Reliability of Assessing Oral Presentations by the University Professionals

K.T.A.C. Kasthuriratna[1], R.M.M.P. Rathnayaka[2], V.N. Kodithuwakkuge[3], C.K. Beneragama[3]

[1]*189/1, 2nd Lane, Werellawatta, Yakkala, Sri Lanka*
[2]*Department of Agricultural Extension, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka*
[3]*Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka*

Abstract: **The fairness and precision of evaluation of Oral Presentations of students by university professionals have become a debatable subject. The effectiveness of the evaluation of PowerPoint presentations was seriously questioned by the students due to its unreliability of scoring procedure. Therefore, it's important to establish a planned evaluation system for oral presentation based on PowerPoint, to guarantee the fairness for every student. To minimize the potential biases, most of the universities presently adopt Objective Structured Evaluation systems to enhance the transparency and the reliability of the assessments. In view of that, the present study analysed the biasness of assessing the oral presentations of a student cohort of a university. For this study, mean score of each student received from each examiner was taken. Single-factor ANOVA tests were conducted to analyse variances to compare three examiner groups; professors, senior lecturers and probationary lecturers. Tukey simultaneous test was conducted to identify mean differences in each comparison. Strong evidence of differences among the three examiner groups was present. Within the most senior level of professionals, a greater degree of variance was also identified. In addition, there is a variance within the senior lecturer group while the probationary lecturer group did not reflect any significant variance. In conclusion, our findings demonstrated statistically significant differences in the marks awarded for the PowerPoint presentations of undergraduates as influenced by examiners' experience and seniority both in between examiners and within the same level of examiners.**

Keywords: **Oral presentations, Assessing reliability, Examiners' effect, University professionals, objective structure**

I. INTRODUCTION

The fairness and precision of evaluation of oral presentations of students by the university professionals has become a debatable subject over the last few decades. The effectiveness of the evaluation of PowerPoint presentations has been occasionally questioned by the students due to its unreliability of the scoring procedure. These have created a lot of uneasiness and unrest among students resulting in adverse moral effects to their lives [11]. The lack of transparency and inferior quality assessments cause serious consequences for students in guaranteeing the assessment quality. Therefore, identification of factors affecting the reliability of these assessments and minimizing such, may improve the quality of assessments enhancing the student's faith on assessments.

PowerPoint presentations are an integral part of the evaluation system in most of the universities [11]. However,

due to various reasons, the reliability of assessing has become questionable. Therefore, it's important to establish a well-planned evaluation system for oral presentation based on PowerPoint, targeting the fairness for each student. The factors affecting the fairness or consistency of assessment could be identified in four major areas as (i) Biasness, (ii) Untrained professionals in assessing presentations, (iii) Lack of commitment and (iv) Subjectivity of the evaluation factors [27].

Due to those different factors affecting evaluation, most of the universities presently adopt objective structured evaluation system to enhance the transparency and the reliability of the assessments [27]. Despite the increasing introduction of objective structures assessment system to assess student competencies, there are higher variability. This evaluation system has shown variability on subject, checklist, and overall outcome as well as a higher examiner-dependent effect. Many previous studies have reported examiner-based variability, "hawk-dove-effect" and the gender-based biasness [8]. Oral presentations can be assessed by oneself, peers and by trained professionals. Those three types of assessment have been studied in comparison to each other in different demographic areas, different languages and academic groups [12,13]. Assessment of oral presentations by professionals in the university system should be studied to see if more accurate and transparent grades are given to the students. Oral presentation skill in professional contexts has been identified as one of the most important 21st Century skills in the present-day context. It is of utmost importance that the timely feedback and the grades should be given to the students targeting their improvement of the skills.

The present study was carried out to see the variations in evaluations among university professionals but not go deep in to find out the reasons for such variances. Research questions were targeted to identify the variance of scoring among examiners and to evaluate the influence of their experience on scoring.

II. BACKGROUND OF THE STUDY

A. Oral Presentations Based on PowerPoint

Oral presentations that are often conducted with the aid of PowerPoint in academic institutes can be categorized under the speaking assessments [12]. Oral presentation skills are

considered as an important 21st Century skills in the professional world [7]. Apart from the writing reports or paper-based assessments, being able to present orally is important for communicating knowledge to other groups. These types of skills are very difficult to analyse and evaluate objectively, because it is hard to come up with clear guidelines to what exactly constitute a good presentation [22].

