

Influence of a revision course and the gender of examiners on the grades of the final ENT exam – a retrospective review of 3961 exams

Abstract

Revision courses should repeat already acquired knowledge and skills and mostly provide a basis for passing the following exam.

Aim: The aim of the study is to investigate the influence of a previously attended revision course on the grades achieved in a final exam (Ear, Nose and Throat Diseases). Additionally we ask the question whether the gender of the examiners plays a role concerning the marks or not.

Methods: 3961 exams at the Department of Ear, Nose and Throat (ENT) Diseases in Vienna were investigated, 725 with revision course (experimental group) and 3236 without previous revision course (comparison group). The revision courses were performed in a standardized way concerning form and content, interactive and case based.

Both groups were examined uniform in regard to topics and time duration. 16 male and 6 female examiners were involved. The grading followed a five-level scale. The examination marks were calculated in the arithmetic mean and median value for the entire sample, gender dependence was calculated according to the Wilcoxon-Mann-Whitney-Test. The inferential statistics included single- and multiple factorial analyses of variance as well as uni- and multivariate regression models.

Results: The experimental group achieved a grade average of 2.54 compared with 2.46 for the comparison group. Splitting up into male and female examiners, an average of 2.54 and 2.58 resp. for the experimental group and 2.44 and 2.61 resp. for the comparison group resulted. Female examiner marked significantly lower grades in comparison to their male colleagues ($P=0.001926$).

Conclusions: The ENT revision course did not improve the grade averages of the final ENT exam. Female examiners grade stricter than male examiners. There was no difference concerning grades 4 (pass) and 5 (fail) but female examiners grade less with mark 1.

Keywords: Revision course, oral exam, grading, examiner, Medical School

Matthäus C. Grasl¹
Rudolf Seemann²
Michael Hanisch¹
Gregor Heiduschka¹
Karl Kremser³
Dietmar Thurnher¹

1 Medical University of Vienna, Department of Ear, Nose and Throat Diseases, Vienna, Austria

2 Medical University of Vienna, Department of Cranio-, Maxillofacial and Oral Surgery, Vienna, Austria

3 Medical University of Vienna, Department for Medical Education, Vienna, Austria

1. Introduction

1.1 Background

In revision courses, students and/or co-workers are being taught selected cognitive and procedural knowledge within a few days or weeks [1]. Focused selection and presentation of topics relevant for the exam are important. Knowingly omitting or emphasizing topics are challenges, which essentially influence the outcome of a revision course [2]. In the Anglo-American regions, revision courses are part of the regular curriculum [3], [4], [5], [6], [7], [8], [9]. In German speaking regions they are mostly optional courses [10], [11], [12], [1]. Most of the studies deal with less than 100 exams and survey periods of mostly one year. Survey periods of 5 years and student numbers of several thousand are missing completely. At

the Medical University of Vienna, the curriculum committee recommended optional revision courses to complete the syllabus to prevent delays of students to pass the exams [13]. Similar courses are being offered commercially by private organizations at high costs. The final ENT exam is an oral general exam including cognitive and procedural knowledge about the most frequent and most threatening diseases.

1.2 Aim of the study

The focused transfer of knowledge in presence based revision courses tries to compensate for knowledge gaps by intensive repetition of subjects. Several authors state improvement of exam results caused by the attendance of revision courses [14], [2], [6], [10], [11].

The relevance of our Study “Influence of a revision course and the gender of the examiners on the grades of the final ENT exam – a retrospective review of 3961 exams” is given by the fact that we want to test the generally valid opinion that revision courses improve the results of following exams.

A possible bias caused by examiners is a frequent risk at exams. Especially the gender bias of examiners at oral exams remains unclear. So far, the international literature shows different results. There is no agreement that examiners prefer candidates of their own gender [15]. Male examiners are not more stringent than female examiners [16]. Male examiners grade female students significantly better than male students [12].

This study especially addresses factors of program evaluation and thus is of general value. It includes understanding, confirmation and improvement of effects of intervention. The study tests whether we do what we think or not. A target-aimed support should result. Publication of our data should serve in comparison with other programs as basis for the decision to make changes or not.

The aim of the study is to test the influence on the results of the ENT final exam after participation in an interactive oral revision course. Additionally we ask the question, whether there are differences between male and female examiners concerning the grading or not.

2. Methods

2.1 Background

The curricular regulation for the doctoral viva („Rigorenordnung“) was issued for the medical schools in Austria at the beginning of the 20th century and remained with few changes in effect until the new curriculum (MCW, Medical Curriculum Vienna) was introduced in 2002. The students had to pass oral exams in all relevant disciplines [17]. The final exam for ENT was introduced in 1979 in Vienna Medical School.

2.2 Study design

The present study is a retrospective analysis of the grades of the final ENT exam in the 3rd section of the medical study at the Medical University of Vienna. A part of the student population has attended the revision course prior to the final ENT exam.

2.3 Description of sample

The administrator for student's affairs at the Department for ENT has digitally recorded grading results of 19260 exams between Oct. 1st, 1997 and Jan. 31st, 2013 without personal reference to the students examined. Additionally he coded and recorded data of the examiners, their gender, students and their gender. Thus, the basic set of data could be used in an anonymous way. Data of about

10% of the students who repeated the exam were recorded additionally, but not analyzed in this study.

For this study, between Oct. 1st 1997 and Jan. 31st, 2009, data from the revision courses and data of 3961 exams were included. 16 male and 6 female examiners were active. The gender of the students was included as a variable.

2.4 Framework conditions of the final exam in the discipline ENT and the parallel revision course

4 hours per week credit hours lecture and 1 credit hour practical course were designed for the discipline ENT. During the period of our investigation the revision course was offered additionally at no charge as an enhancement of the syllabus as a preparation for the final ENT exam. This elective course “System oriented revision course in ENT” was offered ongoing by one teacher and included 7 topics of complexes of symptoms. Based on clinical symptoms, interactively in a “question-answer-format” the revision of the topics was symptom and case oriented. At the beginning of each unit, the teacher informed the students about that the revision course would cover the syllabus of the system complexes (e.g. breathlessness, sore throat, etc.) but not the whole syllabus for the final ENT exam, which was listed in a catalogue for both students and teachers. The syllabus of the revision course covered 50-60% of the syllabus for the final ENT exam. Thus, the topics for the ENT exam were standardized. Also a mandatory minimum examination time of 15 minutes was specified. The students could normally choose an examiner.

The grading scale for the final exam was a 5-part scale: 1=excellent, 2=good, 3=satisfactory, 4=sufficient, 5=not sufficient. 1-4 meant passed, 5 meant failed and the exam needed to be repeated.

A second reviewer guaranteed the objectivity of evaluation and interpretation of the measuring instrument final ENT exam. A second examiner could have improved the objectivity of the performance, but sufficient resources were not available. The examiner takes part in the exam, which results in a low reliability [18] and is reason for low validity [19]. The revision course was not evaluated.

2.5 Statistical analysis

Data analysis was performed using the „Statistical Package for the Social Sciences – SPSS“ version 20.0 for Windows. The examination marks were calculated in the arithmetic mean and median value for the entire sample, for the experimental group with revision course, for the comparison group without revision course as well as in all groups according to the gender of the examiner. Moreover the percentage of the particular examination marks of oral examinations was quantified totally and in terms of gender of the examiners. The impact of the factors revision course (yes/no), gender of the examiner (female vs. male) and the gender of the students (female

vs. male) were determined by means of single factor analysis of variance, recoding the gender for the calculation. Taking into account the multiple testing all these three factors were calculated repeatedly in a multifactorial analysis of variance. Furthermore, the impact on the rating was described by linear regression analysis. The inferential statistics consisting of single- and multiple factorial analyses of variance as well as uni- and multivariate regression models were carried out with the open source statistical program „R version 3.1.2“ [https://www.r-project.org/].

3. Results

3.1 Participants

We analyzed 3961 exams from a single institution, the department of ENT at the Medical University of Vienna, wherefrom all examiners came.

