Ten Years of Objective Structured Clinical Examination at the Medical Faculty of Tübingen, Germany: Item Analysis and Students' Satisfaction

Graf J[1], Smolka R[2], Holderried F[3], Wosnik A[1], Lammerding-Köppel M[4], Mohr D[1], Vlad E[1], Nikendei C[5], Zipfel S[1,6] and Herrmann-Werner A[6,7]

[1]*Dean's Office for Students' Affairs, Medical Faculty Tübingen, University of Tübingen, Germany*

[2]*Hospital for Psychosomatic Medicine and Psychotherapy, DRK Clinic Center Berlin, Germany*

[3]*Department of Gastroenterology, Hepatology and Infectious Diseases, University Hospital Tübingen, Germany*

[4]*Competence Center for University Teaching in Medicine, Medical Faculty Tübingen, University of Tübingen, Germany*

[5]*Department for General Internal Medicine and Psychosomatic Medicine, University Hospital Heidelberg, Germany*

[6]*Department of Psychosomatic Medicine and Psychotherapy, University Hospital Tübingen, Germany*

[7]*Interdisciplinary Training Centre DocLab, Medical Faculty Tübingen, University of Tübingen, Germany*

Abstract

Background: Despite the high importance of communicative and clinical-practical skills for practicing medicine, the examination format named Objective Structured Clinical Examination (OSCE) has also been established at the medical faculties in Germany. **Objectives:** Due to the 10th anniversary of Tübingen OSCE, this article aimed to outline and analyses trends of average grades, students' evaluations, item-total correlation and item difficulty over the disciplines and semesters as well as examination's reliability in the context of quality review and the discussion about measurement of learning success. **Subjects and Methods:** For computation of the average scores and students' evaluation, average values and dispersion measures were calculated over all OSCE examinations for each semester (n=3,261 students), whereby 20 semester cohorts of students of the sixth clinical semester were compared. Furthermore, the mean values and the range of the severity of item-total correlation and item difficulty were calculated from the subject-related values for each semester. In addition the intra-class correlation coefficient was calculated per semester. **Results:** The average scores across all disciplines, examinations and semesters were 20.1 with only slight fluctuations, while the students evaluated the examination format consistently good as well. Item difficulty, item-total correlation and reliability were acceptable over all disciplines over the semesters. **Conclusions:** Due to the very low (non-significant) differences in the average scores over the semesters, an equally high degree of skills acquisition can be assumed. Item difficulty, item-total correlation and reliability corresponded to the results of the literature.

Keywords: Objective Structured Clinical Examination; Communication and practical skills; Trends of the average grades; Student's evaluation; Reliability

Introduction

Communicative skills in medical schools

Despite the high importance for practicing medicine [1,2] communicative and clinical-practical skills still only play a minor role in medical schools, but are increasingly implemented in the teaching curricula [3]. Since the communication of doctors is considered as a complex skill which is not easily acquired, the foundations of the development of subject-specific communicative skills should preferably be taught during the pre-clinical part of medical school [4].

Objective structured clinical examination

Due to this fact, the examination format named Objective Structured Clinical Examination (OSCE) has also been established at the medical faculties in Germany since the turn of the millennium. The OSCE was first described by

Harden et al. [5] in 1975 and has been used in Anglo-American states since the 1980s [6,7]. During the OSCE examination, students complete a course of examination stations similar to circuit training, where they have to demonstrate their communicative and clinical-practical skills. During a set time, different standardized tasks must be solved, for example physical examinations, history talking or diagnoses, and specially trained standardized patients (SP) are used in addition. Every station is equipped with an examiner who assesses the students' examination performance objectively and in a standardized manner [8,9].

