzern beim Progress Test Medizin. [Dissertation]. Berlin: Humboldt-Universität zu Berlin; 2007.
25. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum; 1988.
26. Leonhart R. Effektgrößenberechnung bei Interventionsstudien. Reha. 2004;43:241-246. DOI: 10.1055/s-2004-828293
27. Mayring P. Qualitative Inhaltsanalyse - Grundlagen und Techniken. 11. Auflage. Weinheim, Basel: Beltz Verlag; 2008.
28. Karay Y, Schauber SK, Stosch C, Schuettpelz-Brauns K. Can computer-based assessment enhance the acceptance of formative multiple choice exams? A utility analysis. Med Teach. 2012;34:292-296. DOI: 10.3109/0142159X.2012.652707
29. Flake JK, Barron KE, Hulleman C, McCoach BD, Welsh ME. Measuring cost: The forgotten component of expectancy-value theory. Contemp Educ Psychol. 2015;41:232-244. DOI: 10.1016/j.cedpsych.2015.03.002
30. Blake J, Norman GR, Keane DR, Mueller B, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: Psychometric properties and effect on learning. Acad Med. 1996;71(9):1002-1007. DOI: 10.1097/00001888-199609000-00016
31. Aarts R, Steidel K, Manuel BAF, Driessen EW. Progress testing resource-poor countries: A case from Mozambique. Med Teach. 2010;32(6):461-463. DOI: 10.3109/0142159X.2010.486059
32. Given K, Hannigan A, McGrath D. Red, yellow and green: What does it mean? How the progress test informs and supports student progress. Med Teach. 2016;38(10):1025-1032. DOI: 10.3109/0142159X.2016.1147533
33. Yielder J, Wearn A, Chen Y, Henning M, Weller J, Lillis S, Mogol V, Bagg W. A qualitative exploration of student perceptions of the impact of progress tests on learning and emotional wellbeing. BMC Med Educ. 2017;17(1):148. DOI: 10.1186/s12909-017-0984-2
34. Nouns ZM, Schauber S, Witt C, Kingreen H, Schüttpelz-Brauns K. Development of knowledge in basic medical sciences during undergraduate medical education-A comparison of a traditional and a problem-based curriculum. Med Educ. 2012;46(12):1206-1214. DOI: 10.1111/medu.12047
35. Coelho C, Zahra D, Ali K, Tredwin C. To accept or decline academic remediation: What difference does it make? Med Teach. 2019;41(7):824-829. DOI: 10.1080/0142159X.2019.1585789
36. Lillis S, Yielder J, Mogol V, O'Connor B, Bacal K, Booth R, Bagg W. Progress testing for medical students at the University of Auckland: Results from the first year of assessments. J Med Educ Curr Dev. 2014;1:41-45. DOI: 10.4137/JMECD.S20094
37. Norman G, Neville A, Blake J, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. Med Teach. 2010;32(6):496-499. DOI: 10.3109/0142159X.2010.486063
38. Kastenmeier AS, Redlich PN, Fihn C, Treat R, Chou R, Homel A, Lewis BD. Individual learning plans foster self-directed learning skills and contribute to improved educational outcomes in the surgery clerkship. Am J Surg. 2018;216(1):160-166. DOI: 10.1016/j.amjsurg.2018.01.023
39. Schuwirth LW, van der Vleuten CP. Current assessment in medical education: Programmatic assessment. J Appl Test Technol. 2019;20(S2):2-10.

40. Ringsted C, Hodges B, Scherpbier A. 'The research compass': An introduction to research in medical education: AMEE Guide No. 56, Med Teach. 2011;33(9):695-709. DOI: 10.3109/0142159X.2011.595436

Corresponding author:
Dr. rer. nat. Katrin Schüttpelz-Brauns
Medical Faculty Mannheim at Heidelberg University, Department of Undergraduate Education and Educational Development, Theodor-Kutzer-Ufer 1-3, D-68167 Mannheim, Germany, phone: +49 (0)621/383-71270, fax: +49 (0)621/383-71201
katrin.schuettpelz-brauns@medma.uni-heidelberg.de

Please cite as
Schüttpelz-Brauns K, Karay Y, Arias J, Gehlhar K, Zupanic M. Comparison of the evaluation of formative assessment at two medical faculties with different conditions of undergraduate training, assessment and feedback. GMS J Med Educ. 2020;37(4):Doc41. DOI: 10.3205/zma001334, URN: urn:nbn:de:0183-zma0013341

This article is freely available from
https://www.egms.de/en/journals/zma/2020-37/zma001334.shtml

Received: *2019-08-29*
Revised: *2020-03-10*
Accepted: *2020-04-27*
Published: *2020-06-15*

Copyright
©2020 Schüttpelz-Brauns et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at http://creativecommons.org/licenses/by/4.0/.

OPEN ACCESS

This is the German version.
The English version starts at p. 1.

Vergleich der Bewertung einer formativen Prüfung an zwei medizinischen Fakultäten mit unterschiedlichen Studien-, Prüfungs- und Feedbackbedingungen

Katrin Schüttpelz-Brauns[1]
Yassin Karay[2]
Johann Arias[3]
Kirsten Gehlhar[4]
Michaela Zupanic[5]

1 Medizinische Fakultät Mannheim der Universität Heidelberg, Mannheim, Deutschland

2 Universität zu Köln, Medizinische Fakultät, Köln, Deutschland

3 RWTH Aachen, Medizinische Fakultät, Aachen, Deutschland

4 Carl von Ossietzky Universität Oldenburg, Fakultät für Medizin und Gesundheitswissenschaften, Oldenburg, Deutschland

5 Private Universität Witten/Herdecke gGmbH, Fakultät für Gesundheit, Witten, Germany

Zusammenfassung

Einführung: Sowohl formative als auch summative Prüfungen haben ihre Berechtigung in medizinischen Curricula: formative Prüfungen zur Begleitung des Lernprozesses und summative Prüfungen zur Sicherung des Erreichens von Mindeststandards. Je nach Studien-, Prüfungs- und Feedbackbedingungen wird formativen Prüfungen durch die Studierenden mehr oder weniger Bedeutung beigemessen und entsprechend kann die Erfüllung ihrer Funktion fraglich sein. In dieser Studie wird beschrieben, wie der nicht-bestehensrelevante formative Progress Test Medizin (PTM) an zwei Medizinischen Fakultäten mit partiell unterschiedlichen Rahmenbedingungen eingebettet ist und welche Auswirkungen diese auf das Testbemühen der Studierenden und die Bewertung des Tests, insbesondere der Wahrnehmung von dessen Nutzen und (immateriellen) Kosten, wie Nicht-Teilnahme an zeitgleichen Angeboten oder emotionale Beeinträchtigungen, haben.
Methoden: In dieser Studie wurde der Anteil der nicht ernsthaften PTM-Teilnehmenden an zwei Medizinischen Fakultäten (Gesamtstichproben: N_{F1}=1.410, N_{F2}=1.176) im WS 15/16 sowohl durch die Zahl nicht beantworteter Fragen im Test selbst als auch im Rahmen einer Befragung mit einem standardisierten Instrument (N_{F1}=415, N_{F2}=234) bestimmt. Weiterhin wurden in dieser Befragung offene Fragen zum wahrgenommenen Nutzen bzw. den wahrgenommenen Kosten gestellt, welche mit qualitativen und quantitativen Methoden ausgewertet wurden.
Ergebnisse: Der PTM wird an Fakultät 2 insgesamt besser angenommen. Dies zeigt sich in dem höheren Anteil ernsthafter Testteilnehmenden, den niedrigeren wahrgenommenen Kosten und dem höheren berichteten Nutzen sowie dem größeren Anteil an konstruktiven Kommentaren. Studierende der Fakultät 2 haben das Prinzip des formativen Prüfens besser verstanden und nutzen die Ergebnisse des PTM als Feedback über den eigenen Wissensfortschritt, zur Lernmotivation und zur Reduktion von Prüfungsangst.
Diskussion: Wenn Medizinische Fakultäten formative Prüfungen in das Curriculum integrieren, müssen sie Rahmenbedingungen schaffen, in denen diese Prüfungen als wichtiger Teil des Curriculums wahrgenommen werden. Ansonsten ist es fraglich, ob sie ihrer Funktion der Begleitung des Lernprozesses gerecht werden können.