With the popularization of computer technology within university academic systems, there had been a paradigm shift in digital presentations. Microsoft PowerPoint has emerged as a supportive software for oral presentations. Attractive features of PowerPoint (diagrams, pictures, feasible fonts, animations etc.) and user-friendly platform make it popular among the university students [24].

B. Assessing the Oral Presentations

Evaluators should be well-aware about how students perceive the task of making presentations [22]. Because, presentations play an important role in student success in universities. Evaluators should plan a marking scheme which is based on minimal standards and quality, before grading the students' oral presentations. Moreover, evaluators should have enough experience to judge the quality of presentations as objective as possible [12].

C. Objective Structured Assessment System

Objective structured assessment systems have been well-established targeting reliability and the validity of the assessment systems. This system was first applied for surgical skills in the field of medicine and later developed as a system to apply to other skills as well. There are two strategies called an operation specific checklist and a global rating scale. The global rating scale is the method which can be used to evaluate other skills. It is a common method of evaluation and consists seven evaluation items scored on a five-point scale [25]. This system has become popular in the university systems in evaluating students' skills mainly due to its reliability and transparency [27].

D. Influence of Examiner on the Assessment System

According to [4], there are many evidences for the reliability of objective structured assessment system, especially for different locations/sites and different languages. However, few other studies have identified that, there are variable conditions in this assessment system [16]. Some authors have reported that, content of the examination, student outcome and detailed check lists have a considerable impact on students' grades [8], in addition to the examiner dependency, which can be attributed to examiner's training on the subject, behavior, and time availability for the assessment process.

In some cases, the evaluator can be bias towards a student. Biasness of teachers can be discussed using "halo effect" as well. This effect is based on the prior knowledge of the person (neatness, grades of previous assessments, intelligence, and support in the class). As an example, if some students have earned higher grades in previous assessments and appeared intelligent, with good support in class, they may create a positive halo effect. Moreover, if some students with bad grades for previous work and not pleasant in class may create a negative halo effect [10, 18]. However, not every study reported the halo effect on grading [2]. The bias in written exams following an oral exam by academics has been reported previously [17].

Moreover, there is a traditional idea about the examiners' biasness on gender of the students. However, majority of the studies have reported that, there is no association between gender and students grades [9, 20, 28]. Some studies have revealed the gender bias nature of teachers' evaluation on students in the Europe [21]. In another study, the female students were given substantially less marks than comparable males and this disparity grows with increased proficiency of medical students [1]. Another study has shown that, the female examiners rated higher marks generally for all the students, indicating the gender-related biasness of the evaluator [26].

E. The Study

All of the preceding investigations and conclusions essentially served as the impetus for this study. The present study attempted to find out the difference of scoring of the examiners in students' oral presentations and to examine whether the experience of the examiners has an impact on the marks of the students' presentations. Final year undergraduate students at a public university (following a specialization module) were selected as the study population. The student cohort has conducted a research project for 5 months and they were supposed to present their research findings in the form of an oral presentation with the aid of PowerPoint. The final presentation is one of the key components in assessing the research project in addition to the written thesis since it's an important skill that students should have [13], [29]. Each student was supposed to make a presentation for 15 minutes, followed by a discussion for 5 minutes. The same group of examiners who are the subject experts in the relevant field evaluated the students' presentations.

III. METHODOLOGY

Data were taken from the final year research presentations done by 45 students from the same specialized area. The presentations were evaluated by an academic panel consisting 18 qualified examiners as 7 Professors (*Prof*), 9 Senior Lecturers (*SL*) and 2 Probationary lecturers (*Prob*). A pre-determined marking scheme was employed to examine students' presentations. Marks were given out of 200. The mean value of the marks given by all the examiners was considered as the final mark for students' performances in oral presentations. The presentations lasted for two consecutive days in a physical environment and all the examiners were requested to evaluate each student

individually. Examiners were not allowed to evaluate the students who were supervised by them.

Within each item, assigned marks are expected to be given in accordance with the item, without additional breakdown. Evaluation sheets with criteria were given to each examiner and they were collected at the end of each day. The following pre-determined marking scheme was used to produce the evaluation sheets (Table 1).