3.2 Results of the exams

Table 1 shows the results for all examiners according to the gender of the examiners for the experimental group (with revision course) and comparison group (without previous revision course).

3.2.1 Average grade in the overall collective

In the overall collective of the examiners the experimental group with revision course ($n=725$; Grade: 2.54 ± 1.38) was graded in poor direction compared to the comparison group without revision course ($n=3236$; Grade 2.46 ± 1.34). When used in analysis of single factor variance statistically no significant results were certifiable (ANOVA: $Df=1$; square sum= 4.2 ; root mean square= 4.1559 , $F=2.2865$; $P=0.1306$). In the univariate linear regression resulted a statistically not significant effect of 0.08376 ($P=.131$ by Wald test) for lower grades of students which completed the revision course.

3.2.2 Dependency of the results of the exams on examiners gender

The examination grades of male examiners were on average 2.45 with a standard deviation of 1.36 and a sample size of 3376 . The examination marks of female examiners were on average 2.60 with a standard deviation of 1.28 and a sample size of 585 . Female examiner marked significantly poorer grades in comparison to their male colleagues (Wilcoxon rank sum test: $W=910610$, $P=.001926$).

3.2.3 Dependency of the results of the exams of student's gender

The arithmetic mean of the examination grades of female students were 2.45 ± 1.34 and 2.51 ± 1.36 for their male

colleagues, respectively. In the single factorial analysis of variance, students gender proved not to be a significant influencing factor (ANOVA: $Df=1$; square sum= 3.6 ; root mean square= 3.64 ; $F=2.0028$; $P=0.1571$). In the analogous single variant regression model the effect of female was 0.06233 on the average grade with no significance ($P=.157$ by Wald test).

3.3.4 Multifactorial analysis of variance

Taking into account the multiple testing, the three factors revision course, gender of the examiner and the gender of the students were tested in a multivariate regression analysis (see Table 2), resp. a multifactorial analysis of variance (see Table 3). Only the gender of the examiners proved to be furthermore a significant factor of influence. On the average, female examiners assigned 0.146 points lower grades.

3.2.5 Average grades, number and percent of fails (“not sufficient”) compared with the gender of students

There is a special emphasis on grade 5 (not sufficient). The percentage varied between 12.23% for the combination female candidate/female examiner and 8.13% percent for the combination male candidate/female examiner (see Table 4).

The average over all combinations is 10.2% and thus showed a very homogenous grading for the decision pass or fail.

4. Discussion

4.1. Value of the results and comparison with the existing literature

In the overall collective of the examiners the experimental group with revision course was graded in poor direction compared to the comparison group without revision course. When used in analysis of single factor variance, statistically no significant results were found, and in the univariate linear regression there resulted a statistically not significant effect for lower grades of students who completed the revision course.

Literature shows high efficiency and acceptance by the students for revision courses. On the other hand, the effort, which is balanced by a good preparation for exams or for the job, should not be underestimated [2].

Throughout, revision courses can be seen as useful courses in different forms. Students state that their knowledge is better [6] or even significantly better [14] than before and a revision course is effective as a preparation for state exams [10]. Students are no longer anxious but motivated for further independent learning. Grades improved and significantly less students fail the following exam. The limited period guarantees continuous learning. Weak and good students benefit likewise [1]. Revision courses are efficient, cost-effective and adaptable for

Table 1: Arithmetic mean and median value of all examination results, from the experimental group with revision course and the comparison group without revision course, separated by the gender of the examiner.

Examinations	all			with revision course			without revision course		
	No.	Grade Ø	Mean	No.	Grade Ø	Mean	No.	Grade Ø	Mean
16 Males	3376	2.45	2	622	2.54	2	2754	2.44	2
6 Females	585	2.60	2	103	2.58	2	482	2.61	2
22 Total sum	3961	2.48	2	725	2.55	2	3236	2.46	2

Table 2: Multivariate regression model of the grading estimated by genders of examiners and students as well as the students participation in the revision course. The average grade corresponds to the reference group of male students examined by male examiners.

	Effect	Standard error	t-value	P(> t)
Average grade	2.47462	.03743	66.122	<2e-16***
Female examiner	.14589	.06041	2.415	.0158*
Female students	-.05736	.04406	-1.302	.1931
Revision course	.08500	.05536	1.535	.1247

Significance codes: *** .001 and * .05

Table 3: Multifactorial analysis of variance of grading.

Factor	Degree of freedom	Square sum	Mean square sum	F-value	P(>F)
Female Examiner	1	11.1	11.0746	6.1020	.01354 *
Female Students	1	3.1	3.0749	1.6943	.19312
Revision course	1	4.3	4.2585	2.3464	.12565
Residual value	3957	7181.6	1.8149		

Significance code: * .05

Table 4: Average grades, number and percentage of failed students compared to gender of students and examiners.

	No.	Grades Ø	No. of failed students	Percent
Female student /male examiner	2113	2.43	203	9.61
Female student /female examiner	327	2.61	40	12.23
Male student /male examiner	1263	2.5	140	11.08
Male student /female examiner	258	2.59	21	8.13
Total sum	3961	2.48	404	10.2

high numbers of students [3]. For tests of practical skills performance and success improve [5].

Our results are contrary to relevant literature.

Female students showed an average grade of 2.45 ± 1.34 , their male colleagues 2.51 ± 1.36 . In the single factor analysis of variance student's gender proved not to be a significant influential factor. In the univariate linear regression model the effect of female gender with -0.06233 at the average of marks was classified as not significant. These results comply with literature [12], [15], [16], [17], [18].

4.2 The ENT revision course did not improve the average grades for the following final oral ENT exam covering the whole subject.

The teacher of the revision course was examiner as well. There was no bias concerning participants of the revision course. The average grades of students having attended the revision course were only marginally better. Differentiation between well and poor performing students was not carried out.

Why, contrary to our assumption, students having attended a revision course compared with students not having attended a revision course did not perform better but even slightly poorer, can be explained as follows: students think that the revision course is a substitute for self-studies and not a supplement and revision and a time saving way. The final ENT exam is one of the last exams

of their medical studies. Additionally there could have been a longer time span between the revision course and the exam and the revision course only covered 50-60% of the topics examined in the final exam. It cannot be totally dismissed, that the revision course possibly promoted the effect of "bulimic learning", which means learning for a single test, keeping knowledge in the short time memory, spitting it out at the exam without digestion. We did not ask the students for their motivation to attend the revision course. Possible reasons are:

- Offer to attend a tight teaching/learning program
- Interest in ENT topics
- Structured interactive and case oriented tuition by a teacher with high subject-specific and didactic competence
- Compensation of the missing involvement with the subject
- "Distance learning" without attending lectures
- Promotion of further self-directed learning
- Reduction of fear and uncertainty before the exam

4.3 The average examination grades of female examiners are significantly poorer than those of male examiners

Bias of exam results caused by examiners are a latent risk at exams. Humans behave differently and it is not surprising that examiners show tendencies of preference and discrimination of candidates. Knowing this, measures

should be taken as early as possible. These vary from randomized candidate distribution or a board of examiners or even changing from an oral to a written exam [20]. Different features of the candidates, not yet examined, could cause a possible reason for gender dependent grading. These variables could influence, to a different extent, male and female examiners, which makes it necessary to explain different connections [15]. Free choice of examiners, age of candidates, experience of examiners and sympathy or reservation regarding the candidate. Wiskin et al. [21] describe in their paper about gender as a variable at the OSCE about communication at the level of the last year of studies, that female examiners give better grades. They state that this could be prevented by a strategic gender equality for men and women. Also Boehm et al. [12] report a better grading by female examiners in the final exam Social Medicine. McManus et al. [22] propose to solve the problem of a bias caused by gender or ethnic origin at practical OSCE stations, by deploying 2 examiners per station. Using multifactorial Rasch modelling they estimate the effect that examiners prefer or aggravate, and compensate the effect by pairing toughest and least tough examiners. Despite the possible bias, all our examiners valued this kind of exam. It offered the possibility to examine declarative and procedural knowledge case and symptom oriented together with basic knowledge. Still, there was no significant difference for the fail ratio between male and female examiners.