Schlüsselwörter: Formative Prüfungen, Medizinische Ausbildung, Progress Testing, Testbemühen

Einleitung

Nach der Ärztlichen Approbationsordnung (ÄAppO) §2 Absatz 7 Satz 1 muss die erfolgreiche Teilnahme im vorklinischen Abschnitt mit 17 Leistungsnachweisen (Anlage 2a) und im klinischen Abschnitt mit 40 Leistungsnachweisen (Anlage 2b) nachgewiesen werden. Dieser Nachweis erfolgt entweder mit benoteten Prüfungen, welche das Lernergebnis eines Abschnittes, wie ein Fach oder ein Modul abprüfen oder mit der Bewertung bestanden/nicht bestanden. Daher handelt es sich hierbei um das Prüfen des Lernens bzw. summative Prüfungen [1]. Demgegenüber stehen Prüfungen, welche den Lernprozess begleiten. Diese formativen Prüfungen [1] fördern das kontinuierliche und vertiefende Lernen [2]. Feedback ist ein zentraler Aspekt des kontinuierlichen Lernens, indem Lücken im Lernen identifiziert und gezielt korrigiert werden. Das kontinuierliche Lernen bereitet auf das lebens-

lange Lernen vor, welches aufgrund der Schnelllebigkeit von Wissen und dem ständigen Wechsel von Anforderungen immer mehr verlangt wird [3]. Es gibt bereits einige Studien zu formativen Prüfungen, welche die Wirkung auf das Lernen untersuchen. Dieser sog. Educational Impact ist Teil des Nützlichkeitsmodells für Prüfungen [4] und kann als Indiz dafür gesehen werden, dass die Wirkung auf das kontinuierliche Lernen durch die formative Prüfung bzw. deren Feedback gegeben ist. Wade et al. entwickelten einen Fragebogen, um die Wahrnehmung von Progress Tests – einer Art der formativen Prüfungen (s. u.) – als Lernwerkzeug an zwei verschiedenen medizinischen Fakultäten zu vergleichen und fanden, dass die Lernumgebung einen Einfluss auf die Wertschätzung von Progress Tests als Unterstützung des Lernens hat [5]. Cobb et al. befragten Studierende in semi-qualitativen Interviews zu ihrer Wahrnehmung von DOPS (formativen Direct Observation of Procedural Skills) im Vergleich zu MCQ (summativen Multiple-Choice-Prüfungen) und fanden, dass die formative Prüfung tieferes Lernen förderte, aber die summative Prüfung für die Studierenden wichtiger war [6]. In einer Fragebogenstudie an der Fakultät für Gesundheitswissenschaften in Maastricht wurden summative Blocktests von den Studierenden als lohnender empfunden und die Ergebnisse des Progress Tests nicht für selbstreguliertes Lernen verwendet [7]. Die Einbettung von Feedback durch/mit Progress Tests in ein umfassendes Prüfungsprogramm erhöhte die Verwendung des Progress-Test-Feedbacktools durch die Studierenden und die Integration in das Lernen [8].

Sowohl das kontinuierliche Begleiten des Lernprozesses durch formative Prüfungen als auch das Sicherstellen des Erreichens von Mindeststandards in Form von summativen Prüfungen haben ihre Berechtigung im medizinischen Curriculum.

Es ist anzunehmen, dass die beiden Zielrichtungen des Lernens – Lernen für die Prüfung vs. Prüfungen für das Lernen – von Seiten der Studierenden unterschiedlich intensiv verfolgt werden, welches sich in dem unterschiedlichen Ausmaß an Testbemühen zeigt. Dies lässt sich mit der Erwartungs-Wert-Theorie von Wigfield & Eccles [9] erklären. Die Erwartungs-Wert-Theorie besagt, dass die Motivation zur Erledigung einer Aufgabe von zwei Komponenten abhängig ist: die Erwartung, eine Aufgabe lösen zu können und dem Wert, den diese Aufgabe für das Individuum hat. Wigfield & Eccles [9] unterscheiden vier verschiedene Komponenten, die diesen Wert ausmachen können:

- Leistungswert (die Aufgabe bestmöglich zu meistern),
- Intrinsischer Wert (der Spaß bzw. die Freude beim Lösen der Aufgabe),
- Nutzenwert (wie gut die Aufgabe in zukünftige Pläne passt, also wie nützlich die Aufgabe ist),
- Kosten (Kosten im eigentlichen Sinne, inwiefern Aktivitäten miteinander konkurrieren, aber auch emotionale Kosten).

In Bezug auf summative und formative Prüfungen gibt es einen Unterschied bei dem Wert, der einer Aufgabe, der Prüfung, beigemessen wird. Welchen Wert eine Aufgabe bzw. Prüfung für jede/n Einzelne/n hat, wird auch durch die Rahmenbedingungen an der eigenen Fakultät beeinflusst. Diese hat maßgeblichen Einfluss auf den wahrgenommenen Nutzen und die wahrgenommenen Kosten. Zum Beispiel ist es sehr wahrscheinlich, dass die summativen Prüfungen im Fokus der Studierenden stehen, wenn an der eigenen medizinischen Fakultät die nach ÄAppO vorgeschriebenen Nachweise ausschließlich durch summative Prüfungen nachgewiesen werden müssen. Im ungünstigsten Fall lernen sie äußerst effizient, d. h. alle jeweils geforderten Inhalte kurz vor der entsprechenden Prüfung, das sog. Bulimielernen [10]. Damit wäre aus Sicht der Studierenden der Nutzen - Bestehen der Prüfungen - maximal bei minimalen Kosten. Dieses so erworbene Wissen läuft Gefahr nach der Prüfung „abgehakt“ und bald vergessen zu werden [11].

Gerade was Nutzen und Kosten aus der Erwartungs-Wert-Theorie betrifft, können Fakultäten Rahmenbedingungen schaffen, um die Motivation zur Nutzung von formativen Prüfungen und damit den Einfluss auf das Lernen zu erhöhen. So können formative Prüfungen als zusätzlicher Aufwand gesehen werden, v. a. wenn summative Prüfungen und/oder arbeitsintensive Lehrveranstaltungen (study load) parallel absolviert werden müssen. Wenn dagegen die formative Prüfung als sinnvoller und wertvoller Bestandteil des Gesamtcurriculums wahrgenommen wird und von den Fakultätsangehörigen wertgeschätzt wird, könnte der Nutzen der formativen Prüfung – trotz paralleler benoteter Prüfungen und hoher study load – als hoch angesehen werden.

Eine Möglichkeit zu untersuchen, unter welchen Bedingungen formative Prüfungen – trotz konkurrierender und damit als nützlicher wahrgenommene – summativer Prüfungen erfolgreich implementiert werden können, bietet der formative Progress Tests Medizin.

Progress Tests sind Multiple-Choice-Tests, die das medizinische Wissen einer Berufsanfängerin/eines Berufsanfängers regelmäßig im Verlauf des Studiums erfassen und mit dem Wissensstand der Kommiliton/innen des gleichen Semesters vergleichen, um Lücken im aktuellen Wissensstand zu identifizieren und um das Lernverhalten konstruktiv zu beeinflussen. Alle Arten von Progress Tests geben Feedback, werden jedoch unterschiedlich bzgl. ihrer Bestehensrelevanz verwendet. Im niederländischen Konsortium und in den U.S.A. werden z. B. die Ergebnisse des Progress Tests pro Individuum über mehrere Testzeitpunkte akkumuliert und sind in dieser Form bestehensrelevant [12], [13]. Das heißt, die Progress Tests werden zwar nicht benotet, haben aber dennoch einen Einfluss auf den Fortgang im Studium. In Deutschland und Österreich sind die Teilnahmen zwar verpflichtend, jedoch nicht bestehensrelevant [2], [14]. Im deutschsprachigen Konsortium Progress Test Medizin (PTM) erfolgt die Testerstellung und Auswertung zentral an der Charité-Universitätsmedizin in Berlin. Alle Teilnehmenden des Tests erhalten aus Berlin ca. 4-6 Wochen nach dem Test eine ausführliche Rückmeldung ihrer Ergebnisse im Verlauf der Jahre, aber auch im Vergleich mit ihren Kommiliton/innen,

differenziert nach Organsystemen und Fächern. Das unterschiedliche Ausmaß an Testbemühen an den einzelnen Fakultäten zeigt sich im Anteil ernsthafter Testteilnehmender, der routinemäßig nach jeder Testdurchführung ermittelt wird. Beim nicht bestehensrelevanten PTM zeigt sich, dass es sehr unterschiedliche Anteile an ernsthaften Testteilnehmenden an den verschiedenen Fakultäten gibt. So wurden Anteile von 75-90% an den teilnehmenden Fakultäten berichtet [15].

In der vorliegenden Studie wird untersucht, wie der nicht-bestehensrelevante PTM an zwei Medizinischen Fakultäten eingebettet ist und welchen Einfluss das auf das Testbemühen der Studierenden und die Wahrnehmung des Progress Tests, spezieller die Wahrnehmung der Kosten sowie des Nutzens, als formative Prüfung hat. Die Rahmenbedingungen für den PTM unterscheiden sich in beiden Fakultäten unter anderem in der Einbindung in das jeweilige Curriculum: Die Studien-, Prüfungs- und Feedbackbedingungen im Vergleich sind in Tabelle 1 detailliert dargestellt.

Wenn man die Prüfungs- und Feedbackbedingungen an beiden Fakultäten betrachtet, sollte der Anteil ernsthafter Testteilnehmender und die damit einhergehende Wahrnehmung der Kosten und des Nutzens des PTM vergleichbar sein, da beide Fakultäten Bedingungen aufweisen, welche positiv auf die Motivation und entsprechend auf das Testbemühen wirken sollten.

So haben die Studierenden an der Fakultät 1 eine Wahlmöglichkeit, die an Fakultät 2 nicht vorhanden ist. Sie können wählen, an welchen 8 von 10 PTM sie teilnehmen möchten. Nach der Selbstbestimmungstheorie von Ryan & Deci [16] sollte dies die intrinsische Motivation erhöhen und sich damit auch der *Anteil ernsthafter Testteilnehmender* erhöhen.