Table 1: Evaluation Criteria of the Marking Scheme

Criteria/ item	Criteria/ item in brief	Allocated marks
Content	Originality, technical relevance	50
Quality of presentation	Logical sequence, organization and preparation of the presentation, delivery, oral communication, confidence, interest, eye contact, thorough knowledge of the topic	50
Quality of visual aids	Visual aid design, layout and usefulness, effective use of the visual aids	50
Discussion	Ability to stimulate further thinking on the subject, ability to answer questions	25
Time management	Managing time during the presentation	25
Total		200

Mean score of each student received from each examiner was taken. Single factor ANOVA tests were conducted to analyse variances to compare three examiner groups, compare 7 professors, 9 senior lecturers and 2 probationary lecturers. Tukey simultaneous tests were conducted to identify mean differences in each mentioned comparison. For each comparison, hypotheses were developed as follows;

1. Hypothesises to compare three examiner groups,

H_1: There are no differences among three examiner groups (*prof*, *SL*, and *Prob*)

H_2: At least two mean scores differ among three examiner groups

2. Hypothesises to compare professor group,

H_3: There are no differences among professors in the evaluation process

H_4: At least two mean scores differ among professors in the evaluation process

3. Hypothesises to compare Senior Lecturer group,

H_5: There are no differences among Senior Lecturers in the evaluation process

H_6: At least two mean scores differ among Senior Lecturers in the evaluation process

4. Hypothesises to compare Probationary lecturer group,

H_7: There are no differences among Probationary lecturers in evaluation process

H_8: At least two mean scores differ among Probationary lecturers in evaluation process

Minitab statistical package and Excel (2013) were used to analyse data.

IV. RESULTS & DISCUSSION

Although all examiners were expected to be present for all 45 students' presentations on both days, some did not (Prof 3, Prof 5, SL 2, SL 3, SL 8, and SL 9) (Table 2), indicating the lack of commitment of the examiners on the evaluation process. This may have an impact on the mean score of the students which is completely independent from their performances [24].

As shown in the Table 2, Professor 05 contributed to evaluate 28 students' presentations out of 45. Senior lecturer 2 and 3 evaluated only 18 and 14 presentations respectively which is less than half of the total number of presentations. To minimize the examiner effect on presentations, the marks of these should be removed from the final assessment. All the students received final marks within a range of 127.2 to 164.8 out of 200. As a percentage, the lowest marks received by a student was 63.6% whereas the highest mark was 82.4%.

As shown in the Figure 1, some examiners (e.g. SL 4) have given marks making them outliers in some students' final marks. This may be due to their subjective perception [6] and biases on students as many of students are familiar to them throughout the specialization course program. Moreover, Senior Lecturer 4 appeared to be an outlier in most of the students' cases. The mean score of that particular examiner was 121.9 which is lower than the lowest mark a student received overall. Senior lecturer 4 seems to be a tough examiner as well as highly inconsistent in scoring compared to other examiners. As a result of their subjective judgments, the examiner effect appears to be strong from these types of examiners. This kind of outliers/extreme scores should preferably be eliminated from the final assessment in order to reduce the examiner biaseness.

The findings of the one-way ANOVA tests used to evaluate four set hypotheses are given in the Table 3.

Table 3: Analysis of variance in 4 different comparisons

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Comparison on three examiner groups					
GroupM	2	8939	4469.7	16.18	0.000
Error	665	183736	276.3		
Total	667	192675			
Comparison on Professors					
ExaminerM	6	14125	2354.1	19.01	0.000
Error	262	32448	123.8		
Total	268	46573			

Comparison on Senior Lecturers					
ExaminerM	8	59796.3	7474.5	37.42	1.1102e-16
Error	300	59916.5	199.7		
Total	308	119712.8			
Comparison on Probationary Lecturers					
ExaminerM	1	126.6	129.6	0.65	0.419

Error	88	17320.8	196.8		
Total	89	17450.4			

(Examiner M=Mean score given by the examiner, GroupM=Mean score given by the examiner group)