OSCE exams in Tübingen

At the Medical Faculty of Tübingen, the OSCE format was introduced in early 2004, and since then, OSCE examinations have been taking place each semester for students in their 6th semester. The examination procedure is currently used in 9 clinical subjects (dermatology, general medicine, internal medicine, neurology, orthopaedics,

paediatrics, psychosomatics, radiology and surgery). The OSCE examinations always take place every semester at the beginning of the semester break, with the number of specialist stations varying from 10 to 16. The content of the stations is based on the teaching content of the corresponding curricular examination courses and students can practice what to expect in a mock OSCE one week before the actual one [10]. The OSCE in Tübingen includes history talking stations (e.g. a focused anamnesis for diarrhoea), stations with physical examinations (e.g. of the shoulder), as well as stations where an interpretation of the findings is necessary (e.g. chest X-ray pneumonia), all with 6-minute examination sequences. For the stations with history talking and physical examination, SPs are used whose task is to present with various symptoms and

syndromes to the medical students [11]. Until 2013, the OSCE in Tübingen was paper-based, whereupon, the OSCE was turned into a tablet-based examination [12]. As recommended by the literature, the maximum score to be awarded at each station is 25 points [6-8], while the examination coordinators are relatively free in preparing the tasks and assigning the points, depending on their specialist field. For stations with SPs, the OSCE coordination defined a uniform communication block (5 points), which is why only a maximum of 20 con-tent-related points can be awarded at those stations (Table 1).

Communication	Points (max. 5)
Greeting: Student asks for patient`s name (1 Point)	
Presentation: Student introduces himself with name and function (1 Point)	
Pat. adapted style of speech: Most of the time sufficiently slow, clear language, adequate questions / instructions, no (unexplained) foreign words (1 point)	
Attendance in history taking/in physical examination: Most of the time favorable discussion by means of gesticulation and facial expression, empathetic appearance, eye contact, excuses (1 point)	
Clear structure ("common theme") in the conversation/in the in the physical examination: Logically comprehensible procedure of the anamnesis/in the physical examination (1 point)	

Table 1: Communication block for stations with simulated patients.

Also in Tübingen, the OSCE examination is conducted in order to assess practical and communicative competences of medical students [13]. However, it remains uncertain how reliably the examination in Tübingen can measure and depict students' competences. As with many methods for measuring theoretical constructs (especially in educational research), it remains open how suitable the measuring methods are for measuring the features to be determined.

Aims

To mark the 10th anniversary of the OSCE examination at the Medical Faculty of Tübingen, an analysis was requested of how the average score determined per semester for all OSCE stations, as well as the students' evaluations of the examination format had changed over the course of 10 years in order to be able to make statements about the variability of the level of competence and about the satisfaction.

Material and Methods

Research questions

By determining various psychometric key figures, it should be investigated to what extent the examination is suitable for the measurement of communicative and practical skills (item analysis: item-total correlation and item difficulty) and how accurately and reliably the competences can be verified within the framework of the OSCE examination (reliability). All in all, four research questions should be answered: (1) How did the average score of all OSCE-stations develop over the course of the semesters? (2) What variability was found in the students' evaluations over the course of the semesters? (3) How suitable is the examination for measuring

students' competences and what degree of variability is shown when comparing semesters? (4) How reliable is a measurement of students' competences within the frame-work of the OSCE examination?

Study design

The design was conceived as a longitudinal trend study without any randomization since all students who passed OSCE in the evaluation period were included. Trend studies (also called replicative surveys) represent the third subtype of longitudinal analyses (in addition to cohort and panel studies). A trend study samples different groups of people at different points in time but in the same situation and from the same population [14]. The results of all OSCE examinations conducted by the Medical Faculty of Tübingen within the evaluation period were included in the analysis (summer semester (SS) 2004 to winter semester (WS) 2013/2014), as well as the students' evaluations within the same period. A total of n=3,261 students who passed the OSCE examination during the evaluation period were included in the analysis (average number per semester: n=163). All students were in their regular 6th semester at the time of passing the OSCE examination so the study compares the examination results of students in their 6th semester but from different age cohorts (trend study/ replicative survey). Since the average grades are to be determined for all students and subjects, a total of n=20 semester cohorts were available. During an overall evaluation period of 10 semesters, the students were always asked to evaluate the examination at the end of the OSCE (SS 2004, WS 2004-2005; WS 2007-2008; SS 2008; SS 2011 to WS 2013-2014). A total of 2,756 surveys were available for the analysis (average amount per semester: n=138), which resulted in a response rate of 84.5%.