Außerdem gibt es an der Fakultät 1 ein sofortiges Feedback durch die computer-basierte Administration. Sofortiges Feedback ist wichtig, damit man Aufgaben beendet und mit der Arbeit zufrieden ist [17], [18], [19]. Daher sollte sich auch durch die Bedingung der computer-basierten Administration das Testbemühen erhöhen und damit auch der Anteil ernsthafter Testteilnehmender.

Obwohl die Rückmeldung der Ergebnisse sehr schnell erfolgt, gibt es keinen Dialog über die Ergebnisse an Fakultät 1. Der Dialog über die Ergebnisse ist an der Fakultät 2 in das Mentorenprogramm eingebunden. Der Dialog ist essentiell für ein effektives Feedback und damit für das Funktionieren formativer Prüfungen [20], [21], [22]. Die sollte den *wahrgenommenen Nutzen* des PTM an Fakultät 2 erhöhen.

Da an Fakultät 2 der PTM als Prüfung kommuniziert wird, im Gegensatz zu Fakultät 1, in der er als Evaluation präsentiert wird, sollte der PTM an Fakultät 2 aus einem weiteren Grund als nützlicher wahrgenommen werden. Wie Heeneman et al. in ihrer Studie zeigen konnten, nutzen Studierende das Feedbacksystem des bestehensrelevanten Progress Tests mehr und haben höhere Testwerte, wenn der Progress Test in ein ganzheitliches Prüfungssystem eingebunden ist [8]. Die höheren Testwerte wurden dabei als indirekter Indikator für Testbemühen angesehen. Gleichzeitig werden die *wahrgenommenen Kosten* niedriger, wenn die formative Prüfung Teil des Prüfungssystems ist.

Unter Berücksichtigung der Bedingungen an den beiden Fakultäten und deren theoretischen Einflüsse auf das Testbemühen, gemessen am Anteil ernsthafter vs. nicht ernsthafter Testteilnehmender, sowie auf die wahrgenommenen Kosten und Nutzen des PTM, lassen sich folgende Hypothesen ableiten:

1. *Der Anteil nicht ernsthafter Testteilnehmender an Fakultät 1 (F1) ist niedriger als an Fakultät 2 (F2).*
2. *Die wahrgenommenen Kosten des PTM sind an Fakultät 1 (F1) höher als an Fakultät 2 (F2).*
3. *Der wahrgenommene Nutzen des PTM ist an Fakultät 1 (F1) niedriger als an Fakultät 2 (F2).*

Methodik

Die Studie erfolgt in einem Mixed-Method-Ansatz, bei dem quantitativ die Anteile der nicht ernsthaften Testteilnehmenden bestimmt werden. Im qualitativen Teil werden die Themen identifiziert, die für die Studierenden bezogen auf den PTM in beiden Fakultäten bzgl. des wahrgenommenen Nutzen und der Kosten jeweils von Relevanz sind.

Stichprobe

Im Wintersemester 2015/16 nahmen N=1.410 (F1) bzw. N=1.176 (F2) Medizinstudierende am PTM teil. Das entspricht einem Anteil von 50 % der immatrikulierten Medizinstudierenden in F1 bzw. 61% in F2. Der Frauenanteil der Studierenden an den beiden Fakultäten beträgt 62% (F1) bzw. 68% (F2).

Material

Der Anteil nicht ernsthafter Testteilnehmender wurde auf zwei verschiedene Wege ermittelt. Zum einen wurden die Studierenden, die beim Ausfüllen des Tests im WiSe 2015/16 bei allen Fragen die „weiß nicht“-Option gewählt oder alle Fragen übersprungen haben, als nicht ernsthaft identifiziert, da selbst im 1. Semester die eine oder andere Frage beantwortet werden kann. Zum anderen wurde das Testbemühen mit Hilfe der Test-Effort Short Scale (TESS) [23] ermittelt. TESS besteht aus drei fünf-stufigen Likert-Items mit den Abstufungen 1 bis 5, die den Leistungswert („Ich möchte beim PTM die bestmöglichen Ergebnisse erreichen.“), den Nutzenwert („Ich finde den PTM sinnvoll.“) und die wahrgenommenen Kosten („Der PTM ist ein wertvoller Teil meines Studiums.“) erfragen. Aus den Antworten auf alle drei Fragen wird der Mittelwert berechnet. Studierende, die der jeweiligen Aussage nicht zustimmten und alle Fragen mit 1 beantworteten (entspricht einem TESS-Score von 1), werden als nicht ernsthafte Testteilnehmende kategorisiert. Beide Verfahren haben jeweils einen methodischen Nachteil, der die Validität mindern könnte. Der Nachteil bei der Selbstauskunft ist, dass es einen unbekannten Prozentsatz von

Tabelle 1: Studien-, Prüfungs- und Feedbackbedingungen an den beiden Fakultäten

Merkmal	Fakultät 1	Fakultät 2
Studiengang	Modellstudiengang	Modellstudiengang
Anzahl der Studierenden	2900	1250
Studienstruktur	Modellstudiengang mit 3 Studienabschnitten 1. Vorklinik, dann M1 2. Klinik, dann M2 3. Praktisches Jahr, dann M3	Modellstudiengang mit 4 Studienabschnitten 1. 1. und 2. Sem. 2. 3.-6. Sem., dann M1 3. 7.-10. Sem, dann M2 4. Praktisches Jahr, dann M3
PTM seit	WiSe 2003	WiSe 2003
Testformat	Computer-basiert	Papier-basiert mit Option der Computer-basierten Durchführung
Teilnahme	Ab 1. Semester Pflicht, in der VKL mind. 3x Teilnahme, in der Klinik mind. 5x Teilnahme (=8x pro Medizinstudium)	Teilnahmeverpflichtung ab 1. Sem. und ein PTM pro Semester als Zulassungsvoraussetzung in den nächsten Studienabschnitt. Im 3. Studienabschnitt nur eine Teilnahme im 8. oder 9. Semester (Wahlfreisemester), Planung vermeidet zeitgleiche Veranstaltungen
Konsequenzen	Bei Nicht-Teilnahme keine Zulassung zu weiteren Kursen	Bei Nicht-Teilnahme keine Zulassung zu weiteren Kursen/zum nächsten Studienabschnitt
Einbindung des PTM	In das Qualitätsmanagement (Evaluationsordnung)	In das Prüfungssystem (Prüfungsordnung)
Information über den PTM	Informiert werden die Studierenden per E-Mail über die Auswertung. Aufgeklärt über den PTM werden die Studierenden in den Einführungsveranstaltungen des ersten Semesters (Power-Point-Präsentation), danach nicht mehr.	Informationen über Termine erfolgt per E-Mail. Nutzen und Allgemeines über den PTM wird jedes Semester in einem Vortrag vor der jeweiligen Durchführung vermittelt.
Erhalt der Ergebnisse	Die Studierenden erhalten unmittelbar nach dem Test eine Schnellauswertung. Sie erhalten ihr Testergebnis und können ihre Antworten mit dem Lösungsschlüssel vergleichen. Am Ende des Semesters können sich die Studierenden die Berliner Auswertung herunterladen.	Papierbasierte Tests bieten ein Lösungsheft zum direkten Feedback an. PDF-Auswertung erfolgt über ein Online-Portal, darüber hinaus werden die PTM Ergebnisse in einem Feedbacktool („HIP-How I Perform“) veröffentlicht, in dem die PTM-Ergebnisse den summativen Kursergebnissen gegenübergestellt werden.
Dialog über die Ergebnisse	Keine	In der nächsten Gruppenvorstellung zu Beginn des Semesters werden die Ergebnisse des Vorsemesters kurz analysiert. Anhand der Ergebnisse finden ggfs. Mentorengespräche statt.

Studierenden gibt, die sozial erwünscht antworten. Das bedeutet, dass sie ein höheres Ausmaß an Testbemühen angeben könnten, als dies tatsächlich der Fall ist. Der Nachteil bei der Identifikation über die „weiß nicht“-Option liegt darin, dass es auch sog. Musterkreuzer geben kann. Dies sind Testteilnehmende, welche zwar alle Fragen beantworten, dies jedoch ohne Kenntnis des Fragentextes tun [24]. Aufgrund dieser Nachteile sind beide Verfahren parallel angewendet worden.

Um die wahrgenommenen Kosten und Nutzen messbar zu machen, haben wir offene Fragen gestellt. Sowohl der Begriff der Kosten als auch der Begriff des Nutzens sind sehr abstrakt. Daher haben wir formal balancierte Fragen gestellt, die Antwortmöglichkeiten provozieren, die sich diesen beiden Begriffen zuordnen lassen. Dies sind zum einen die Fragen nach Nachteilen und Vorteilen des PTM, aber auch Fragen direkt zum Nutzen des PTM. So sollten Studierende, welche den PTM nutzen, auch mit anderen Personen über ihre Ergebnisse reden, wie z. B. mit ihrem/ihrer Mentor/in, um mit ihm bzw. ihr das eigene Lernverhalten zu ändern.