Table 2: Descriptive statistics of the marks given by each examiner to student

	Prof 1	Prof 2	Prof 3	Prof 4	Prof 5	Prof 6	Prof 7	SL 1	SL 2	SL 3	SL 4	SL 5	SL 6	SL 7	SL 8	SL 9	Prob 1	Prob 2
Mean	148.4	154.1	159.7	139.7	160.3	158.8	147.0	158.4	175.3	155.9	121.8	157.6	137.8	155.5	151.4	146.2	158.6	161.0
SD	10.01	10.50	9.96	15.92	7.55	10.34	10.16	14.00	9.43	12.00	20.28	9.00	13.23	14.55	14.63	12.26	15.19	12.76
NSE	39	42	33	44	28	40	43	40	18	14	44	40	45	40	35	33	45	45
SE	1.60	1.62	1.73	2.40	1.43	1.64	1.55	2.21	2.22	3.21	3.06	1.42	1.97	2.30	2.47	2.13	2.27	1.90
Min	125	129	135	105	142	135	125	120	157	130	40	143	102	110	110	120	127	129
Max	163	181	183	165	178	175	169	191	190	169	165	180	162	178	178	170	195	181
Range	38	52	48	60	36	40	44	71	33	39	125	37	60	68	68	50	68	52

(Mean=Mean of examiner, SD=Standard Deviation, NSE=Number of students evaluated, SE=Standard Error, Min= Minimum score given by the evaluator, Max= Maximum score given by the evaluator, Range= Difference between Min and Mix score)
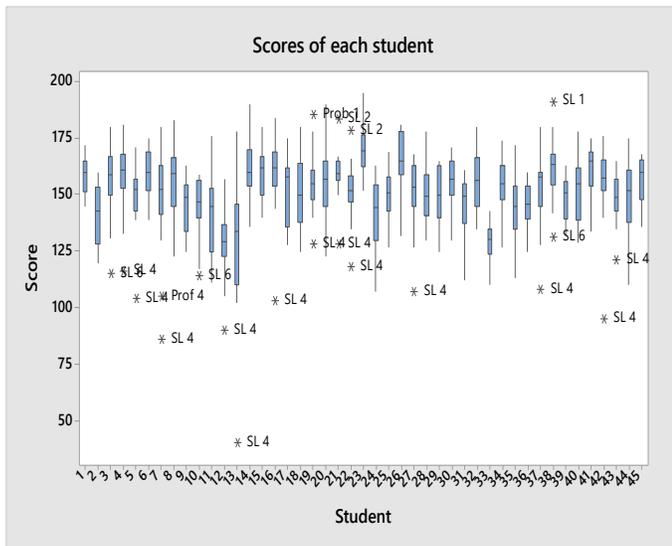

Figure 1. Comparison of students' final marks with mean marks given by each examine

According to the results from three examiner group comparison (Table 03), H1 can be rejected and H2 is acceptable. There is strong evidence of differences among the three examiner groups (p=0.000). This could be due to their differences in seniority and experiences [5]. A Tukey simultaneous test was conducted to identify the significant differences between each group. Probationary lecturers' scores are at least 6.60 and at most 15.92 points higher than that of senior lecturers, on average difference of 11.26 which is considerably high. Professors' results were at least 3.20 and at most 12.68 points higher than those of probationary

lecturers, with an average difference of 7.94 points. This might be due to the lack of experience of probationary lecturers than other two groups and it directs them to give scores to students without being 'extreme'. However, professors' scores were at least 6.57 and at most 0.08 points higher than that of Senior Lecturers, which was unexpected, but could be due their seniority.

Table 4: Tukey simultaneous test for differences in means in evaluators

Difference of Type Levels	Difference of Means	SE of Difference	Simultaneous 95% CI	T-Value	Adjusted P-Value
2 - 1	-3.32	1.39	(-6.57, -0.08)	-2.40	0.044
3 - 1	7.94	2.02	(3.20, 12.68)	3.92	0.000
3 - 2	11.26	1.99	(6.60, 15.92)	5.66	0.000

Individual confidence level = 98.04%

1 - Professors, 2 – Senior Lecturers, 3 – Probationary Lecturers)

As per the Table 4, the results from professor group comparison have shown differences among the professors (p=0.000) which rejects the null hypothesis (H3). To clarify further, Tukey simultaneous test was conducted for differences in means. Evaluation of professor 1 is significantly differ from professor 3 (p=0.000), 4 (p=0.008), 5 (p=0.000), and 6 (p=0.001). Evaluation of professor 3 is significantly different from professor 7 (p=0.000), evaluation of professor 4 is significantly differ from professor 2 (p=0.000), 3 (p=0.000), 5 (p=0.000), 6 (p=0.000), and 7 (p=0.037), evaluation of professor 7 is significantly differ from professor 5 (p=0.000) and 6 (p=0.000). There was a greater degree of variance even among the most senior level

of professionals. Each professor is specialized in different subject areas and their seniority and experiences may vary among themselves as well. This could be the reason for the observed variances within the professors with respect to the scores given.