4.4 Strengths and weaknesses of the study

4.4.1 Strengths

The authors consider the clear concept of this retrospective study as strength. The high number of cases allow clear statistical statements. During the long data collection period, 22 examiners were included.

4.4.2 Weaknesses

The design is simple and not very robust, but easily done and cost saving. There was no structural feedback from students and examiners, which could have given insight into individual-related bias caused by the interaction between examiners and candidates. We have no data about how often Students attended the revision course and how close the attendance was prior to the exam. We did not consider especially the repeaters (10%) in our calculations. Having performed an evaluation of the revision course, we could have distinguished between "reaction level" and "learning level" [23], [24].

5. Conclusions

We conducted this study to find out the advantage and the influence of the gender of examiners of the final ENT exam at the Medical University of Vienna. The ENT revision

course did not improve the average grades at the final ENT exam. The average grades of female examiners were significantly worse than the grades of male examiners. The main difference between the grading of female and male examiners was for grades 1 and 2 and not for grade 5 (fail). All examiners were very close for the decision between 4 (pass) and 5 (fail). As a relative limit of this study, a low reliability can be seen which also causes a low validity. After the introduction of a new medical curriculum in 2002, which does not include final ENT exams any more, also the revision courses for the final ENT exam were discontinued. However, in the first year of the new curriculum several teachers offer optional revision courses close to the exam at the end of the school year, without registration.

Acknowledgements

The authors give thanks to all teachers of the ENT department of the Medical University of Vienna, who participated at this study as examiners.

Competing interests

The authors declare that they have no competing interests.

References

1. Störmann S, Chiapponi C, op den Winkel M, Wöck M, Bender J, Kern AB, Gebhardt C, Angstwurm M. Mit dem Internet zum Examen - Prüfungsvorbereitung mit dem virtuellen Staatsexamen Repetitorium der LMU München. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bochum, 23.-25.09.2010. Düsseldorf: German Medical Science GMS Publishing House; 2010. Doc10gma143. DOI: 10.3205/10gma143
2. Kühn J, Jabs WJ. Das Lübecker Repetitorium "Innere Kompakt": ein Pilotprojekt zur Vorbereitung auf das neue zweite Staatsexamen. *GMS Z Med Ausbild.* 2007;24(3):Doc148. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2007-24/zma000442.shtml>
3. Hibbert EJ, Lambert T, Carter JN, Learoyd DL, Twigg S, Clarke S. A randomized controlled pilot trial comparing the impact of access to clinical endocrinology video demonstrations with access to usual revision resources on medical student performance of clinical endocrinology skills. *BMC Med Educ.* 2013;13:135. DOI: 10.1186/1472-6920-13-135
4. Lymn JS, Mostyn A. Audience response technology: Engaging and empowering non-medical prescribing students in pharmacology learning. *BMC Med Educ.* 2010;10:73. DOI: 10.1186/1472-6920-10-73
5. Meade O, Bowskill D, Lymn JS. Pharmacology as a foreign language: A preliminary evaluation of podcasting as a supplementary learning tool for non-medical prescribing students. *BMC Med Educ.* 2009;9:74. DOI: 10.1186/1472-6920-9-74
6. Mole G, Gillespie L. Those who can, teach. Accessing medical students' perception of a finals revision programme delivered by foundation and core trainees: a cross-sectional study. *BMJ Open.* 2012;2:e001544. DOI: 10.1136/bmjopen-2012-001544

7. Freimanis AK. Successful teaching of radiology of medical students and interns. With emphasis on third and fourth-year programs. *Radiology*. 1970; 97(3):669-680. DOI: 10.1148/97-3-669
8. Swartz TH, Lin JJ. A clinical refresher course for medical scientist trainees. *Med Teach*. 2014;36(6):475-479. DOI: 10.3109/0142159x.2014.886767
9. Brown G, Manogue M. AMEE Medical Education Guide No. 22: Refreshing lecturing: a guide for lecturers. *Med Teach*. 2001;23(3):231-244. DOI: 10.1080/01421590120043000
10. Rengier F, Rauch PJ, Partovi S, Kirsch J, Nawrotzki R. A three-day anatomy revision course taught by senior peers effectively prepares junior students for their national anatomy exam. *Ann Anat*. 2010;192(6):396-399. DOI: 10.1016/j.aanat.2010.02.008
11. Bredemeier S, Pabst R, Nave H. Der "Makro-Marathon": Ergebnisse der Evaluation eines Repetitoriums der makroskopischen Anatomie zur Vorbereitung auf das Physikum. *GMS Z Med Ausbild*. 2005;22(4):Doc197. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000197.shtml>
12. Boehm G, Bernhard G, Kwizda-Gredler B, Kunze U, Rathmanner T, Rieder A, Schoberberger R, Schwarz B, Vutuc C, Kunze M. Einfluss von Geschlecht und Studiengebühren auf die Noten bei Rigorosum-Prüfungen im Prüfungsfach Sozialmedizin. Wien: Didaktik; 2001. Zugänglich unter/available from: <http://didaktik-on.net/cgi-bin/didaktik.cgi?id=0000053>
13. Bundesministerium für Wissenschaft und Forschung. Studienplan für die Studieneinrichtung Medizin an der Medizinischen Fakultät der Universität Wien: § 11 Repetitorien: Auf Empfehlung der Studienkommission sind Repetitorien einzurichten, wenn dies aus Gründen der Vollständigkeit des Lehrangebotes (§ 58 lit.b UOG) oder zur Vermeidung von Studienverzögerungen (§ 58 lit.j UOG) erforderlich ist. Wien: Bundesministerium für Wissenschaft und Forschung; 1992.
14. Webb AL, Choi S. Interactive radiological anatomy eLearning solution for first year medical students: Development, integration, and impact on learning. *Anat Sci Edu*. 2014;7(5):350-360. DOI: 10.1002/ase.1428
15. Denney ML, Freeman A, Wakeford R. MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source? *Br J Gen Pract*. 2013;63(616):e718-725. DOI: 10.3399/bjgp13X674396
16. McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. *BMC Med Educ*. 2013;13:103. DOI: 10.1186/1472-6920-13-103
17. Ministerium für Kultus und Unterricht. Das Studium an der medizinischen Fakultät. Die medizinische Rigorosenordnung. Verordnung des Ministers für Kultus und Unterricht vom 21. Dezember 1899, RGBl. Nr. 271. Wien: Ministerium für Kultus und Unterricht; 1899.
18. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach*. 2005;27(4):294-297. DOI: 10.1080/01421590500126437
19. Schuhwirth LW, van der Leuten CP. Quality control: assessment and examinations. *Z Hochschuldidaktik*. 1996;20(1):66-76.
20. Niehaus DJ, Jordaan E, Koen L, Mashile M, Mall S. Applicability and fairness of the oral examination in undergraduate psychiatry training in South Africa. *Afr J Psychiatry (Johannesbg)*. 2012;15(2):119-123. DOI: 10.4314/ajpsy.v15i2.16
21. Wiskin CM, Allan TF, Skelton JR. Gender as a variable in the assessment of final year degree-level communication skills. *Med Educ*. 2004;38(2):129-137. DOI: /10.1111/j.1365-2923.2004.01746.x
22. McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Med Educ*. 2006;6:42. DOI: 1186/1472-6920-6-42
23. Jünger J, Just I. Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin. *GMS Z Med Ausbild*. 2014;31(3):Doc34. DOI: 10.3205/zma000926
24. Kirkpatrick D. *Evaluating Training Programs: The Four Levels*. 2nd ed. San Francisco/CA: Berrett-Koehler Publishers Inc.; 1998.