Die wahrgenommenen Kosten wurden über zwei offene Fragen abgefragt:

- Fühlen Sie sich emotional durch den PTM eingeschränkt? (Frage 1)
- Welche Nachteile sehen Sie beim PTM? (Frage 2)

Der wahrgenommene Nutzen wurde über fünf Fragen (zwei geschlossene und drei offene Fragen) zu verschiedenen Aspekten ermittelt:

- Dialog mit anderen Personen über die Ergebnisse beim PTM mit den Unterfragen:
  - Ich rede mit Kommilitonen über meine Ergebnisse beim PTM. (Likert-Item mit 1 „trifft nicht zu“ bis 5 „trifft zu“),
  - Ich rede mit meinem Mentor über meine Ergebnisse beim PTM. (Likert-Item mit 1 „trifft nicht zu“ bis 5 „trifft zu“),
  - Ich rede mit anderen Menschen über meine Ergebnisse beim PTM. Und zwar mit ... (offene Frage, Frage 3).

- Nutzen Sie die Ergebnisse des PTM noch anderweitig? (offene Frage, Frage 4)
- Welche Vorteile sehen Sie beim PTM? (offene Frage, Frage 5)

Es gab keine Einschränkung hinsichtlich der Anzahl der Kommentare, welche die Studierenden zu den offenen Fragen abgeben konnten.
Zusätzlich wurden im Fragebogen das Geschlecht und die Studiensemester abgefragt, um die Vergleichbarkeit beider Gruppen prüfen zu können.

Versuchsdurchführung

An Fakultät 1 und 2 fand der PTM in den ersten Wochen des Semesters in den Räumlichkeiten der Hochschule und unter Aufsicht statt. Für jede Kohorte wurden mindestens zwei überschneidungsfreie Termine geplant, die von den Studierenden selbstständig gewählt werden konnten. An beiden Fakultäten erfolgte die Testung computerbasiert. An Fakultät 2 wurden zusätzlich Termine für eine papierbasierte Testung angeboten.
Die Studierenden an beiden Fakultäten nahmen regulär am PTM-Termin teil. Zu Beginn des Tests wurde neben der regulären Einführung auch über die Gesamtstudie informiert. Die Gesamtstudie untersucht die Motivation beim PTM und seinen Einfluss auf das Lernen beim PTM. Daher enthielt der Fragebogen mehr als die hier angegebenen Fragen. Die Teilnehmenden wurden in der regulären Einführung gebeten, den Fragebogen im Anschluss an die Testbearbeitung auszufüllen und wurden darauf hingewiesen, dass diese Teilnahme freiwillig und anonym ist. Die Ethikkommission der Medizinischen Fakultät Mannheim, Universität Heidelberg, stellte ein positives Ethikvotum für die Gesamtstudie aus (2015-542-N-MA).

Auswertung

Der Anteil nicht ernsthafter Testteilnehmender pro Fakultät wurde jeweils mit einem χ^2-Test auf Unabhängigkeit überprüft. Da die Stichprobe sehr groß ist und demzufolge auch kleine Unterschiede signifikant werden können, wurde zusätzlich die Effektstärke Cohens w für Kontingenztabellen und Cohens d für metrische Daten (s. u.) berechnet, um die inhaltliche Relevanz von Unterschieden zu beurteilen [25]. Die Effektstärke w ist kategorisiert als kein Effekt mit $w<0,1$, kleiner Effekt mit $w<0,3$, moderater Effekt mit $w<0,5$ sowie großer Effekt mit $w\geq0,5$ [25].
Für den Vergleich der TESS-Scores zwischen den beiden Fakultäten wurde ein t-Test für unabhängige Stichproben mit ungleichen Varianzen berechnet, sowie die Effektstärke d nach Cohen [25], wobei die gepoolte Standardabweichung nach Leonhart (2004) berechnet wurde [26]. Die Kategorisierung von d lautet wie folgt: $d<0,2$ kein Effekt, $d<0,5$ kleiner Effekt, $d<0,8$ moderater Effekt und $d\geq0,8$ großer Effekt.
Die Auswertung der beiden Likert-Items („Ich rede mit Kommilitonen...“ bzw. „...mit meinem Mentor über meine Ergebnisse beim PTM“) wurden umkodiert, sodass Angaben von 4 oder 5 als Zustimmung gewertet wurden.

Zur Auswertung der offenen Fragen zu den Kosten und Nutzen des PTM wurden qualitative und quantitative Methoden verwendet. Die Daten des Fragebogens wurden in drei Schritten analysiert: Zunächst untersuchten zwei Autorinnen (KG, MZ) alle Kommentare der offenen Fragen und codierten sie unabhängig voneinander anhand der thematischen Inhaltsanalyse [27]. In einem zweiten Schritt nach gemeinsamer Diskussion von Diskrepanzen und neuen Perspektiven wurden diese Codes wiederum unabhängig voneinander in Kategorien gruppiert und eine Kategorienliste erstellt. Im dritten Schritt wurde diese Kategorienliste hinsichtlich der Inter-Coder-Reliabilität überprüft mit perfekten Übereinstimmungen (je 100%) für die offenen Fragen 1 (8 Kategorien), 3 (7 Kategorien) und 4 (4 Kategorien). Sehr gute Übereinstimmungen ergaben sich für die offene Frage 2 (94%, 9 Kategorien) und Frage 5 (97%, 12 Kategorien), so dass diese Kategorienliste in den weiteren Analysen verwendet wurde. Im Ergebnisteil wird die Anzahl der Nennungen pro Kategorie angegeben. Die dazugehörigen Prozentzahlen beziehen sich jeweils auf die Gesamtanzahl der Nennungen für die angegebene Frage.

Ergebnisse

Deskriptive Statistik

415 Studierende an F1 sowie 453 Studierende an F2 nahmen an der Befragung teil. 234 Studierende der F1 beantworteten die in die Analyse einbezogenen Fragen (57% weiblich, Responder=56% der Stichprobe, 234/415). An F2 beantworteten 248 Studierende diese Fragen (71% weiblich; Responder=55% der Stichprobe, 248/453). Eine Übersicht findet sich in Tabelle 2.

Tabelle 2: Teilnehmende an der Befragung zum PTM

Merkmal	Abstufung	Fakultät 1	Fakultät 2
Anzahl der Teilnehmenden	Gesamt	415 (100%)	453 (100%)
	Davon Responder*	234 (56%)	248 (55%)
Geschlecht	Weiblich	133 (57%)	176 (71%)
	Männlich	101 (43%)	72 (29%)

Anmerkung: *bzgl. der Fragestellung

Die beiden Universitäten unterschieden sich statistisch signifikant hinsichtlich der Verteilung der Geschlechter ($\chi^2=10,52$, df=1, p<,001) mit einem höheren Frauenanteil bei F2, jedoch nicht in der Verteilung der Studierenden auf den vorklinischen und klinischen Abschnitt des Studiums (n. s.). In Voranalysen zeigten sich keine statistisch signifikanten Effekte, so dass die Variable Geschlecht in den Auswertungen nicht als Kovariate einbezogen wurde.

Tabelle 3: Kategorisierung der Nennungen auf die offene Frage „Fühlen Sie sich emotional durch den PTM eingeschränkt? Inwiefern?“

	F1	F2	Beispielkommentar
Anzahl der Kommentare[A]	55 (24%)	19 (8%)	
davon eher negative Kommentare bzgl. PTM[B]	26 (47%)	4 (21%)	
- Zeitverschwendung	13	0	*„Es stresst mich, kurz vor den Klausuren einen Nachmittag mit dem PTM zu verschwenden.“*
- PTM nervt	10	4	*„Man ist genervt/frustriert weil man erst so wenig weiß.“*
- Umstände	3	0	*„Ich könnte mit meinem einzigen freien Tag etwas sinnvolleres machen.“*
davon eher positive bzw. konstruktive Kommentare bzgl. PTM[B]	29 (53%)	15 (79%)	
- Ergebnis	11	7	*„Der PTM zeigt auf, wie viele Defizite man in seinem Wissen hat und je näher man dem Examen rückt, desto mehr belastet es einen, wenn man schlecht abschneidet.“*
- Überhaupt nicht	12	2	
- Akzeptanz	4	1	*„Der PTM ist gut um seinen Wissensstand nachzuverfolgen, extra Lerneinheiten für den PTM werden von mir nicht durchgeführt.“*
- Vergleich	2	5	*„Eine klare Vergleichbarkeit von verschiedenen Universitäten und Studiengängen (Regel- vs. Modellstudiengang).“*

Anmerkung: Angegeben ist die Anzahl der Nennungen für diese Frage. [A]Prozentangaben beziehen sich auf die Anzahl der Gesamtnennungen (Responder N_{F1}=234 bzw. N_{F2}=248), [B]Prozentangaben beziehen sich auf die Anzahl der Kommentare für diese eine Frage.

Anteile nicht ernsthafter Testteilnehmender beim PTM

Unabhängig von der Berechnung und der gewählten Operationalisierung zeigt sich, dass in Fakultät 1 der Anteil nicht ernsthafter Testteilnehmender deutlich höher ist als in Fakultät 2.

An F1 gibt es N_{F1}=173/1.410 (12%) Studierende, die beim PTM alle Fragen mit „weiß nicht“ bzw. gar nicht beantwortet haben, an F2 N_{F2}=5/1.191 (<1%). Dies ist ein signifikanter Unterschied mit $\chi^2(1)=142{,}20$; $p<0{,}001$ und einem kleinen Effekt von $w=0{,}23$.