Ethical approval

In accordance with the requirements of the university, no approval from the ethics committee had to be obtained, as only students but no patients were interviewed. Patient data were not used in the present study. Consent to participate was sought and gained from all participants in the study.

Questionnaires

For all semesters of the evaluation period, the examination statistics of every discipline (single grades of all students per subject) were available, as well as the students' standardized surveys consisting of 3 questions. The survey was developed in-house and is now also being used in courses which use simulated patients to prepare students for examining patients. In the survey, the students were asked to evaluate organization, content and SP (Table 2). The assessment was based on a 6-step Likert-scale (1=satisfactory to 6=unsatisfactory).

Organization								
Organization (How well did you know about the test? How well was the exam organized?)	appropriate	1	2	3	4	5	6	inappropriate
Evaluation of OSCE's stations								
Content (severity of the task, wording, materials)	appropriate	1	2	3	4	5	6	inappropriate
SP	very good	1	2	3	4	5	6	inadequate

Table 2: Survey for the students' evaluation.

Implementation

Using the examination results provided by the departments, the average score for all students across all subjects was calculated for each semester. The practical and communicative competences were not considered separately, but evaluated interdisciplinary since the comparison of the average total score should be presented (research question 1). The students' evaluations of the OSCE examination always took place after the completion of the examination, and an average value was determined for each semester. The survey results were transferred to an Excel spreadsheet and then evaluated descriptively (research question 2). As part of the usability test of the OSCE examination to determine the communicative and practical skills, item- and reliability analyses were carried out (research questions 3 and 4). In each semester, the item difficulties and the item-total correlation were calculated separately for all subjects. To determine the item-total correlation, the individual items in each subject were correlated with the overall result. Due to examination law reasons, the interdisciplinary averages were then calculated for each of the mentioned psychometric key figures for each semester. The calculation of reliability was carried out per semester by comparing the individual subject items.

Statistics

In order to check how the average scores for each semester of all OSCE stations, as well as how the students' evaluations of the examination format developed over the 10-year course, a mean value analysis was carried out. For each semester, a mean total was calculated from the respective average score per subject, which was presented together with the largest and the smallest subject-related mean value. Finally, the respective mean values per semester were compared. The analysis of the identified differences for statistical significance was carried out by using the univariate analysis of variance (ANOVA).

The variables were checked in advance for normal distribution (using the Kolmogorov-Smirnov adaptation test) and variance homogeneity (using the Levene test) (research question 1). The students' surveys were also evaluated descriptively, and the mean values and the standard deviation were determined for each case. Here, too, any differences were analyzed for statistical significance by using the univariate ANOVA (research question 2). In order to check how the psychometric key figures developed in a comparison across all semesters, the item-total correlation and the item difficulty were determined separately every semester for each examination. Then, the averages were determined for each area and following this, total mean values were determined from all subject-related mean values for all 20 semesters, which are then presented as a synopsis in order to provide a direct comparison of semesters (here, too, the respective total mean value is shown, as well as the largest and smallest subject-related mean value as a dispersion measure) (research question 3). Finally, the intra-class correlation coefficients as well as the corresponding 95% confidence intervals were determined for each semester as part of a reliability analysis (research question 4). For all statistical tests, a p-value of <0.05 was considered to be statistically significant bilaterally ($\alpha = 0.05$). All statistical calculations were done with Statistical Package for the Social Sciences (SPSS) 22, while the images were created in MS Excel 2010.

Results

Development of the average scores

Table 3 and Figure 1 show the development of the average scores over the past 10 years. When only looking at the descriptive frequency distribution, the average score across all subjects and students (with a maximum score of 25 points) increased slightly over the course of 10 years, from 19.5 points to a current score between 20 and 20.5 points. However, there were no significant differences between the semesters in the context of ANOVA.