Corresponding author:

Ao. Univ.-Prof. Dr. Matthäus C. Grasl, MME
Medical University of Vienna, Department of Ear, Nose
and Throat Diseases, Währinger Gürtel 18-20, A-1090
Vienna, Austria
matthaeus.grasl@meduniwien.ac.at

Please cite as

Grasl MC, Seemann R, Hanisch M, Heiduschka G, Kremser K, Thurnher D. Influence of a revision course and the gender of examiners on the grades of the final ENT exam – a retrospective review of 3961 exams. *GMS Z Med Ausbild*. 2015;32(4):Doc38. DOI: 10.3205/zma000980, URN: urn:nbn:de:0183-zma0009802

This article is freely available from

<http://www.egms.de/en/journals/zma/2015-32/zma000980.shtml>

Received: 2014-09-18

Revised: 2015-09-14

Accepted: 2015-09-18

Published: 2015-10-15

Copyright

©2015 Grasl 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/>.

Beeinflussung der Prüfungsergebnisse durch ein Repetitorium als Vorbereitung auf das mündliche HNO-Rigorosum und Einfluss des Geschlechts der PrüferInnen – eine retrospektive Aufarbeitung von 3961 Prüfungen

Zusammenfassung

Repetitorien sind Lehrveranstaltungen, die bereits erworbenes Wissen und Kenntnisse wiederholen und meist als Vorbereitung zum Bestehen eines Examens dienen.

Zielsetzung: Zweck der Studie ist es, die Beeinflussung des Ergebnisses bei der mündlichen Gesamtprüfung (Rigorosum) im Fach Hals-Nasen-Ohrenheilkunde zu untersuchen, wenn zuvor ein interaktives mündliches Repetitorium besucht wurde. Zudem wird der Frage nachgegangen, ob es Unterschiede zwischen weiblichen und männlichen Prüfern bei der Notenvergabe gibt.

Methode: 3961 Einzelprüfungen an der Universitätsklinik für HNO-Krankheiten in Wien wurden untersucht, 725 mit Repetitorium (Versuchsgruppe) und 3236 ohne Repetitorium (Vergleichsgruppe). Das Repetitorium wurde als inhaltlich standardisiertes Wahlfach interaktiv und fallorientiert gestaltet.

Beide Gruppen wurden nach einheitlichen Kriterien (Inhalte, Dauer) geprüft. 16 männliche Prüfer und 6 weibliche Prüferinnen waren tätig. Die Benotung erfolgte in einer 5-teiligen Skala und wurde als arithmetischer Durchschnittswert mit Median, die Geschlechtsabhängigkeit der Notenvergabe wurde mit dem Wilcoxon-Rangsummentest berechnet. Die schließende Statistik bestand aus ein- und multifaktoriellen Varianzanalysen sowie uni- und multivarianten Regressionsmodellen.

Ergebnisse: Die Versuchsgruppe erzielte einen Notendurchschnitt von 2,54 gegenüber jener der Vergleichsgruppe mit 2,46. Nach Aufteilung in männliche und weibliche Prüfer ergibt sich ein Notendurchschnitt für die Versuchsgruppe von 2,54 bzw. 2,58 und für die Vergleichsgruppe von 2,44 bzw. 2,61. Der Unterschied ist statistisch signifikant ($p=0,001926$).

Schlussfolgerung: Das HNO-Repetitorium hat nicht zu einer Verbesserung des Notendurchschnittes bei der nachfolgenden mündlichen Prüfung über Inhalte des gesamten Faches geführt. Weibliche Prüfer benoten strenger als männliche. Der wesentliche Unterschied zwischen männlichen und weiblichen PrüferInnen in der Notengebung besteht bei der Vergabe der Noten 1 und 2 und nicht bei der Note 5 (nicht bestanden).

Schlüsselwörter: Repetitorium, mündliche Prüfung, Benotung, Prüfer, Medizinstudium

Matthäus C. Grasl¹
Rudolf Seemann²
Michael Hanisch¹
Gregor Heiduschka¹
Karl Kremser³
Dietmar Thurnher¹

1 Medizinische Universität
Wien, Universitätsklinik für
Hals-, Nasen- und
Ohrenkrankheiten, Wien,
Österreich

2 Medizinische Universität
Wien, Universitätsklinik für
Mund-, Kiefer- und
Gesichtschirurgie, Wien,
Österreich

3 Medizinische Universität
Wien, Department für
Medizinische Aus- und
Weiterbildung, Wien,
Österreich

1. Einleitung

1.1 Hintergrund

Repetitorien sind Lehrveranstaltungen, bei denen Studierenden und/oder graduierten Mitarbeitern aus allen Fächern in wenigen Tagen bis über mehrere Wochen hinweg kognitives oder prozedurales Wissen in unterschiedlicher Gewichtung vermittelt wird [1]. Die knappe, aber prägnante Darstellung der prüfungsrelevanten Aspekte steht im Vordergrund. Bewusstes Weglassen einiger Themen und das Hervorheben anderer sind die Herausforderungen, die das Gelingen eines Repetitoriums maßgeblich beeinflussen [2]. Als Prüfungsvorbereitung sind Repetitorien besonders im angloamerikanischen Raum im Curriculum fest verankert [3], [4], [5], [6], [7], [8], [9]. Im deutschsprachigen Raum sind sie allerdings meist freiwillige Lehrveranstaltungen [10], [11], [12], [1].

Die Mehrzahl der Studien umfasst Studierendenzahlen von bis zu einigen Hundert und Erhebungszeiträume von meist nur einem Studienjahr. Erhebungszeiträume von 5 Jahren und Studierendenzahlen von mehreren Tausend fehlen gänzlich.

An der Medizinischen Universität Wien wurden auf Empfehlung der Studienkommission kostenlose Repetitorien für Studierende aus Gründen der Vollständigkeit des Lehrangebotes und zur Vermeidung von Studienverzögerungen eingerichtet [13]. Ähnliche Repetitorien werden von kommerziellen Anbietern gegen nicht unerhebliche Honorare angeboten. Das Fach Hals-Nasen-Ohrenheilkunde wird als mündliche Gesamtprüfung (Rigorosum) durchgeführt und prüft kognitives und prozedurales Wissen zu den häufigsten und gefährlichsten Erkrankungen.

1.2 Ziel der Studie

Die komprimierte Wissensvermittlung in Repetitorien als Präsenzunterricht versucht Wissenslücken, die im Studium entstanden sind, durch intensive Wiederholung des Stoffes auszugleichen. In der Literatur finden sich vielfach Hinweise auf bessere Prüfungsergebnisse, wenn Repetitorien besucht werden [14], [2], [6], [10], [11].

Wir sehen als Begründung für die Relevanz unserer Studie „Beeinflussung der Prüfungsergebnisse durch ein Repetitorium als Vorbereitung auf das mündliche HNO-Rigorosum“ die allgemein gültige Meinung zu prüfen, dass Repetitorien den nachfolgenden Prüfungserfolg steigern.

Möglicher Prüfer-Bias stellt ein häufiges Risiko bei jeder Prüfung dar. Insbesondere das Geschlecht der Prüfer bei mündlichen Prüfungen führt zu Verzerrungen deren Klärung noch nicht restlos erfolgt ist. Die bisherige internationale Literatur dazu zeigt durchaus unterschiedliche Ergebnisse. Prüfer zeigen keine generelle Tendenz Prüflinge ihres Geschlechts zu bevorzugen [15]. Männliche Prüfer sind nicht strenger als weibliche Prüfer [16]. Männliche Prüfer benoten weibliche Studentinnen signifikant besser als deren männliche Kollegen [12].

Die Studie beschäftigt sich vor allem mit Faktoren der Programmevaluation und ist über das lokale Setting hin-

aus interessant. Sie beinhaltet das Verstehen, das Bestätigen und Steigern von Effekten der Intervention. Es wird überprüft, ob man das macht, was man zu machen glaubt. Eine zielführende Unterstützung wird angestrebt. Die Veröffentlichung der gewonnenen Daten soll im Vergleich zu anderen Programmen als Grundlage zur Beibehaltung und Vervielfältigung andernorts dienen.