Im Fragebogen wurden an F1 von 291/415 (70%) Studierenden folgende durchschnittliche TESS-Werte, die das selbsteingeschätzte Testbemühen widerspiegeln, angekreuzt: M_{F1}=2,51; SD_{F1}=1,08 und an F2 von 409/453 (90%) Studierenden M_{F2}=3,63, SD_{F2}=0,88. Dieser Unterschied ist ebenfalls signifikant mit $T(543{,}80)=14{,}68$; $p<0{,}001$ und einem großen Effekt von $d=-1{,}19$. Das Testbemühen der Studierenden an F2 war demnach deutlich größer als an F1. Wenn man die Testteilnehmenden in ernsthaft vs. nicht ernsthaft kategorisiert, gibt es N_{F1}=52/415 (13%) bzw. N_{F2}=3/453 (<1%) nicht ernsthafte Testteilnehmende. Dieser Unterschied ist ebenfalls signifikant mit $\chi^2(1)=68{,}96$; $p<0{,}001$ und einem moderaten Effekt ($w=0{,}31$).

Wahrgenommene Kosten in Fakultäten mit unterschiedlichen Prüfungs- und Feedbackbedingungen

Insgesamt berichteten die Studierenden aus F1 häufiger von wahrgenommenen Kosten des PTM. An F2 wurden insgesamt mehr positive, konstruktive Nennungen gegeben als in F1. Bei der Beantwortung der Frage 1 *„Fühlen Sie sich emotional durch den PTM eingeschränkt?“* waren Mehrfachnennungen möglich. In F1 gab es 55 Nennungen (24% der 234 Responder) zu dieser Frage, von denen 53% (29/55) konstruktiv waren. Von den 19 Nennungen (8% der 248 Responder) aus F2, welche diese Frage beantworteten, waren 15/19 (79%) Nennungen konstruktiv.

Die Zuordnung von Nennungen zu den einzelnen Kategorien pro Fakultät ist in Tabelle 3 aufgeführt.

Die Frage 2 *„Welche Nachteile sehen Sie beim PTM?“* ergab an F1 241 Nennungen (103% der 234 Responder; d. h. einige Mehrfachnennungen), davon 43% (104/241) konstruktive Nennungen. An F2 waren von 105 Nennungen (42% der 248 Responder) 65/105 (62%) konstruktiv. Die Zuordnung von Nennungen zu den einzelnen Kategorien pro Fakultät ist in Tabelle 4 aufgeführt.

Wahrgenommener Nutzen in Fakultäten mit unterschiedlichen Prüfungs- und Feedbackbedingungen

163 (39%) der Studierenden aus F1 sowie 309 (68%) der Studierenden aus F2 reden mit anderen Personen über ihre Ergebnisse beim PTM. Dabei stimmten 84 (20%) der Studierenden der F1 der Aussage zu, dass sie mit ihren Kommiliton/innen über ihre Ergebnisse beim PTM reden. An der F2 waren dies 147 (32%). Der Aussage, dass sie mit ihrer Mentorin / ihrem Mentor über die eigenen Ergebnisse beim PTM reden, stimmten 4 (1%) der Studierenden der F1 zu und 16 (4%) der F2. Mit anderen Personen reden insgesamt 75 (18%) der teilnehmenden Medizinstudierenden der F1 sowie 146 (32%) der F2 über ihre PTM Ergebnisse. Die Häufigkeit der Zustimmung zu den beiden geschlossenen Fragen sowie die Zuordnung von Nennungen zu den einzelnen Kategorien pro Fakultät für die anderen Personen (offene Frage 3) sind in Tabelle 5 aufgeführt.

Tabelle 4: Kategorisierung der Nennungen auf die offene Frage „Welche Nachteile sehen Sie beim PTM?“

	F1	F2	Beispielkommentar
Anzahl der Kommentare[A]	241 (103%)[C]	105 (42%)	
davon eher destruktive Kommentare bzgl. PTM[B]	137 (57%)	40 (38%)	
- Zeitverschwendung	50	13	*„überflüssig; nicht sinnvoll, bringt nichts; Zeitverschwendung; löst Verwirrung aus.“*
- Frustration	46	17	*„Frustrierend, wenn/da kaum Fragen beantwortet werden können; Überforderung; Frust für niedrige Semester oder Regelstudiengänge, da v.a. klinische Fragen; demotivierend; zu wenig kann beantwortet werden.“*
- Konsequenzlos	41	10	*„Pflichtveranstaltung; wird nicht ernst genommen, daher keine Aussage; keine Vergleichbarkeit.“*
davon eher konstruktive Kommentare bzgl. PTM[B]	104 (43%)	65 (62%)	
- Testformat	59	26	*„Testformat: zu lang, zu viele Fragen, zu wenig Zeit; Fragengebiete durcheinander; Fragen unterschiedlich schwer; Fragen wiederholen sich; Punktabzug bei falsch beantworteten Fragen (man traut sich nicht mehr, zu kreuzen); Auswertungsweise; keine Vorbereitung möglich; keine Angabe, wie viele Fragen man lösen können müsste.“*
- Organisation	19	10	*„Universitäre Planung: schlechte Platzierung (z. B. Wochenende, zu früh im Semester); nicht im Curriculum integriert; schlechte Durchführung (Lärm, Austausch untereinander).“*
- Übt Druck aus	9	12	*„Übt Druck aus, fördert Konkurrenzdenken und Frust.“*
- Keinen Nachteil	8	10	*„Von zu Hause; keine Nachbesprechung; Ergebnisse ohne Begründung (kein Lerneffekt).“*
- Nicht mit M2 vergleichbar	7	6	*„Nicht mit 2. StEx vergleichbar (keine "weiß-nicht-Option" im StEx, keine Fallfragen oder Bildfragen im PTM), führt zur Fehleinschätzung des eigenen Wissens.“*
- Keine Einsichtnahme	2	1	*„Keine gute Einsichtnahme möglich; Ergebnismitteilung zu spät, zu kompliziert, gar nicht; keine Einsichtnahme.“*

Anmerkung: Angegeben ist die Anzahl der Nennungen für diese Frage, [A]Prozentangaben beziehen sich auf die Anzahl der Gesamtnennungen (Responder N_{F1}=234 bzw. N_{F2}=248), [B]Prozentangaben beziehen sich auf die Anzahl der Kommentare für diese eine Frage, [C]Prozente > 100% ergeben sich durch Mehrfachnennungen pro Person.

Tabelle 5: Dialog mit anderen Personen über die Ergebnisse beim PTM

	F1	F2
Kommiliton/innen[A]	84 (20%)	147 (32%)
Mentor/in[B]	4 (1%)	16 (4%)
Anzahl der Kommentare[C]	75 (18%)	146 (32%)
- Freunde	18	39
- Familie	16	35
- Eltern	13	47
- Partner/-in	13	23
- niemandem	12	2
- sonstiges	3	0

Anmerkung: Angegeben ist die Anzahl der Zustimmungen auf die Aussagen
- [A] „Ich rede mit Kommilitonen über meine Ergebnisse beim PTM.“
- [B] „Ich rede mit meinem Mentor über meine Ergebnisse beim PTM.“
- [C] „Ich rede mit anderen Menschen über meine Ergebnisse beim PTM. Und zwar mit …“

Prozentangaben beziehen sich auf die Anzahl der Studierenden der Fakultät 1 und 2.

Bei der Frage 4 *„Nutzen Sie die Ergebnisse des PTM noch anderweitig? Wenn ja, wie?“* gab es 72/234 Nennungen an F1 (31% der Responder), und 33/248 Nennungen (13% der Responder) an F2. Obwohl es an der F1 insgesamt mehr Nennungen für eine anderweitige Nutzung gab als an der F2, waren die Nennungen von F1 zu einem hohen Prozentsatz eher Kategorien mit negativer Konnotation (70/72 Nennungen, 97%) zuzuordnen, im Vergleich dazu gab es an F2 nur 22/33 Nennungen (67%) mit eher negativer Konnotation, wie in Tabelle 6 dokumentiert.
Auf die Frage 5 *„Welche Vorteile sehen Sie beim PTM?“* gab es an beiden Fakultäten knapp über 200 Nennungen (F1 mit 207/234, 88% bzw. F2 mit 202/248, 81% der Responder). In F1 konnten 163/234 (79%) der Nennungen positiven Kategorien zugeordnet werden, in F2 198/248 (98%), wie in Tabelle 7 ersichtlich.

Diskussion

Formative Prüfungen sind wichtig als essentieller Teil des Prüfens für das Lernen. Wenn formative Prüfungen nicht bestehensrelevant sind, können sie im Empfinden der Studierenden hohe Kosten und/oder geringeren Nutzen im Vergleich zu summativen Prüfungen haben. In diesen Fällen kann der Anteil nicht ernsthafter Testteilnehmender hoch sein. In der vorliegenden Studie wurde untersucht, ob unterschiedliche Rahmenbedingungen an zwei Fakultäten einen Einfluss auf das Testbemühen sowie die wahrgenommenen Kosten und Nutzen einer formativen Prüfung – des Progress Tests Medizin (PTM) – haben. Die unterschiedlichen Rahmenbedingungen finden sich in der geforderten Anzahl der Teilnahmen, der Darbietung des PTM, der Rückmeldung der Ergebnisse sowie der universitären Einbindung. Obwohl an beiden medizinischen Fakultäten Maßnahmen zur Erhöhung der Akzeptanz des PTM durchgeführt werden, um das Testbemühen zu steigern, wird der PTM von den Studierenden an Fakultät 2 besser angenommen als an Fakultät 1. Dies zeigt sich in dem höheren Anteil ernsthafter Testteilnehmender, den niedrigeren wahrgenommenen Kosten und dem höheren berichteten Nutzen sowie dem größeren Anteil an konstruktiven Kommentaren.