Development of the student satisfaction

As shown in Figure 2, the average scores for student satisfaction are at a consistently high level. Overall, the

evaluation was able to show consistently high satisfaction scores for the examination format on the part of the students, whereby the differences across the semesters were so small that no statistically significant differences could be identified in the context of ANOVA, so here, these were random variations, too.
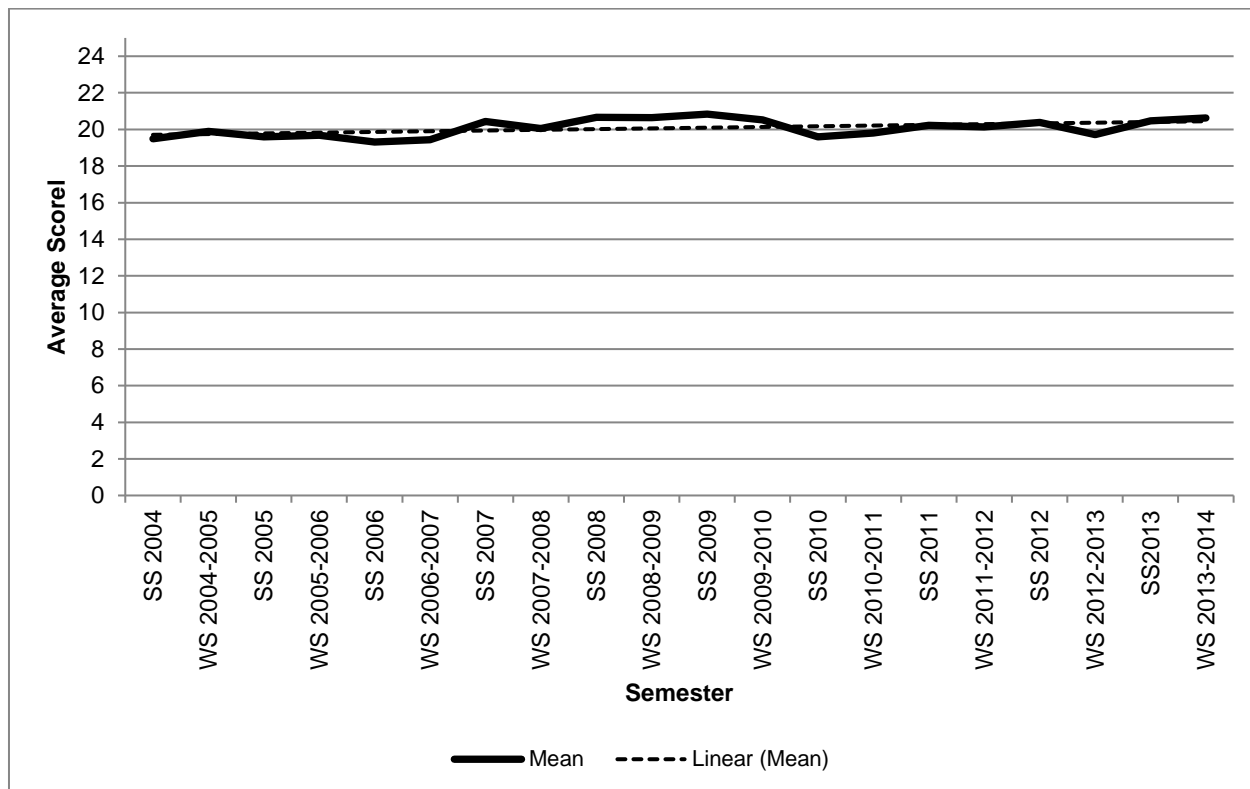