Zweck der Studie ist es, die Beeinflussung des Ergebnisses bei der mündlichen Gesamtprüfung (Rigorosum) im Fach Hals-Nasen-Ohrenheilkunde zu untersuchen, wenn zuvor ein interaktives mündliches Repetitorium zur Prüfungsvorbereitung besucht wurde. Außerdem wird der Frage nachgegangen, ob es Unterschiede zwischen weiblichen und männlichen Prüfern bei der Notenvergabe gibt.

2. Methoden

2.1 Hintergrund

Die Rigorosenordnung als Form der mündlichen Prüfung zur Erlangung eines akademischen Grades an medizinischen Fakultäten in Österreich wurde zu Beginn des 20. Jahrhunderts erlassen. Sie regelte den bis zur Etablierung des Neuen Wiener Curriculums 2002 im Wesentlichen unveränderten Fächerkanon mit mündlichen Prüfungen - Rigorosen - in allen dafür vorgesehenen Fächern [17]. Diese Prüfungsform für das Fach Hals-Nasen-Ohrenheilkunde wurde in Wien 1979 eingeführt und wird auslaufend bis dato durchgeführt.

2.2 Studiendesign

Die hier dargestellte Studie ist eine retrospektive Analyse der Benotung des Rigorosums im Fach Hals-Nasen-Ohrenheilkunde von Studierenden der Humanmedizin im 3. Studienabschnitt an der Medizinischen Universität Wien. Ein Teil der Studierenden hat vor Ablegung des Rigorosums an einem prüfungsvorbereitenden Repetitorium teilgenommen.

2.3 Stichprobenbeschreibung

Der Administrator für Studienangelegenheiten an der Universitätsklinik für Hals-Nasen-Ohrenheilkunde hat vom 1.10.1997 bis 31.1.2013 ohne Personenbezug fortlaufend die Benotung von 19260 Prüfungsereignissen digital aufgezeichnet. Gleichzeitig zeichnete er verschlüsselt die einzelnen Prüfer und deren Geschlecht, sowie die Prüflinge und deren Geschlecht auf, sodass der Basisdatensatz für die weitere Bearbeitung und Auswertung vollkommen anonym geführt werden konnte. Ein Prozentsatz von ca. 10% Prüfungswiederholern wurde in die Studie mit eingeschlossen aber in der Analyse nicht extra berücksichtigt. Für diese Studie wurde der Zeitraum vom 1.3.2004 bis 31.7.2009, in dem gleichzeitig Repetitorien stattgefunden haben, ausgewählt und die Daten von 3961 Prüfungen eingeschlossen. Die Rigorosumsprüfung im Beobachtungs-

zeitraum wurde von 16 Männern und 6 Frauen abgenommen. Das Geschlecht der Prüflinge als Variable wurde in die Studie einbezogen.

2.4 Rahmenbedingungen des mündlichen Rigorosums aus Hals-Nasen-Ohren-Heilkunde und des dazugehörigen Repetitoriums

Als Lehrveranstaltungen für das Fach Hals-, Nasen-, und Ohrenheilkunde waren vier Semesterwochenstunden Vorlesungen und eine Semesterwochenstunde Pflichtpraktikum vorgesehen. Im Untersuchungszeitraum wurde gleichzeitig zur Erweiterung dieses Lehrangebotes ein kostenloses Repetitorium zur Vorbereitung auf das mündliche HNO-Rigorosum eingerichtet. Dieses Wahlfach „Symptomorientiertes Repetitorium aus HNO-Heilkunde“ wurde laufend angeboten und immer vom selben Vortragenden - in sich wiederholenden Einheiten von 7 verschiedenen „Symptomkomplexen“ - abgehalten. Anhand von klinischen Krankheitsbildern erfolgte interaktiv im „Frage-Antwort-Format“ eine symptom- und fallorientierte Stoffwiederholung. Die Studenten wurden am Beginn jeder Einheit darüber informiert, dass die Unterrichtseinheiten zwar den Lehrstoff der Symptomenkomplexe (z.B. Atemnot, Halsschmerzen etc.) abdecken, nicht aber den gesamten Lehrstoff der „HNO-Prüfung“, der in Form eines Katalogs mit ca. 200 Fragenbereichen sowohl allen Studierenden als auch allen Prüfern zur Verfügung stand. Die Lerninhalte des Repetitoriums überdeckten sich in ca. ca. 50-60% mit jenen des Rigorosum. Insgesamt ergab sich eine Standardisierung der Prüfungsinhalte, die zusätzlich noch durch die in der Rigorosenordnung festgelegte Mindestprüfzeit von 15 Minuten unterstützt wurde [18]. Studierende konnten einen Wunschprüfer angeben, im Allgemeinen wurde dies auch respektiert.

Die Benotung der studentischen Leistung beim Rigorosum erfolgte nach einer 5-teiligen Skala: 1=„sehr gut“; 2=„gut“, 3=„befriedigend“, 4=„genügend“ und 5=„nicht genügend“. Die Prüfung war bestanden bei Note 1 bis 4, bei Note 5 war eine Wiederholung notwendig.

Die Auswertungs- und Interpretationsobjektivität des Messinstruments HNO-Rigorosum ist durch die Beurteilung unabhängiger Zweitbegutachter gegeben. Die Durchführungsobjektivität hätte durch den Einsatz von Zweitprüfern verbessert werden können. Dafür waren allerdings keine Ressourcen vorhanden. Die niedrige Reliabilität liegt zum Teil daran, dass der Prüfer aktiv am Prüfungsgeschehen teilnimmt [18] und ist auch Grund für die niedrige Validität [19]. Eine Evaluation des Repetitoriums wurde nicht durchgeführt.

2.5 Statistische Auswertung

Die Datenanalyse erfolgte mit dem „Statistical Package for the Social Science – SPSS“ für Windows in der Version 20. Die Prüfungsnoten der Prüfungsereignisse wurden im arithmetischen Mittel und Median für das gesamte Kollektiv, für die Versuchsgruppe mit Repetitorium, für die Vergleichsgruppe ohne Repetitorium sowie bei allen

Gruppen nach Geschlecht der Prüfer, errechnet. Zudem wurde der prozentuelle Anteil der einzelnen Notenwerte der Rigorosumprüfungen gesamt und nach dem Geschlecht des Prüfers bestimmt. Der Einfluss der Faktoren Repetitorium (ja/nein), Geschlecht des/der Prüfenden (weiblich vs. männlich) und Geschlecht des Studenten (weiblich vs. männlich) wurde mittels einfaktorieller Varianzanalyse getestet, wobei das Geschlecht für die Berechnung umkodiert wurde. Dem multiplen Testen Rechnung tragend wurden alle diese drei Faktoren zudem in einer multifaktoriellen Varianzanalyse wiederholt getestet. Der Einfluss auf die Notengebung wurde des Weiteren in linearen Regressionsanalysen dargestellt. Die schließende Statistik bestehend aus ein- und multifaktoriellen Varianzanalysen sowie uni- und multivariaten Regressionsmodellen wurde mit dem Open Source Statistikprogramm „R version 3.1.2“ durchgeführt [https://www.r-project.org/].

3. Ergebnisse

3.1 Teilnehmer

Insgesamt wurden in dieser Studie 3961 Einzel-Prüfungsergebnisse aus einer Institution (Universitätsklinik für HNO-Krankheiten der MedUni Wien) ausgewertet. Aus derselben Institution stammen auch die 16 männlichen und 6 weiblichen Prüfer.

3.2 Prüfungsergebnisse

Die Prüfungsergebnisse sind für alle Prüfer und getrennt nach deren Geschlecht für die Versuchsgruppe (mit Repetitorium) und Vergleichsgruppe (ohne Repetitorium) in Tabelle 1 dargestellt.