Ernsthafte Testteilnahme

Die Hypothese „Der Anteil nicht ernsthafter Testteilnehmender an Fakultät 1 ist niedriger als an Fakultät 2.“ konnte nicht bestätigt werden. Entgegen dieser Hypothese ist der Anteil ernsthafter Testteilnehmender an Fakultät 1 kleiner als an Fakultät 2, trotz mehr Wahlmöglichkeiten und einem unmittelbaren Feedback. Obwohl an anderer Stelle gezeigt werden konnte, dass bei computer-basierter Administration der Anteil ernsthafter Testteilnehmender höher ist als bei papier-basierter Administration [28], haben bereits verschiedene Studien gezeigt, dass mehrere Faktoren Einfluss auf das Testbemühen haben. Unizentrische Studien können deshalb nur einen marginalen Erklärungsbeitrag liefern, um das multifaktorielle Bedin-

Tabelle 6: Kategorisierung der Nennungen auf die offene Frage „Nutzen Sie die Ergebnisse des PTM noch anderweitig? Wenn ja, wie?“

	F1	F2	Beispielkommentar
Anzahl der Kommentare[A]	72 (31%)	33 (13%)	
Davon eher negative Kommentare bzgl. des PTM[B]	70 (97%)	22 (67%)	
- nein	67	21	
- Abwehr	3	1	*„Ich finde es lächerlich sich mit anderen Komillitonen zu vergleichen. Was soll gerade in unserem Studium dieser Pseudo Druckaufbau!!! Das hat doch nichts mit vernünftiger Fehlerkultur zu tun!“*
Davon eher positive bzw. konstruktive Kommentare bzgl. des PTM[B]	2 (3%)	11 (33%)	
- Lernziele	2	8	*„Ich schaue mir an in welchen Fächern ich schlecht war + nehme mir vor daran etwas zu ändern“*
- Bewerbungen	0	3	*„Einreichung bei der Studienstiftung des dt. Volkes.“*

Anmerkung: Angegeben ist die Anzahl der Nennungen für diese Frage, [A]Prozentangaben beziehen sich auf die Anzahl der Gesamtnennungen (Responder N_{F1}=234 bzw. N_{F2}=248), [B]Prozentangaben beziehen sich auf die Anzahl der Kommentare für diese eine Frage.

Tabelle 7: Kategorisierung der Nennungen auf die offene Frage „Welche Vorteile sehen Sie beim PTM?“

	F1	F2	Beispielkommentar
Anzahl der Kommentare[A]	207 (88%)	202 (81%)	
Davon eher negative Kommentare bzgl. des PTM[B]	44 (21%)	4 (2%)	
- kein Vorteil	20	1	*„Kein Vorteil!“*
- kein Vorteil in Vorklinik	12	1	*„Kein Vorteil in der Vorklinik, eher frustrierend; nur für klinische Semester sinnvoll.“*
- Auswertung verfügbar	5	1	*„Kein Vorteil, da Fragen nicht (komplett), nicht einsehbar, nicht zu Hause nachzusehen/nachzuarbeiten; keine gute Auswertung verfügbar.“*
- kein Anreiz	4	0	*„Erfüllt lediglich universitäre Verpflichtung; kein Anreiz, sich ernsthaft mit dem Test zu beschäftigen.“*
- keine IMPP-Fragen	2	0	*„Keine IMPP-Fragen.“*
Davon eher positive bzw. konstruktive Kommentare bzgl. des PTM[B]	163 (79%)	198 (98%)	
- Feedback	109	128	*„(Regelmäßiges) Feedback/Überprüfung eigener Wissensstand bzw. -fortschritt; sehen, was bereits gelernt wurde, Rückmeldung zum Stand in einzelnen Fächern, zur Lernstrategie; sehen, was noch gelernt werden muss; Darstellung des Examenswissens; bereits fest vorhandenes Wissen wird abgefragt.“*
- Vergleich mit Kommilitonen	21	32	*„Vergleich mit anderen Studierenden (Kommilitonen, anderen Universitäten, anderen Ländern), Rückschluss auf die Qualität der Lehre an der eigenen Universität; Vergleich der Universitäten/Studiengangsmodelle.“*
- Prüfungsangst abbauen	14	23	*„Vorbereitung /Übung (zum 2. Staatsexamen); Klinische Fragen; Fragen und Prüfungsdauer kennenlernen; Prüfungsangst abbauen.“*
- Lernmotivation	14	7	*„Lernmotivation, Erfolgserlebnis; Lerneffekt.“*
- Prüfung ohne Druck	5	6	*„Prüfung ohne Prüfungsdruck/negative Folgen.“*
- andere Prüfung	0	2	*„Andere, nicht universitäre Prüfung; objektiver(er), anonymer Vergleich, andere Fragenautoren.“*

Anmerkung: Angegeben ist die Anzahl der Nennungen für diese Frage, [A]Prozentangaben beziehen sich auf die Anzahl der Gesamtnennungen (Responder N_{F1}=234 bzw. N_{F2}=248), [B]Prozentangaben beziehen sich auf die Anzahl der Kommentare für diese eine Frage.

gungsgefüge für das Testbemühen in formativen Prüfungen vollständig aufzuklären.

Kosten

Die vorliegende Studie konnte die Hypothese „Die wahrgenommenen Kosten des PTM sind an Fakultät 1 höher als an Fakultät 2." bestätigen. An Fakultät 1 werden von den Teilnehmenden mehr Kosten dargelegt als an Fakultät 2. Die Kommentare der Teilnehmenden spiegeln Befunde aus der Literatur wider, dass die Kosten des PTM als hoch wahrgenommen werden, wenn die Studierenden einschätzen, dass sie zur gleichen Zeit keine höher bewerteten Alternativen durchführen können, wie Lernen auf „richtige" Prüfungen oder wenn sie emotionalen Stress beim Ausfüllen des Tests empfinden [29].

Nutzen

Die Ergebnisse zur Überprüfung der Hypothese „Der wahrgenommene Nutzen des PTM ist an Fakultät 1 niedriger als an Fakultät 2." müssen differenzierter betrachtet werden. Obwohl mehr Studierende an Fakultät 2 über ihre PTM-Ergebnisse sprechen, handelt es sich hierbei zur Hälfte um fakultätsferne Personen. Dies ist verwunderlich, da der PTM ein Teil des Studiums sein sollte und daher zu erwarten wäre, dass die Studierenden hauptsächlich mit ihren Kommiliton/innen und Mentor/innen über die Ergebnisse reden. Eine Mentorin/ein Mentor wurde bei der Beantwortung dieser Frage jedoch nur selten genannt, obwohl es an F2 ein Mentorenprogramm gibt. Bei der Frage, ob die Studierenden die Ergebnisse des PTM noch anderweitig nutzen, war der Anteil der Kommentare der Studierenden an Fakultät 1 höher als bei den Studierenden der Fakultät 2. Allerdings handelt es sich hierbei um einen sehr hohen Anteil an Kommentaren mit negativer Konnotation bzw. Kommentare, die zeigen, dass die Ergebnisse nicht anderweitig genutzt werden. Studierende der Fakultät 2 haben das Prinzip des formativen Prüfens besser verstanden und nutzen die Ergebnisse des PTM als Feedback über den eigenen Wissensfortschritt, zur Lernmotivation und zur Reduktion von Prüfungsangst. Obwohl an Fakultät 2 die Einstellung gegenüber dem PTM positiver ist, haben an beiden Fakultäten Studierende nur in den seltensten Fällen erwähnt, dass sie den PTM als Lerninstrument verwenden (10 Nennungen von insgesamt 482 Studierenden, die den Fragebogen ausgefüllt haben). Daher ist die Wirkung auf das Lernen fraglich. Dies wäre jedoch ein Qualitätskriterium für die Nützlichkeit einer Prüfung [4], v. a. bei formativen Prüfungen, deren Funktion das Prüfen als eine Anregung und Rückmeldung zum Lernen ist. Die Lernwirkung muss genauer in weiteren Studien untersucht werden, zumal auch bei bestehensrelevanten Progress Tests die Wirkung auf das Lernen fraglich ist. So konnte nur eine moderate Rolle des Progress Tests bei der Identifikation von Stärken und Schwächen ausgemacht werden [30]. Aarts et al. zeigten, dass eine Mehrheit der Studierenden die Ergebnisse des bestehensrelevanten Progress Tests zum Monitoren ihres Wissens nutzten, jedoch war nicht klar, ob dies auch einen direkten Einfluss auf das Lernen hatte [31]. Dies zeigte sich auch bei Given et al. Sie fanden in semi-strukturierten Interviews heraus, dass sich die Studierenden zwar über ihre Stärken und Schwächen informiert fühlten, das Feedback jedoch keinen Einfluss auf das zukünftige Lernen hatte [32]. Auch Yielder et al. fanden in Fokusgruppen heraus, dass bei jüngeren Studierenden das zukünftige Lernen durch den Progress Test beeinflusst wird, jedoch nicht durch das Feedback, sondern durch den Inhalt des Tests [33]. Studierende in höheren Fachsemestern nutzen den Progress Test eher als Erinnerungen daran, dass sie überhaupt lernen müssen. Der Anteil der Kommentare zu den Vorteilen des PTM in der vorliegenden Studie ist in beiden Fakultäten ungefähr vergleichbar, jedoch zeigt sich auch hier, dass die Studierenden der Fakultät 2 den PTM positiver beurteilen. Daher kann gefolgert werden, dass die Hypothese zum wahrgenommenen Nutzen des PTM bestätigt werden kann, einschränkend kommt jedoch hinzu, dass die Wirkung als Lerninstrument auch an Fakultät 2 fraglich ist.