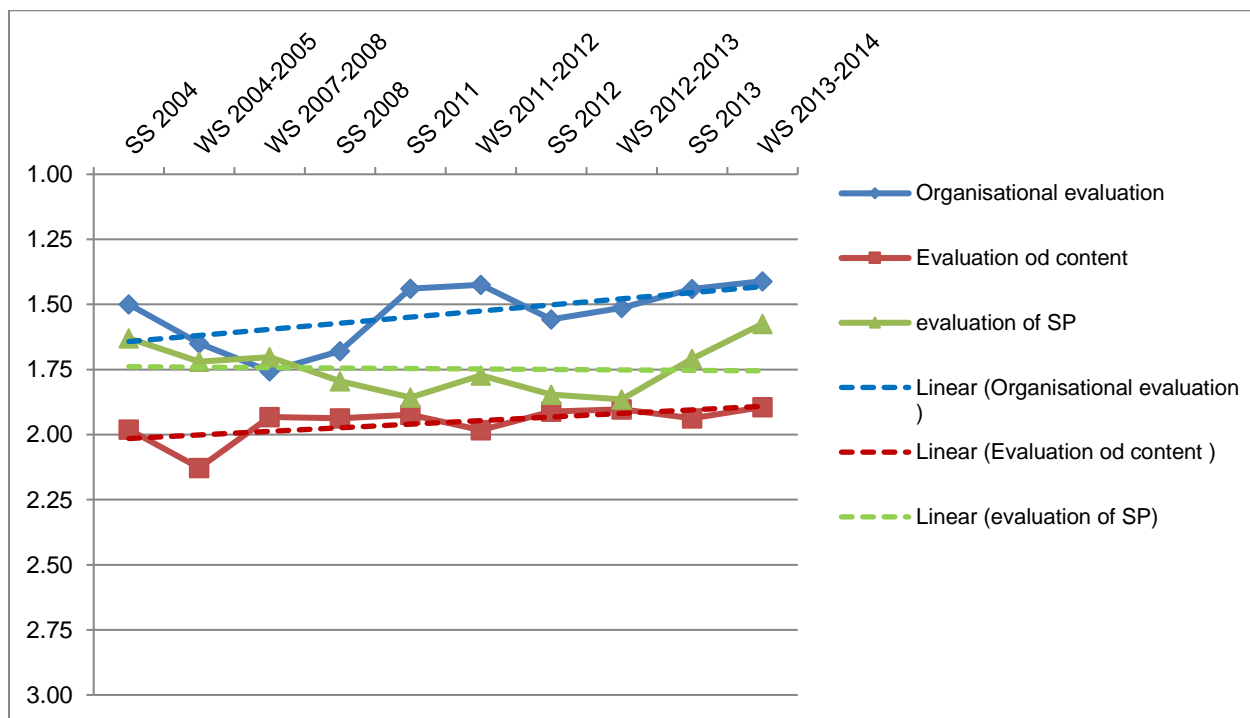**Figure 1:** Development of the average score.

Figure 2: Satisfaction with the exam format over time (1=appropriate/very good, 6=inappropriate/in-adequate).

Graf J, Smolka R, Holderried F, et al. (2018) Ten Years of Objective Structured Clinical Examination at the Medical Faculty of Tübingen, Germany: Item Analysis and Students' Satisfaction J Health Sci Educ 2: 133.

Semester	Score Mean	Score Median	Score SD (Min/ Max)	p-value (ANOVA)
2004	19.49	19.50	1.98 (16.9/22.5)	0.064
2004-2005	19.89	20.20	0.93 (18/21.1)	
2005	19.60	19.48	1.30 (17.7/22.3)	
2005-2006	19.67	19.70	0.53 (18.6/20.4)	
2006	19.31	19.44	0.52 (18.6/19.8)	
2006-2007	19.43	19.27	1.00 (18.4/21.4)	
2007	20.57	20.22	1.89 (19.5/21)	
2007-2008	20.05	19.84	0.56 (19.3/20.9)	
2008	20.66	20.67	0.88 (19.3/21.7)	
2008-2009	20.65	20.65	0.76 (19.8/21.9)	
2009	20.84	20.90	0.56 (19.8/21.5)	
2009-2010	20.53	20.69	0.81 (19/21.5)	
2010	19.60	20.22	2.16 (14.5/21.6)	
2010-2011	19.80	19.90	0.97 (18.3/21.2)	
2011	20.23	20.38	0.86 (18.9/21.7)	
2011-2012	20.13	20.00	0.60 (19.1/20.9)	
2012	20.37	20.38	0.86 (19.5/21.8)	
2012-2013	19.71	19.43	0.70 (18.8/20.8)	
2013	20.48	20.47	0.89 (19.2/21.8)	
2013-2014	20.63	20.70	0.45 (19.9/21.2)	

Table 3: Development and variability of the average score.