3.2.1 Notendurchschnitt im Gesamtkollektiv

Im Gesamtkollektiv der Prüfer wurde die Versuchsgruppe mit Repetitorium ($n=725$; Note: $2,54 \pm 1,38$) gegenüber der Vergleichsgruppe ohne Repetitorium ($n=3236$; Note: $2,46 \pm 1,34$) tendenziell schlechter benotet - statistisch war in der einfaktoriellen Varianzanalyse aber keine Signifikanz festzustellen (ANOVA: $Df=1$; Quadratsumme=4,2; Mittel der Quadrate=4,1559; $F=2,2865$; $p=0,1306$). In der univariaten linearen Regression ergab sich ein statistisch nicht signifikanter Effekt von 0,08376 (p des Wald-Tests=0,131) für schlechtere Noten bei Studenten die das Repetitorium absolviert haben.

3.2.2 Abhängigkeit der Prüfungsergebnisse vom Geschlecht der Prüfer

Die Notenwerte männlicher Prüfer waren im arithmetischen Mittel 2,45 mit einer Standardabweichung von 1,36 bei einer Fallzahl von 3376. Die Notenwerte weiblicher Prüfer waren im arithmetischen Mittel 2,60 mit einer Standardabweichung von 1,28 bei einer Fallzahl von 585.

Tabelle 1: Arithmetisches Mittel und Median aller Prüfungsereignisse, der Versuchsgruppe mit Repetitorium und der Vergleichsgruppe ohne Repetitorium, getrennt nach Geschlecht der Prüfer.

Prüfungen	alle			mit Repetitorium			ohne Repetitorium		
	n	Noten Ø	Median	n	Noten Ø	Median	n	Noten Ø	Median
16 Männer	3376	2,45	2	622	2,54	2	2754	2,44	2
6 Frauen	585	2,60	2	103	2,58	2	482	2,61	2
Summe	3961	2,48	2	725	2,55	2	3236	2,46	2

Weibliche Prüfer vergaben signifikant schlechtere Noten als ihre männlichen Kollegen (Wilcoxon-Rangsummentest: $W=910610$, $p=0,001926$).

3.2.3 Abhängigkeit der Prüfergebnisse vom Geschlecht der Studenten

Studentinnen wiesen einen Notendurchschnitt von $2,45 \pm 1,34$ auf, ihre männlichen Kollegen $2,51 \pm 1,36$. In der einfaktoriellen Varianzanalyse erwies sich das Geschlecht der Studenten nicht als signifikanter Einflussfaktor (ANOVA: $Df=1$; Quadratsumme=3,6; Mittel der Quadrate=3,64; $F=2,0028$; $p=0,1571$). Im univariaten linearen Regressionsmodell war der Effekt des weiblichen Geschlechts mit 0,06233 auf den Notendurchschnitt als nicht signifikant einzustufen (p des Wald-Test=0,157).

3.2.4 Multifaktorielle Varianzanalyse

Dem multiplen Testen Rechnung tragend wurden alle drei Faktoren in einer multivariaten Regressionsanalyse (siehe Tabelle 2) bzw. einer multifaktoriellen Varianzanalyse (siehe Tabelle 3) getestet. Lediglich das Geschlecht der Prüfer erwies sich weiterhin als signifikanter Einflussfaktor. Durchschnittlich vergaben weibliche Prüferinnen um 0,146 Punkte schlechtere Noten.

3.2.5 Notendurchschnitt, Anzahl und Prozent der „Nicht genügend“ dargestellt nach Geschlecht der Kandidaten und Prüfer

Eine besondere Gewichtung kommt der Note 5 (nicht bestanden) zu. Der Prozentsatz variiert von maximal 12,23% bei der Kombination weibliche Kandidaten/weiblicher Prüfer bis 8,13% bei der Kombination männliche Kandidaten/weiblicher Prüfer (siehe Tabelle 4). Durchschnittlich über alle Kombinationen ist der Prozentsatz 10,2 und damit Ausdruck für eine sehr konstante Bewertung aller Prüfer wenn es darum geht ob der Prüfling bestanden hat oder nicht.

4. Diskussion

4.1 Bedeutung der Ergebnisse und Vergleich mit der existierenden Literatur

Im Gesamtkollektiv der Prüfer wurde die Versuchsgruppe mit Repetitorium ($n=725$, Note: $2,54 \pm 1,38$) gegenüber der Vergleichsgruppe ohne Repetitorium ($n=3236$, Note:

$2,46 \pm 1,34$) tendenziell schlechter benotet, statistisch war in der einfaktoriellen Varianzanalyse aber keine Signifikanz festzustellen und in der univariaten linearen Regression ergab sich dementsprechend ein statistisch nicht signifikanter Effekt von 0,08376 schlechteren Noten bei Studenten mit Repetitorium.

In der Literatur wird im Zusammenhang mit der Durchführung von Repetitorien immer von einer hohen Effizienz und einem hohen Grad der Annahme durch die Teilnehmer berichtet. Allerdings wird auch darauf hingewiesen, dass der damit verbundene Aufwand nicht zu unterschätzen sei. Dieser werde aber durch die gute Prüfungs- bzw. Berufsvorbereitung der Studenten aufgewogen [2].

Repetitorien können durchwegs als nützliche Lehrveranstaltungen in verschiedenen Formen betrachtet werden. Studierende geben an, ihr Wissen sei besser [6], bzw. signifikant besser [14] als vorher und ein Repetitorium sei effektiv als Vorbereitung für staatliche Prüfungen [10]. Die Studierenden seien nicht verängstigt sondern motiviert zum eigenständigen Weiterlernen. Die Noten seien besser und signifikant weniger Studierende hätten die nachfolgende Prüfung nicht bestanden. Die zeitliche Befristung Sorge für kontinuierliches Lernen. Schwache und gute Studenten würden gleichermaßen gefördert [1]. Repetitorien seien effizient, kostengünstig und anpassbar für eine große Anzahl von Lernenden [3]. Bei der Prüfung von praktischen Fähigkeiten sei es zur Verbesserung der Prüfungs-Performance mit signifikanter Steigerung des Erfolges gekommen [5]. Unsere Ergebnisse stehen im Gegensatz zur einschlägigen Literatur.

Studentinnen wiesen einen Notendurchschnitt von $2,45 \pm 1,34$ auf, ihre männlichen Kollegen $2,51 \pm 1,36$. In der einfaktoriellen Varianzanalyse erwies sich das Geschlecht der Studenten nicht als signifikanter Einflussfaktor. Im univariaten linearen Regressionsmodell war der Effekt des weiblichen Geschlechts mit -0,06233 auf den Notendurchschnitt als nicht signifikant einzustufen. Diese Ergebnisse stimmen mit bereits in der Literatur erwähnten überein [12], [15], [16], [17], [18].

4.2 Das HNO-Repetitorium hat nicht zu einer Verbesserung des Notendurchschnittes bei der nachfolgenden mündlichen Prüfung über Inhalte des gesamten Faches geführt.

Der Dozent des Repetitoriums war auch gleichzeitig Prüfer. Ein Bias im Sinne „Vorlieben des Prüfers erkennen beim Repetitorium“ lag nicht vor. Die durchschnittlichen Notenwerte im Vergleich zu Studenten ohne Repetitorium waren nur unwesentlich besser.

Tabelle 2: Multivariantes Regressionsmodell der Noten geschätzt durch Geschlechter von Prüfer und Student sowie Teilnahme am Repetitorium. Die Durchschnittsnote entspricht der Referenzgruppe männlicher Studentengeprüft durch männliche Prüfer.

	Effekt	Standardfehler	t-Wert	p(> t)
Durchschnittsnote	2,47462	0,03743	66,122	<2e-16***
Prüferinnen	0,14589	0,06041	2,415	0,0158*
Studentinnen	-0,05736	0,04406	-1,302	0,1931
Repetitorium	0,08500	0,05536	1,535	0,1247

*** Signifikanzgrenze: 0,001 * Signifikanzgrenze: 0,05

Tabelle 3: Multifaktorielle Varianzanalyse der Noten. Die Spalten sind in der Ordnung von links nach rechts: Faktor, Freiheitsgrade, QS Quadratsumme, MQS mittlere Quadratsumme, F-Wert, p(>F) Wahrscheinlichkeitswert.