H5 (null) hypothesis could also be rejected with the findings from the comparison among Senior lecturers. It indicated a significant variance than other two groups (p=0.000). Tukey test supported these findings with many differences in mean comparisons, except in a few insignificant pairwise comparisons between some senior lecturers. Similar to Professors, most senior lecturers reflected a higher variance which possibly creates a greater examiner effect on the oral presentation evaluation process. Since the number of Professors and Senior Lecturers was high in the evaluation panel, the impact on the overall final marks will be severe. However, when comparing Probationary lectures, the variance is insignificant (p=0.419) whereas null hypothesis (H7) can be accepted.

All the comparisons indicated that the differences in variances between three examiner groups and within each examiner group except one, are not favourable for a neutral and unbiased evaluation process. By structuring the valuation process further with clearly defined and specific criteria [3], the objective evaluation of students' oral presentations can be warranted while eliminating subjective effects of examiners [23]. In terms of eliminating the effect of examiners' seniority and experiences in the assessment, providing training on how to outlines the marking criteria, teaching curriculum, and expected level of student performance in communication and examination skills as a strategy would help [15]. This kind of strategies will provide a framework for all the levels of examiners to do the evaluation process to examine students' true performances.

V. CONCLUSION & RECOMMENDATIONS

Despite the fact that all examiners are required to participate in the assessment process, certain evaluators' lack of consistent dedication to the process has resulted in poor overall grades for students. In the evaluation process, at least two mean scores differ among three examiner groups, at least two mean scores differ among professors, and at least two mean scores differ among Senior Lecturers. In the evaluation process, however, there were no variations in assessment amongst Probationary lecturers.

Our findings demonstrated small but statistically significant differences in the marks awarded for the PowerPoint presentations of undergraduate students based on examiner's seniority both in between examiners and within same level of examiners. Moreover, the examiners' lack of commitment to the evaluation process and subjective biases and extremes could result in wrong assessments. We believe that our data highlight the need for specific strategies to encourage more objective marking by examiners. We recommend training that outlines the marking criteria,

teaching curriculum, and expected level of student performance in communication and examination skills as a strategy to reduce bias in power point presentation assessment. This would allow examiners to mark students in a way that reflects their true performance, irrespective of examiner seniority, biasness or experience with assessment

This variability may introduce errors into ratings, which are independent of the student's performance. Through training, examiners should be made aware of potential bias, for example by implementation of methods like role-playing. This might have positive influence on examiner bias and should be further investigated in order to get fair results during exams.

REFERENCES

[1] Axelson, R. D. *et al.* (2010) 'Assessing implicit gender bias in medical student performance evaluations', *Evaluation and the Health Professions*, 33(3), pp. 365–385. doi: 10.1177/0163278710375097.

[2] Batten, J. *et al.* (2013) 'The influence of reputation information on the assessment of undergraduate student work', *Assessment & Evaluation in Higher Education*. Taylor & Francis, 38(4), pp. 417–435.

[3] Bone, A. (1999). *Ensuring successful assessment* (p. 32). National Centre for Legal Education.

[4] Brailovsky, C. A., Grand'Maison, P. and Lescop, J. (1992) 'A large-scale multicenter objective structured clinical examination for licensure', *Academic Medicine*. LWW, 67(10), pp. S37-9.

[5] Brown, G. A., Bull, J., & Pendlebury, M. (2013). *Assessing student learning in higher education*. Routledge.

[6] Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of political Economy*, 97(5), 1232-1254.

[7] Campbell, K. S. *et al.* (2001) 'Peer versus self assessment of oral business presentation performance', *Business Communication Quarterly*. Sage Publications Sage CA: Thousand Oaks, CA, 64(3), pp. 23–40.

[8] Chesser, A. *et al.* (2009) 'Sources of variation in performance on a shared OSCE station across four UK medical schools', *Medical Education*, 43(6), pp. 526–532. doi: 10.1111/j.1365-2923.2009.03370.x.