Development of item difficulty and item-total correlation

Table 4 shows the interdisciplinary item difficulties and the item-total correlation per semester, as well as the respective minimum and maximum values resulting from the subject-specific analyses per semester. The item difficulty was 0.80 on average across all semesters. The subject-specific scattering was also rather low across almost all semesters, but it was significantly higher than the semester-specific scattering. The item-total correlation reached a value greater than 0.3 across almost all semesters, but with a greater semester-specific scattering than in terms of the item difficulty. Most of the time, the subject-specific scattering per semester was also not negligible, with the highest range of item-total correlation with a value of 0.47 shown in the evaluation results of the winter semester 2004-2005 (min: 0.10, max: 0.57).

Development of intra-class correlation

Significant differences were also shown when comparing the intra-class correlation coefficients (Table 4): in 15 out of 20 semesters, this resulted in a value of ≥ 0.7 and thus in a high correlation.

Discussion

Main results

Neither the average score (research question 1) nor the dimensions of the students' evaluations (research question 2) showed statistically significant differences when comparing the semesters. This points to a small variability consisting of random variations and thus to a degree of competence acquisition that is equally high throughout the semesters and a

consistent examination quality from the perspective of the students. Furthermore, the examination seems to be suitable in a psychometric sense for the measurement of students' competences (research question 3): The item-total correlation, which indicates how well the differentiation between positive (=good students) and negative (=bad students) works, was almost consistently within an acceptable range (>0.3), because in principle, item-total correlations of >0.3 are recommended [15].

However, a significant variability was shown both in the subject-specific sense (within the semester) and in the semester comparison. Even with the item difficulty, almost all semesters showed high averages, the variability was significantly lower than with the item-total correlation. Finally, the high intra-class correlations found indicate a high reliability and thus the potential of the OSCE examination to be able to reliably measure students' competences (research question 4).

Semester	Item-total correlation: *Mean (Min; Max)*	Item difficulty: *Mean (Min; Max)*	Intraclass correlation coefficient [95%-CI]
2004	0.43 (0.30; 0.49)	0.78 (0.68; 0.90)	0.73 [0.66; 0.79]
2004-2005	0.34 (0.10; 0.57)	0.80 (0.72; 0.87)	0.74 [0.68; 0.80]
2005	0.45 (0.25; 0.55)	0.78 (0.71; 0.89)	0.76 [0.69; 0.81]
2005-2006	0.45 (0.37; 0.53)	0.79 (0.74; 0.82)	0.75 [0.68; 0.80]
2006	0.84 (0.71; 0.92)	0.77 (0.75; 0.79)	0.96 [0.95; 0.97]
2006-2007	0.50 (0.37; 0.59)	0.78 (0.74; 0.86)	0.79 [0.73; 0.83]
2007	0.52 (0.41; 0.71)	0.82 (0.79; 0.87)	0.80 [0.75; 0.84]
2007-2008	0.25 (0.08; 0.35)	0.80 (0.77; 0.84)	0.53 [0.38; 0.65]
2008	0.63 (0.47; 0.73)	0.83 (0.77; 0.87)	0.85 [0.82; 0.88]
2008-2009	0.54 (0.33; 0.72)	0.83 (0.79; 0.88)	0.80 [0.72; 0.86]
2009	0.43 (0.23; 0.55)	0.83 (0.79; 0.86)	0.71 [0.64; 0.78]
2009-2010	0.38 (0.25; 0.56)	0.82 (0.76; 0.86)	0.67 [0.58; 0.74]
2010	0.44 (0.29; 0.68)	0.78 (0.58; 0.86)	0.75 [0.69; 0.80]
2010-2011	0.49 (0.36; 0.69)	0.79 (0.73; 0.85)	0.79 [0.74; 0.83]
2011	0.60 (0.36; 0.76)	0.81 (0.75; 0.87)	0.86 [0.83; 0.89]
2011-2012	0.30 (0.07; 0.49)	0.81 (0.77; 0.84)	0.59 [0.48; 0.68]
2012	0.46 (0.30; 0.60)	0.81 (0.76; 0.86)	0.78 [0.72; 0.83]
2012-2013	0.87 (0.74; 0.92)	0.79 (0.76; 0.83)	0.97 [0.96; 0.97]
2013	0.35 (0.17; 0.55)	0.82 (0.77; 0.87)	0.66 [0.57; 0.73]
2013-2014	0.49 (0.26; 0.68)	0.83 (0.80; 0.85)	0.78 [0.72; 0.82]