Faktor	Freiheitsgrad	QS	MQS	F-Wert	P(>F)
Prüferinnen	1	11,1	11,0746	6,1020	0,01354 *
Studentinnen	1	3,1	3,0749	1,6943	0,19312
Repetitorium	1	4,3	4,2585	2,3464	0,12565
Restwert	3957	7181,6	1,8149		

* Signifikanzgrenze: 0,05

Tabelle 4: Notendurchschnitt, Anzahl und Prozent der der „Nicht genügend“ dargestellt nach Geschlecht der Kandidaten und Prüfer

	n	Noten Ø	n: nicht genügend	Prozent
Weibliche Kandidaten/männliche Prüfer	2113	2,43	203	9,61
Weibliche Kandidaten/weibliche Prüfer	327	2,61	40	12,23
Männliche Kandidaten/männliche Prüfer	1263	2,5	140	11,08
Männliche Kandidaten/weibliche Prüfer	258	2,59	21	8,13
Summe	3961	2,48	404	10,2

Eine Differenzierung in leistungsstarke und -schwache Studenten wurde nicht durchgeführt.

Warum in unserer Studie der Notendurchschnitt der Versuchsgruppe, die am Repetitorium teilnahm, im Vergleich zur Vergleichsgruppe, die kein Repetitorium besuchte, keineswegs - entgegen der Erwartung - besser war, sondern tendenziell sogar schlechter, erklären sich die Autoren folgendermaßen: Studenten betrachteten das Repetitorium als Ersatz für das Selbststudium und nicht als Ergänzung und Wiederholung, und als Zeitersparnis - immerhin ist das HNO-Rigorosum eine der letzten Prüfungen im Studium. Außerdem könnte ein relativ großer zeitlicher Abstand zwischen der Teilnahme am Repetitorium und der Rigorosum-Prüfung entstanden sein und die für die Prüfung notwendigen Prüfungsinhalte wurden nur zu 50-60% im Repetitorium präsentiert. Es ist nicht von der Hand zu weisen, dass möglicherweise der Effekt des „Bulimielernens“ durch das Repetitorium gefördert wurde. Das bedeutet: Lernen für eine Einzelprüfung mit Abspeicherung im Kurzzeitgedächtnis, „Ausspucken“ des Wissens bei der Prüfung, um danach ohne „Verdauung“, wieder zu vergessen. Die Motivation für die Teilnahme am Repetitorium wurde nicht explizit erfragt. In Betracht kommen dafür:

- Angebot, in kurzer Zeit ein straffes Lehr/Lernprogramm zu erfahren
- Interesse am HNO-Fach
- strukturierten interaktiven, fallorientierten Unterricht bei einem Dozenten mit hoher fachlicher und didaktischer Kompetenz erleben zu wollen
- Ausgleich der fehlenden Auseinandersetzung mit Lehrinhalten
- „Fernstudium“ ohne Besuch der Vorlesungen
- Förderung des eigenständigen Weiterlernens
- Beseitigung von Angst und Unsicherheit vor der Prüfung.

4.3 Die durchschnittlichen Notenwerte bei weiblichen Prüfern sind signifikant schlechter als bei männlichen Prüfern.

Verzerrungen der Prüfungsergebnisse durch Prüfer sind ein latentes Risiko bei Prüfungen. Menschen unterscheiden sich in vielen Verhaltensweisen und es ist nicht überraschend, dass Prüfer Tendenzen zur Bevorzugung oder Benachteiligung bei der Benotung von Kandidaten zeigen. Wenn dies bekannt ist, sollte frühzeitig dagegen Maßnahmen ergriffen werden. Diese reichen von randomisierter Kandidatenzuteilung bis zu kommissionellen Prüfungen oder sogar der Aufgabe der mündlichen zugunsten einer schriftlichen Prüfung [20]. Eine mögliche Ursache für unterschiedliche geschlechtsspezifische Benotung könnte in der Wirkung einer oder mehrerer anderer Probandenmerkmale liegen, die bisher nicht untersucht wurden. Diese Variablen könnten in unterschiedlichem Maße bei männlichen und weiblichen Prüfern wirken, sodass unterschiedliche Zusammenhänge zu erklären wären [15]. In Betracht kommen die freie Prüferwahl, das Alter der Kandidaten, die Prüfungserfahrung der Prüfer, die Sympathie oder Vorbehalte dem Prüfling gegenüber. Wiskin et al. [21] beschreiben in ihrer Publikation zum Thema Gender als Variable beim OSCE der Kommunikationsfähigkeit im Niveau des letzten Ausbildungsjahres, dass weibliche Prüfer besser benoten und geben als Behebung die strategische Gleichstellung von männlichen und weiblichen Prüfern an. Auch Boehm et al. [12] berichten über das Rigorosum im Fach Sozialmedizin eine bessere Benotung durch weibliche Prüfer. McManus et al. [22] sehen als Lösung der Problematik, bei einer spezifischen Verzerrung zum Nachteil von Personen mit individuellen Eigenschaften wie Geschlecht oder ethnischer Zugehörigkeit bei einer praxisbezogenen Prüfung pro Station gleichzeitig zwei Prüfer einzusetzen. Mit dem

multi-faktoriellen Rasch-Modell schätzen sie den Effekt, dass Prüfer bevorzugen oder verschärfen, und gleichen durch die Paarung der am wenigsten und am höchsten strengen Prüfer aus. Trotz vieler möglicher Bias haben die Rigorosumsprüfer, die an dieser Studie teilnahmen, diese Prüfungsform geschätzt. Denn sie bot die Möglichkeit, deklaratives und prozedurales Wissen gleichzeitig symptom- oder fallbezogen gemeinsam mit Grundlagenwissen zu prüfen. Immerhin ergab sich bei der Durchfallquote kein signifikanter Unterschied zwischen weiblichen und männlichen Prüfern.

4.4 Stärken und Schwächen der Studie

4.4.1 Stärken

Als Stärken der retrospektiven Studie sehen die Autoren das klare Konzept. Die insgesamt hohe Anzahl der Probanden ermöglichte klare statistische Aussagen. Im langen Beobachtungszeitraum von 5 Jahren wurden alle 22 tätigen Prüfer mit einbezogen.

4.4.2 Schwächen

Das Design ist zwar einfach und wenig robust, jedoch schnell durchzuführen und kostengünstig. Ein strukturiertes Feedback von Studierenden und Prüfern wurde nicht durchgeführt. Damit hätte man näher auf personenabhängige Verzerrungen bei der Interaktion zwischen Prüfern und Geprüften eingehen können. Es wurde nicht festgestellt, wie oft Studierende am Repetitorium teilnahmen und in welchem Abstand zum Prüfungstermin. Die ca. 10% Prüfungsrepetenten wurden nicht gesondert in der Auswertung berücksichtigt. Mit einer Evaluation des Repetitoriums wäre die „Reaktionsebene“ getrennt von der „Lernebene“ [23], [24] möglich geworden.

5. Schlussfolgerung

Die Begründung für diese Studie war es, den Nutzen und die Beeinflussungen durch das Geschlecht der Prüfer des HNO-Rigorousums an der Medizinischen Universität Wien zu untersuchen. Das HNO-Repetitorium hat nicht zu einer Verbesserung des Notendurchschnittes beim Rigorosum zum Fach HNO-Heilkunde geführt. Die durchschnittlichen Notenwerte bei weiblichen Prüfern sind signifikant schlechter als bei männlichen Prüfern. Der wesentliche Unterschied zwischen männlichen und weiblichen Prüfern lag bei der Vergabe der Noten 1 und 2 und nicht bei der Vergabe der Note 5 (nicht bestanden). Sehr konstant gelang bei der Prüfung allen Prüfern jene Kandidaten zu bestimmen, die nicht bestanden haben. Als relative Limitierung sind die niedrige Reliabilität der Studie und damit auch der Grund für eine niedrige Validität zu sehen. Aufgrund der Einführung der neuen Studienordnung für Medizin 2002, deren Prüfungsordnung keine Rigorosen mehr vorsieht, werden dafür keine Repetitorien mehr abgehalten. Allerdings gibt es im 1. Studienjahr Repetito-

rien für die Jahrgangsprüfung bei denen die Studierenden bisher nicht namentlich erfasst werden.