[9] Colliver, J. A. *et al.* (1993) 'Effects of examinee gender, standardized-patient gender, and their interaction on standardized patients' ratings of examinees' interpersonal and communication skills.', *Academic Medicine*. Lippincott Williams & Wilkins.

[10] Cooper, W. H. (1981) 'Ubiquitous halo.', *Psychological bulletin*. American Psychological Association, 90(2), p. 218.

[11] De Grez, L., Valcke, M. and Roozen, I. (2009) 'The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education', *Computers and Education*. Elsevier Ltd, 53(1), pp. 112–120. doi: 10.1016/j.compedu.2009.01.005.

[12] Dobson, S. (2006) 'The assessment of student powerpoint presentations-attempting the impossible?', *Assessment and Evaluation in Higher Education*, 31(1), pp. 109–119. doi: 10.1080/02602930500262403.

[13] Dorée, S., Jardine, R. and Linton, T. (2007) 'Let's talk about student presentations', *Primus*, 17(4), pp. 338–352. doi: 10.1080/10511970601131589.

[14] Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education*, 31(2), 115-128.

[15] Gibbs, G. (1992). Developing teaching: teaching more students: No. 4 Assessing more students. Published by The Polytechnics and Colleges Funding Council.

[16] Kift, S. (2002). Harnessing assessment and feedback to assure quality outcomes for graduate capability development: A legal education case study. In *AARE 2002 Conference Papers* (pp. 1-24). AARE Inc.

[17] Mäkinen, M. *et al.* (2010) 'Assessment of CPR-D skills of nursing students in two institutions: reality versus recommendations in the guidelines', *European Journal of Emergency Medicine*. LWW, 17(4), pp. 237–239.

[18] Malouff, J. M. *et al.* (2014) 'Preventing halo bias in grading the work of university students', *Cogent Psychology*. Cogent, 1(1). doi: 10.1080/23311908.2014.988937.

[19] Malouff, J. M., Emmerton, A. J. and Schutte, N. S. (2013) 'The Risk of a Halo Bias as a Reason to Keep Students Anonymous During Grading', *Teaching of Psychology*, 40(3), pp. 233–237. doi: 10.1177/0098628313487425.

[20] McManus, I. C., Elder, A. T. and Dacre, J. (2013) 'Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP (UK) PACES and nPACES examinations', *BMC medical education*. BioMed Central, 13(1), pp. 1–11.

[21] McManus, I. C., Thompson, M. and Mollon, J. (2006) 'Assessment of examiner leniency and stringency ('hawk-dove effect') in the MRCP(UK) clinical examination (PACES) using multi-facet Rasch modelling', *BMC Medical Education*, 6, pp. 1–22. doi: 10.1186/1472-6920-6-42.

[22] Mengel, F., Sauermann, J. and Zölitz, U. (2019) 'Gender bias in teaching evaluations', *Journal of the European Economic Association*. Oxford University Press, 17(2), pp. 535–566.

[23] Miles, R. (2014) 'The learner's perspective on assessing and evaluating their oral presentations', *Proceedings of CLaSIC 2014*, pp. 337–352.

[24] Morgan, C., Dunn, L., Parry, S., & O'Reilly, M. (2004). *The student assessment handbook: New directions in traditional and online assessment*. Southern Cross University.

[25] Niamtu 3rd, J. (2001). The power of PowerPoint. *Plastic and reconstructive surgery*, *108*(2), 466-484.

[26] Niitsu, H. *et al.* (2013) 'Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room', *Surgery Today*, 43(3), pp. 271–275. doi: 10.1007/s00595-012-0313-7.

[27] Schleicher, I. *et al.* (2017) 'Examiner effect on the objective structured clinical exam - A study at five medical schools', *BMC Medical Education*. BMC Medical Education, 17(1), pp. 1–7. doi: 10.1186/s12909-017-0908-1.

[28] Turner, J. L. and Dankoski, M. E. (2008) 'Objective structured clinical exams: a critical review', *Fam Med*, 40(8), pp. 574–578.

[29] Wiskin, C. M. D., Allan, T. F. and Skelton, J. R. (2004) 'Gender as a variable in the assessment of final year degree-level communication skills', *Medical education*. Wiley Online Library, 38(2), pp. 129–137.

[30] Živković, S. (2014). The importance of oral presentations for university students. *Mediterranean Journal of Social Sciences*, *5*(19), 468-468.