Table 4: Development and variability of average scores of item and reliability analysis.

Classification of the results

In principle, the OSCE examination format is of high relevance for examining the practical and communicative skills of medical students, since the communicative competence of physicians in the health care system is becoming increasingly important. However, in order to further promote the communicative and clinical-practical skills, it is important to analyze, reflect on and possibly modify the outcomes of the teaching format as well as the quality of the outcomes since communicative and practical competences are essential for a successful doctor-patient relationship and are thus crucial for patient safety and efficiency of medical treatment [1-8]. In the light of the fact that OSCE examinations require a great deal of human and financial resources, it is thus important to know how reliably the competences which need to be evaluated can be measured with the examination. OSCE examinations are considered extremely valid and reliable compared to other examination formats [16,17], and a high degree of reliability could also be shown in this publication. In terms of item-total correlation and item difficulty, the results thus matched those of other studies [18,19]. Although there are many studies that examined the validity, reliability, and methodological quality of the OSCE exam [16-19], we found no studies that also reflected average grades and the results of the evaluation.

Limitations

There were some limitations to the validity of the results. It remains unclear whether the OSCE examination is equally suitable for measuring communicative and practical competences, and to what extent differences in reliability between communicative and practical competences exist, since the departments only provided the overall scores without the separately assigned standardized communication scores of each student. Separate research is needed to better assess the suitability of the examination to measure communicative and/or practical skills. In particular, the objective of the present study was the presentation of item difficulty, item-total correlation, and reliability, which does not allow any conclusions to be drawn on the objectivity and validity of the examination, but would be necessary in order to fully assess the suitability of the examination for measuring students' competences.

Outlook

Due to the high importance of the examination, the excellent practicability, the good results and the high satisfaction rates, further development of the examination format is planned for the coming semesters. In the medium term, there are plans to also implement the examination format in other subjects. It would also be desirable to implement an OSCE examination in which higher-level competences are examined, such as those defined in the National Competence-Based Learning Catalog Medicine (NKLM) adopted in 2015 [20], for example in the context of an emergency room.

Conclusions

There were no statistically significant differences between the semesters in relation to aver-age scores and the results of the students' evaluations. The Tübinger OSCE seems to be suitable and reliably for the measurement of communicative and practical skills.

Author Contributions

Conceptualization: JG, AHW.

Data curation: JG.

Formal analysis: JG.

Methodology: JG, CN, AHW.

Project administration: AHW, SZ.

Visualization: JG.

Writing – original draft: JG, AHW.

Writing – review & editing: JG, RS, FH, AW, MLK, DM, EV CN, SZ, AHW.

All authors read and approved the final manuscript.

Conflict of Interest

The authors declare that they have no conflicts of interest in connection with this article.

Acknowledgments

References

1. Foronda C, MacWilliams B, McArthur E (2016) Interprofessional communication in healthcare: An in-tegrative review. Nurse Educ Pract 19: 36-40.

2. Levinson W, Lesser CS, Epstein RM (2010) Developing physician communication skills for patient-centered care. Health Aff (Millwood) 29(7): 1310-1318.

3. Jünger J, Köllner V, Lengerke von T, et al. Kompetenzbasier-ter Lernzielkatalog „Ärztliche Gesprächsführung. Z Psychosom Med Psychother 62(1): 5-19.