Stärken und Schwächen

In der vorliegenden Studie konnte gezeigt werden, dass verschiedene Prüfungs- und Feedbackbedingungen mit unterschiedlichen Anteilen ernsthafter Testteilnehmender und damit einer erhöhten Varianz des Testbemühens verbunden sein können. Zudem wurde ersichtlich, dass die Kosten und der Nutzen des Progress Tests an beiden Fakultäten unterschiedlich wahrgenommen werden. An Fakultät 2 waren nicht nur mehr ernsthafte Testteilnehmende vorhanden, sondern der PTM wurde auch bzgl. Kosten und Nutzen positiver wahrgenommen als an Fakultät 1.

Der Vorteil der vorliegenden Studie liegt im direkten Vergleich zweier Medizinischer Fakultäten, bei denen der PTM zum gleichen Zeitpunkt vor über 15 Jahren eingeführt wurde. In vielen Punkten sind die Bedingungen an beiden Fakultäten vergleichbar: beide haben einen Modellstudiengang und drei Staatsprüfungen, die Einfluss auf die PTM-Ergebnisse haben können [34]. Bei beiden Fakultäten gibt es vergleichbare Implementationsbedingungen des PTM, wie den gleichen Test, Teilnahmeverpflichtung sowie keine Zulassung zu weiteren Kursen, wenn am PTM nicht teilgenommen wird. Andererseits unterscheiden sich die beiden Bedingungen zur Implementierung des PTM durch die unterschiedliche Einbindung in das Qualitätsmanagementsystem bzw. in das Prüfungssystem sowie in der Rückmeldung (sofortige Rückmeldung der eigenen Ergebnisse bei computer-basierter Testung vs. auf Nachfrage Vergleich mit dem Lösungsheft). Neben den vergleichbaren Bedingungen an den beiden Fakultäten bietet die vorliegende Studie als weiteren methodischen Vorteil den Mixed-Methods-Ansatz, der sowohl quantitative als auch qualitative Auswertungen erlaubt. Somit wurde ein besserer Einblick in die Wahrnehmung des PTM an den beiden Fakultäten gewonnen und es konnte zudem quantitativ gezeigt werden,

dass sich der Anteil ernsthafter Testteilnahmen stark zwischen den beiden Fakultäten unterscheidet.
Die Methoden zur Bestimmung des Anteils ernsthafter bzw. nicht ernsthafter Testteilnahmen haben jede einzelne für sich Einschränkungen in ihrer Zuverlässigkeit, wie ein unbekanntes Ausmaß an Sensitivität/Spezifität („objektive Kriterien“) und die fragliche Aussagekraft der Selbstauskunft (TESS-Score). Um die Validität der Ergebnisse zu erhöhen, wurde mit Hilfe von Triangulation das Testbemühen mit verschiedenen Methoden gemessen. Da beide Methoden zu der gleichen Schlussfolgerung führen, kann davon ausgegangen werden, dass das Testbemühen an Fakultät 2 höher ist als an Fakultät 1. Außerdem lassen die Antworten aus den offenen Fragen ebenfalls diese Schlussfolgerung zu, da an Fakultät 2 mehr konstruktive Antworten gegeben wurden und auch ein höherer Nutzen und weniger Kosten berichtet wurden. Nach der Erwartungs-Wert-Theorie sollte daher die Motivation, diese Aufgabe zu erledigen, also beim PTM das Testbemühen, bei Fakultät 2 höher sein als bei Fakultät 1.

Fazit

Der formative PTM als Prüfung für das Lernen soll den Studierenden Rückmeldung über den Stand des eigenen medizinischen Wissens geben, im Vergleich zum Absolvierendenniveau und im Vergleich zu den Kommiliton/innen des gleichen Studienstandes, um im Rahmen des kontinuierlichen Lernens den Lernprozess zu begleiten und zu modulieren. Er soll dadurch ein Gegenpol zum sog. Bulimielernen sein, welches durch zu viele benotete Leistungsnachweise vermehrt auftreten kann [2]. Wie auch bei anderen nicht-bestehensrelevanten Tests gibt es beim PTM große Schwankungen im Testbemühen und damit eine fragliche Wirkung auf das Lernen. Es ist anzunehmen, dass Maßnahmen zur Senkung der wahrgenommenen Kosten und zur Erhöhung des wahrgenommenen Nutzens das Testbemühen und langfristig auch die Wirkung auf das Lernen positiv beeinflussen können. Auch wenn bei bestehensrelevanten Progress Tests mutmaßlich kein Problem mit dem Testbemühen auftritt, zeigt sich in Studien die eingeschränkte Lernwirkung. Daher sollten Rahmenbedingungen identifiziert werden, welche die wahrgenommenen Kosten und Nutzen formativer Prüfungen positiv beeinflussen und damit langfristig auch auf die Lernmodulation wirken. Da durch den PTM zwar Daten für eine Rückmeldung zum Leistungsstand bzw. zum Leistungsverlauf geliefert werden, jedoch die Nutzung des PTM als Lernsteuerungsinstrument bei den Studierenden liegt, sollte der PTM und die Verwendung der Ergebnisse für das eigene Lernen in das Curriculum eingebettet werden. Dies kann durch die Einbettung in das Prüfungssystem geschehen, sowohl als Teil der Prüfungsordnung als auch in der Darstellung der Informationen und der Ergebnisse, wie an Fakultät 2. Weitere Möglichkeiten, wahrgenommene Kosten und Nutzen als Fakultät zu beeinflussen, wären keine parallelen summativen Prüfungen zu den formativen Prüfungen [6], [8], eine Einbindung in das Mentorensystem für alle Studierende und nicht nur als Identifikation für notwendige Förderung leistungsschwacher Studierender [13], [30], [31], [35], [36], [37]. Denkbar wäre auch das Verwenden von formativen Prüfungen zur Erstellung und zum Nachhalten von Lernplänen, zusammen mit dem Mentor/der Mentorin [38]. Wenn formative Prüfungen verwendet werden, um kontinuierlich Feedback zum Wissensstand zu sammeln, mit dem Mentor/der Mentorin zu besprechen und das künftige Lernen darauf auszurichten, wie es im programmatischen Assessment vorgesehen ist [8], [39], dann werden sie ihrem Zweck gerecht. Und erst dann werden die Studierenden den Wert formativer Prüfungen erkennen.
Obwohl formative Prüfungen immer mehr an Bedeutung gewinnen, reicht es nicht, diese als Add-on in das Curriculum einzuführen. Neue Prüfungsformate benötigen vielmehr auch die entsprechenden Rahmenbedingungen, damit sie den gewünschten Effekt erzielen. Bei formativen Prüfungen müssen daher Bedingungen geschaffen werden, in denen die Ergebnisse einen Wert haben und zwar als Leitfaden durch das Studium und als Lenkung des Lernverhaltens. Nur wenn formativen und summativen Prüfungen gleich viel Bedeutung beigemessen wird, werden die wahrgenommenen Kosten und Nutzen vergleichbar sein und das Testbemühen ebenfalls. Somit kann der Fokus der Studierenden auf das kontinuierliche Lernen gelenkt werden, weg vom Bulimielernen, denn es ist anzunehmen, dass Studierende, welche ihr Handeln auf das bloße Bestehen von MC-Prüfungen ausrichten, den Wert von formativen Prüfungen überhaupt nicht erkennen können.
Nicht-bestehensrelevante Prüfungen sind eine gute Möglichkeit zu lernen, unter welchen Bedingungen Prüfungen für das Lernen funktionieren und wie diese effektiv in die bestehenden Curricula eingebettet werden können. Dazu sollte in weiteren Studien das Ausmaß der Einzelmaßnahmen und deren Wechselwirkung untersucht werden. Dies ist eine große Herausforderung, da die Untersuchung realer Bedingungen in der medizinischen Ausbildung erschwert wird durch die vielen, häufig unkontrollierbaren Bedingungen [40].

Interessenkonflikt

Die Autor*innen erklären, dass sie keinen Interessenkonflikt im Zusammenhang mit diesem Artikel haben.