Anmerkung

In dieser Publikation wurde für Personen der Einfachheit halber die männliche Form verwendet. Es sind aber selbstverständlich immer beide Geschlechter gemeint.

Danksagung

Die Autoren danken allen Lehrenden der Universitätsklinik für Hals-Nasen-Ohrenheilkunde der Medizinischen Universität Wien, die sich an dieser Studie als Prüfer beteiligt haben.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Literatur

1. Störmann S, Chiapponi C, op den Winkel M, Wöck M, Bender J, Kern AB, Gebhardt C, Angstwurm M. Mit dem Internet zum Examen - Prüfungsvorbereitung mit dem virtuellen Staatsexamen-Repetitorium der LMU München. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA). Bochum, 23.-25.09.2010. Düsseldorf: German Medical Science GMS Publishing House; 2010. Doc10gma143. DOI: 10.3205/10gma143
2. Kühn J, Jabs WJ. Das Lübecker Repetitorium "Innere Kompakt": ein Pilotprojekt zur Vorbereitung auf das neue zweite Staatsexamen. GMS Z Med Ausbild. 2007;24(3):Doc148. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2007-24/zma000442.shtml>
3. Hibbert EJ, Lambert T, Carter JN, Learoyd DL, Twigg S, Clarke S. A randomized controlled pilot trial comparing the impact of access to clinical endocrinology video demonstrations with access to usual revision resources on medical student performance of clinical endocrinology skills. BMC Med Educ. 2013;13:135. DOI: 10.1186/1472-6920-13-135
4. Lynn JS, Mostyn A. Audience response technology: Engaging and empowering non-medical prescribing students in pharmacology learning. BMC Med Educ. 2010;10:73. DOI: 10.1186/1472-6920-10-73
5. Meade O, Bowskill D, Lynn JS. Pharmacology as a foreign language: A preliminary evaluation of podcasting as a supplementary learning tool for non-medical prescribing students. BMC Med Educ. 2009;9:74. DOI: 10.1186/1472-6920-9-74
6. Mole G, Gillespie L. Those who can, teach. Accessing medical students' perception of a finals revision programme delivered by foundation and core trainees: a cross-sectional study. BMJ Open. 2012;2:e001544. DOI: 10.1136/bmjopen-2012-001544
7. Freimanis AK. Successful teaching of radiology of medical students and interns. With emphasis on third and fourth-year programs. Radiology. 1970; 97(3):669-680. DOI: 10.1148/97-3-669
8. Swartz TH, Lin JJ. A clinical refresher course for medical scientist trainees. Med Teach. 2014;36(6):475-479. DOI: 10.3109/0142159x.2014.886767

9. Brown G, Manogue M. AMEE Medical Education Guide No. 22: Refreshing lecturing: a guide for lecturers. *Med Teach*. 2001;23(3):231-244. DOI: 10.1080/01421590120043000
10. Rengier F, Rauch PJ, Partovi S, Kirsch J, Nawrotzki R. A three-day anatomy revision course taught by senior peers effectively prepares junior students for their national anatomy exam. *Ann Anat*. 2010;192(6):396-399. DOI: 10.1016/j.aanat.2010.02.008
11. Bredemeier S, Pabst R, Nave H. Der "Makro-Marathon": Ergebnisse der Evaluation eines Repetitoriums der makroskopischen Anatomie zur Vorbereitung auf das Physikum. *GMS Z Med Ausbild*. 2005;22(4):Doc197. Zugänglich unter/available from: <http://www.egms.de/static/de/journals/zma/2005-22/zma000197.shtml>
12. Boehm G, Bernhard G, Kwizda-Gredler B, Kunze U, Rathmanner T, Rieder A, Schoberberger R, Schwarz B, Vutuc C, Kunze M. Einfluss von Geschlecht und Studiengebühren auf die Noten bei Rigorosum-Prüfungen im Prüfungsfach Sozialmedizin. *Wien: Didaktik*; 2001. Zugänglich unter/available from: <http://didaktikon.net/cgi-bin/didaktik.cgi?id=0000053>
13. Bundesministerium für Wissenschaft und Forschung. Studienplan für die Studieneinrichtung Medizin an der Medizinischen Fakultät der Universität Wien: § 11 Repetitorien: Auf Empfehlung der Studienkommission sind Repetitorien einzurichten, wenn dies aus Gründen der Vollständigkeit des Lehrangebotes (§ 58 lit.b UOG) oder zur Vermeidung von Studienverzögerungen (§ 58 lit.j UOG) erforderlich ist. Wien: Bundesministerium für Wissenschaft und Forschung; 1992.
14. Webb AL, Choi S. Interactive radiological anatomy eLearning solution for first year medical students: Development, integration, and impact on learning. *Anat Sci Edu*. 2014;7(5):350-360. DOI: 10.1002/ase.1428
15. Denney ML, Freeman A, Wakeford R. MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source? *Br J Gen Pract*. 2013;63(616):e718-725. DOI: 10.3399/bjgp13X674396
16. McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP(UK) PACES and nPACES examinations. *BMC Med Educ*. 2013;13:103. DOI: 10.1186/1472-6920-13-103
17. Ministerium für Kultus und Unterricht. Das Studium an der medizinischen Fakultät. Die medizinische Rigorosenordnung. Verordnung des Ministers für Kultus und Unterricht vom 21. Dezember 1899, RGBl. Nr. 271. Wien: Ministerium für Kultus und Unterricht; 1899.
18. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach*. 2005;27(4):294-297. DOI: 10.1080/01421590500126437
19. Schuhwirth LW, van der Leuten CP. Quality control: assessment and examinations. *Z Hochschuldidaktik*. 1996;20(1):66-76.
20. Niehaus DJ, Jordaan E, Koen L, Mashile M, Mall S. Applicability and fairness of the oral examination in undergraduate psychiatry training in South Africa. *Afr J Psychiatry (Johannesbg)*. 2012;15(2):119-123. DOI: 10.4314/ajpsy.v15i2.16
21. Wiskin CM, Allan TF, Skelton JR. Gender as a variable in the assessment of final year degree-level communication skills. *Med Educ*. 2004;38(2):129-137. DOI: /10.1111/j.1365-2923.2004.01746.x
22. McManus IC, Thompson M, Mollon J. Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling. *BMC Med Educ*. 2006;6:42. DOI: 1186/1472-6920-6-42
23. Jünger J, Just I. Empfehlungen der Gesellschaft für Medizinische Ausbildung und des Medizinischen Fakultätentags für fakultätsinterne Leistungsnachweise während des Studiums der Human-, Zahn- und Tiermedizin. *GMS Z Med Ausbild*. 2014;31(3):Doc34. DOI: 10.3205/zma000926
24. Kirkpatrick D. *Evaluating Training Programs: The Four Levels*. 2nd ed. San Francisco/CA: Berrett-Koehler Publishers Inc.; 1998.

Korrespondenzadresse:

Ao. Univ.-Prof. Dr. Matthäus C. Grasl, MME
Medizinische Universität Wien, Universitätsklinik für Hals-, Nasen- und Ohrenkrankheiten, Währinger Gürtel 18-20, A-1090 Wien, Österreich
matthaeus.grasl@meduniwien.ac.at

Bitte zitieren als

Grasl MC, Seemann R, Hanisch M, Heiduschka G, Kremser K, Thurnher D. Influence of a revision course and the gender of examiners on the grades of the final ENT exam – a retrospective review of 3961 exams. *GMS Z Med Ausbild*. 2015;32(4):Doc38. DOI: 10.3205/zma000980, URN: urn:nbn:de:0183-zma000980

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2015-32/zma000980.shtml>

Eingereicht: 18.09.2014

Überarbeitet: 14.09.2015

Angenommen: 18.09.2015

Veröffentlicht: 15.10.2015

Copyright

©2015 Grasl 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/>.