4. Fragstein von M, Silverman J, Cushing A, et al. (2008) UK consensus statement on the content of communication curricula in undergraduate medical education. Med Educ 42(11): 1100-1107.

5. Harden RMcG, Stevenson M, Wilson Downie W, et al. (1975) Assessment of clinical competence using objective structured examination. Brit Med J 1(5955): 447-451.

6. Gormley G (2011) Summative OSCEs in undergraduate medical education. Ulster Med J 80(3): 127-312.

7. Khan KZ, Ramachandran S, Gaunt K, et al. (2013) The Objective structured clinical examination (OSCE): AMEE

Guide No. 81. Part I: An historical and theoretical perspective. Med Teach 35(9): e1437-1446.

8. Daniels VJ, Pugh D (2017) Twelve tips for developing an OSCE that measures what you want. Med Teach 25: 1-6.

9. Zayyan M (2011) Objective structured clinical examination: The assessment of choice. Oman Med J 26(4): 219-222.

10. Mohr D, Popp R, Zipfel S, Herrmann-Werner A (2015) Einführung studentisch organisierter OS-CE-Übungstage an der Medizinischen Fakultät Tübingen. German Medical Science, GMS Publishing House, DocP11-153.

11. Ryan CA, Walshe N, Gaffney R, et al. (2010) Using standardized patients to assess communication skills in medical and nursing students. BMC Med Educ 17: 24.

12. Mohr D, Graf J, Brass K, et al. (2014) From paper & pencil to tablet: Implementierung einer elektronisch unterstützten OSCE-Prüfung an der Medizinischen Fakultät Tübingen. German Medical Science, GMS Publishing House, DocP175.

13. Graf J, Smolka R, Simoes E, et al. (2017) Communication skills of medical students during the OSCE: Gender-specific differences in a longitudinal trend-study. BMC Med Educ 17: 75.

14. Van der Kamp LTP, Bijleveld CCJH (1998) Methodological issues in longitudinal research. In: Bijleveld CCJH, van der Kamp LTP et al. (Ed.). Longitudinal data analyses. Designs, models and methods. Cromwell Press Ltd. Wiltshire, pp: 1-45 (especially pp. 22f.).

15. Cömert M, Zill JM, Christalle E, et al. (2016) Assessing communication skills of medical students in Objective Structured Clinical Examinations (OSCE) - A systematic review of rating scales. PLoS ONE 11(3): e0152717.

16. Setyonugroho W, Kennedy KM, Kropmans TJ (2015) Reliability and validity of OSCE checklists used to assess the communication skills of undergraduate medical students: A systematic review. Patient Educ Couns 98(12): 1482-1491.

17. Brannick MT, Erol-Korkmaz HT, Prewett M (2011) A systematic review of the reliability of objective structured clinical examination scores. Med Educ 45(12): 1181-1189.

18. Iblher P, Zupanic M, Härtel C, et al. (2011) Im Notfall bereit? OSCE in der studentischen Notfallausbildung. Anästh Intensivmed 52: 812-822

19. Auewarakul C, Downing SM, Praditsuwan R, et al. (2005) Item analysis to improve reliability for an internal medicine undergraduate OSCE. Adv Health Sci Educ Theory Pract 10(2): 105-113.

20. MFT Medizinischer Fakultätentag der Bundesrepublik Deutschland e. V. (2015) Nationaler Kompetenz-basierter Lernzielkatalog Medizin (NKLM). Kiel, Germany.

*Corresponding author: Joachim Graf, M.A., M.Sc., Dean's Office for Students' Affairs, Medical Faculty Tübingen, Geissweg 5/1, 72076 Tübingen, Germany, Tel: +49 7071 29-87118; Email: joachim.graf@med.uni-tuebingen.de

Citation: Graf J, Smolka R, Holderried F, Wosnik A, Lammerding-Köppel M, Mohr D, Vlad E, Nikendei C, Zipfel S, Herrmann-Werner A (2018) Ten Years of Objective Structured Clinical Examination at the Medical Faculty of Tübingen, Germany: Item Analysis and Students' Satisfaction. *J Health Sci Educ* 2(2): 133.