Literatur

1. Schuwirth LW, van der Vleuten CP. Programmatic assessment: From assessment of learning to assessment for learning. Med Teach. 2011;33(6):478-485. DOI: 10.3109/0142159X.2011.565828
2. Schuwirth LW, van der Vleuten CP. The use of progress testing. Perspect Med Educ. 2012;1(1):24-30. DOI: 10.1007/s40037-012-0007-2

3. Berkhout JJ, Helmich E, Teunissen PW, van der Vleuten CP, Jaarsma AD. Context matters when striving to promote active and lifelong learning in medical education. Med Educ. 2018;52(1):34-44. DOI: 10.1111/medu.13463

4. van der Vleuten CP. The assessment of professional competence: Developments, research and practical implications. Adv Health Sci Educ Theory Pract. 1996;1(1):41-67. DOI: 10.1007/BF00596229

5. Wade L, Harrison C, Hollands J. Student perceptions of the progress test in two settings and the implications for test deployment. Adv Health Sci Educ Theory Pract. 2012;17(4):573-583. DOI: 10.1007/s10459-011-9334-z

6. Cobb KA, Brown G, Jaarsma DA, Hammond RA. The educational impact of assessment: a comparison of DOPS and MCQs. Med Teach. 2013;35(11):e1598-e1607. DOI: 10.3109/0142159X.2013.803061

7. van Berkel HJ, Nuy HJ, Geerlings T. The influence of progress tests and block tests on study behaviour. Instruct Sci. 1995;22(4):317-333. DOI: 10.1007/BF00891784

8. Heeneman S, Schut S, Donkers J, van der Vleuten CP, Muijtjens A. Embedding of the progress test in an assessment program designed according to the principles of programmatic assessment. Med Teach. 2017;39(1):44-52. DOI: 10.1080/0142159X.2016.1230183

9. Wigfield A, Eccles JS. Expectancy-value theory of achievement motivation. Contemp Educ Psychol. 2000;25(1):68-81. DOI: 10.1006/ceps.1999.1015

10. Gast L. "Kein Ort. Nirgends?" Das Subjekt der Erkenntnis und die Idee der Universität. Einige Gedanken aus psychoanalytischer Perspektive. Psychol Gesellschaftskritik. 2010;33/34(4/1):153-171.

11. Zeigarnik BV. Das Behalten erledigter und unerledigter Handlungen. Psychol Forsch. 1927;9:1-85.

12. Albano MG, Cavallo F, Hoogenboom R, Magni F, Majoor G, Manenti F, Schuwirth L, Stiegler I, van der Vleuten C. An international comparison of knowledge levels of medical students: the Maastricht Progress Test. Med Educ. 1996;30(4):239-45. DOI: 10.1111/j.1365-2923.1996.tb00824.x

13. van der Vleuten CP, Verwijnen GM, Wijnen WH. Fifteen years of experience with progress testing in a problem-based learning curriculum. Med Teach. 1996;18(2):103-109. DOI: 10.3109/01421599609034142

14. Nouns ZM, Georg W. Progress testing in German-speaking countries. Med Teach. 2010;32(6):467-470. DOI: 10.3109/0142159X.2010.485656

15. Osterberg K, Kölbel S, Brauns, K. Der Progress Test Medizin: Erfahrungen an der Charité Berlin. GMS Z Med Ausbild. 2006;23(3):Doc46. Zugänglich unter/available from: https://www.egms.de/static/de/journals/zma/2006-23/zma000265.shtml

16. Ryan RM, Deci EL. Intrinsic and extrinsic motivations: Classic definitions and new directions. Cont Educ Psychol. 2000;25:54-67. DOI: 10.1006/ceps.1999.1020

17. Hackman JR, Oldham GR. Motivation through the design of work: Test of a theory. Organ Behav Hum Perform. 1976;16(2):250-279. DOI: 10.1016/0030-5073(76)90016-7

18. Kulik JA, Kulik CLC. Timing of feedback and verbal learning. Rev Educ Res. 1988;58(1):79-97. DOI: 10.3102/00346543058001079

19. Tuten TL, Galesic M, Bosnjak M. Effects of immediate versus delayed notification of prize draw results and announced survey duration on response behavior in web surveys: An experiment. Soc Sci Comput Rev. 2004;22(3):377-384. DOI: 10.1177/0894439304265640

20. Irons A. Enhancing learning through formative assessment and feedback. London: Routledge Taylor & Francis Group; 2008. DOI: 10.4324/9780203934333

21. Nicol D, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006;31(2):199-218. DOI: 10.1080/03075070600572090

22. Smyth K. The benefits of students learning about critical evaluation rather than being summatively judged. Ass Eval High Educ. 2004;29(3):369-377. DOI: 10.1080/0260293042000197609

23. Schüttpelz-Brauns K, Kadmon M, Kiessling C, Karay Y, Gestmann M, Kämmer JE. Identifying low test-taking effort during low-stakes tests with the new Test-taking Effort Short Scale (TESS) - Development and Psychometrics. BMC Med Educ. 2018;18(1):101. DOI: 10.1186/s12909-018-1196-0

24. Brauns K. Identifikation von Musterkreuzern beim Progress Test Medizin. [Dissertation]. Berlin: Humboldt-Universität zu Berlin; 2007.

25. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale: Erlbaum; 1988.

26. Leonhart R. Effektgrößenberechnung bei Interventionsstudien. Reha. 2004;43:241-246. DOI: 10.1055/s-2004-828293

27. Mayring P. Qualitative Inhaltsanalyse - Grundlagen und Techniken. 11. Auflage. Weinheim, Basel: Beltz Verlag; 2008.

28. Karay Y, Schauber SK, Stosch C, Schuettpelz-Brauns K. Can computer-based assessment enhance the acceptance of formative multiple choice exams? A utility analysis. Med Teach. 2012;34:292-296. DOI: 10.3109/0142159X.2012.652707

29. Flake JK, Barron KE, Hulleman C, McCoach BD, Welsh ME. Measuring cost: The forgotten component of expectancy-value theory. Contemp Educ Psychol. 2015;41:232-244. DOI: 10.1016/j.cedpsych.2015.03.002

30. Blake J, Norman GR, Keane DR, Mueller B, Cunnington J, Didyk N. Introducing progress testing in McMaster University's problem-based medical curriculum: Psychometric properties and effect on learning. Acad Med. 1996;71(9):1002-1007. DOI: 10.1097/00001888-199609000-00016

31. Aarts R, Steidel K, Manuel BAF, Driessen EW. Progress testing resource-poor countries: A case from Mozambique. Med Teach. 2010;32(6):461-463. DOI: 10.3109/0142159X.2010.486059

32. Given K, Hannigan A, McGrath D. Red, yellow and green: What does it mean? How the progress test informs and supports student progress. Med Teach. 2016;38(10):1025-1032. DOI: 10.3109/0142159X.2016.1147533

33. Yielder J, Wearn A, Chen Y, Henning M, Weller J, Lillis S, Mogol V, Bagg W. A qualitative exploration of student perceptions of the impact of progress tests on learning and emotional wellbeing. BMC Med Educ. 2017;17(1):148. DOI: 10.1186/s12909-017-0984-2

34. Nouns ZM, Schauber S, Witt C, Kingreen H, Schüttpelz-Brauns K. Development of knowledge in basic medical sciences during undergraduate medical education-A comparison of a traditional and a problem-based curriculum. Med Educ. 2012;46(12):1206-1214. DOI: 10.1111/medu.12047

35. Coelho C, Zahra D, Ali K, Tredwin C. To accept or decline academic remediation: What difference does it make? Med Teach. 2019;41(7):824-829. DOI: 10.1080/0142159X.2019.1585789

36. Lillis S, Yielder J, Mogol V, O'Connor B, Bacal K, Booth R, Bagg W. Progress testing for medical students at the University of Auckland: Results from the first year of assessments. J Med Educ Curr Dev. 2014;1:41-45. DOI: 10.4137/JMECD.S20094

37. Norman G, Neville A, Blake J, Mueller B. Assessment steers learning down the right road: impact of progress testing on licensing examination performance. Med Teach. 2010;32(6):496-499. DOI: 10.3109/0142159X.2010.486063

38. Kastenmeier AS, Redlich PN, Fihn C, Treat R, Chou R, Homel A, Lewis BD. Individual learning plans foster self-directed learning skills and contribute to improved educational outcomes in the surgery clerkship. Am J Surg. 2018;216(1):160-166. DOI: 10.1016/j.amjsurg.2018.01.023

39. Schuwirth LW, van der Vleuten CP. Current assessment in medical education: Programmatic assessment. J Appl Test Technol. 2019;20(S2):2-10.

40. Ringsted C, Hodges B, Scherpbier A. 'The research compass': An introduction to research in medical education: AMEE Guide No. 56, Med Teach. 2011;33(9):695-709. DOI: 10.3109/0142159X.2011.595436

Korrespondenzadresse:
Dr. rer. nat. Katrin Schüttpelz-Brauns
Medizinische Fakultät Mannheim der Universität Heidelberg, GB für Studium und Lehrentwicklung, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim, Deutschland, Tel.: +49 (0)621/383-71270, Fax: +49 (0)621/383-71201
katrin.schuettpelz-brauns@medma.uni-heidelberg.de

Bitte zitieren als
Schüttpelz-Brauns K, Karay Y, Arias J, Gehlhar K, Zupanic M. Comparison of the evaluation of formative assessment at two medical faculties with different conditions of undergraduate training, assessment and feedback. GMS J Med Educ. 2020;37(4):Doc41. DOI: 10.3205/zma001334, URN: urn:nbn:de:0183-zma0013341

Artikel online frei zugänglich unter
https://www.egms.de/en/journals/zma/2020-37/zma001334.shtml

Eingereicht: *29.08.2019*
Überarbeitet: *10.03.2020*
Angenommen: *27.04.2020*
Veröffentlicht: *15.06.2020*

Copyright
©2020 Schüttpelz-Brauns et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe http://creativecommons.org/licenses/by/4.0/.