Impact of Assessment Formats on Student Performance in Oral Histology: A Comparative Analysis of Multiple-Choice (MCQ) and Very Short-Answer (VSA) Questions

Zavala Walther David [1], Garas Silvina[2], Bermejo Adriana[3], Matilla Victor[4], Stoehr Rodrigo[5], Bricco Oscar [6].

[1]Profesor Titular de la Cátedra de Histología. Facultad de Odontología de la UNCUYO, Argentina.

[2,3,4,5,6] Cátedra de Histología. Facultad de Odontología de la UNCUYO, Argentina

DOI: https://doi.org/10.47772/IJRISS.2026.10100449

Received: 07 January 2026; Accepted: 14 January 2026; Published: 12 February 2026

ABSTRACT

This study aimed to compare the academic performance and discriminatory power of questions on oral mucosa and dental tissue development Multiple-Choice Questions (MCQs; 2024 cohort) versus Very Short-answer Questions (VSAs; 2025 cohort) formats among first-year dentistry students. Ten identical questions from two cohorts (n =74 each) were analyzed for difficulty (p-index) using chi-square and t-test statistics. Results revealed higher average performance in MCQs (85.08%) than VSAs (79.48%; p=0.292), with significant differences in 3 questions (p<0.05), mostly easy ones (p value >061). We conclude that both formats are suitable for assessing foundational knowledge; however, the MCQ format provides greater reliability, while the VSA format reduces guessing bias

Keywords: Multiple-Choice Questions (MCQ), Very short-answer Questions (VSA), Dental education, academic assessment, Bloom´s taxonomy

INTRODUCTION

Assessment constitutes a fundamental component of the educational process, particularly in health professional training, as it enables the demonstration of learning objective achievement and, ultimately, ensures clinical competence (1). In medical education, evaluation not only measures knowledge but also influences students' study methods and learning styles, known as the washback effect (2). Students tend to prioritize content they know will be assessed, and the examination format partly determines how and what they learn (3).

Multiple Choice Questions (MCQs) offer high reliability and straightforward automated grading but provide orienting cues and promote recognition-based learning (RBL). Very short-answer questions (VSAs), which are open-ended items requiring brief responses, may help overcome these limitations (4). Although the use of VSAs in medical assessments is increasing, nearly all research on their reliability and validity in medical education has been conducted by a single experienced research group (5).

Generally, MCQs are preferred for assessing basic knowledge levels, whereas VSAs are often preferred for knowledge requiring greater clinical application. The primary objective of this study was to compare the academic performance and discriminatory capacity of first-year dentistry students regarding basic knowledge of two topics—oral mucosa and dental tissue formation—between two consecutive cohorts who answered identical questions in MCQ format (2024 cohort) versus VSA format (2025 cohort), and to explore the relationship with the difficulty level of each question.

METHODOLOGY

Study Design and Population

A retrospective, comparative, quasi-experimental cohort study was designed. Ten questions exploring the topics of oral mucosa and the process of dental tissue formation were selected from 74 partial examinations

(n=74) administered to two consecutive cohorts (2024 and 2025) of first-year dentistry students. The sample included 100% of the examinations from the 2025 cohort and 90% from the 2024 cohort to balance the sample sizes (n=74 per group).

Procedure

The selected questions addressed the same topics and required the same answers; however, for the 2024 cohort, they were presented in Multiple Choice (MCQ) format, while for the 2025 cohort, they were formulated as Very Short Answer (VSA) questions. Both cohorts possessed a similar academic profile, had the same study time, and received instruction from the same faculty. Prior academic performance, determined by an evaluation conducted in the first semester, showed comparable results between both cohorts.

Data Analysis

The selected questions were grouped by difficulty level, and responses were compared between the 2024 (MCQ) and 2025 (VSA) cohorts. Difficulty was determined both qualitatively and quantitatively. Qualitatively, clarity of the prompt and the elimination of ambiguity were considered. Quantitatively, the difficulty index p (proportion of correct answers) was applied to each question after the exam. This index ranges from 0 to 1, where values close to 1 indicate easy questions and values close to 0 indicate difficult questions. The scale used was: 0.00–0.20 (very difficult), 0.21–0.40 (difficult), 0.41–0.60 (medium difficulty), 0.61–0.80 (easy), and 0.81–1.00 (very easy) (6).

Internal Validity

Evaluators were calibrated, and VSA questions were designed to have a unique correct answer to reduce examiner bias. To guarantee scoring reliability for the VSAs, inter-rater agreement was calculated on a random sample of 20% of the exams (n=15) using Cohen's Kappa coefficient, yielding a value of 0.87 (95% CI: 0.79–

0.95)

Statistical Analysis

Descriptive statistics were performed. To compare the proportions of correct answers between the two independent cohorts for each individual question, the Chi-square test was used. The Student's t-test was used to compare the means of the two groups. An N=74 was used for each cohort. The null hypothesis (H0) posited no difference in the proportion of correct answers between formats. The significance level was set at alpha=0.05.

RESULTS

Tables 1 and 2 present the data obtained for the 5 questions on "oral mucosa" and the 5 questions on "dental tissues," expressed as the percentage of correct and incorrect responses (n=74), in relation to the difficulty degree established for each question. Notably, the majority of questions corresponded to a low level of difficulty (Easy/Very Easy), with only one question per topic corresponding to medium difficulty.

Regarding the questions on **oral mucosa** (Table 1), in Question 1, students in the MCQ cohort achieved significantly higher success rates than the VSA cohort ($p < 0.05$). For the remaining questions (2, 3, 4, and 5), there were no significant differences between the MCQ and VSA exams ($p > 0.05$).

Regarding the questions on **dental development** (Table 2), statistically significant differences ($p < 0.05$) between MCQ and VSA were found in Questions 1 and 3. Questions 2, 4, and 5 did not register significant differences ($p > 0.05$).

The comparison of the arithmetic means of the two groups (MCQ = 85.08%, SD 13.06; VSA = 79.48%, SD 9.77%) across all 10 questions using the Student's t-test yielded a p-value of 0.292, indicating no statistically significant difference in overall performance.

See Tables 3 and 4 for the detailed Chi-square statistical analysis.

Tables

Table 1Oral Mucosa: Correct and Incorrect Responses by Difficulty Level and Test Type			
Question	**Difficulty (p)**	**Correct Responses (N=74)**	**Incorrect Responses (N=74)**
1 **MCQ**	0.96	71 (96%)	3 (4%)
1 **VSA**	0.74	55 (74.3 %)	19 (25.7%)
2 **MCQ**	0.86	64 (86.4%)	10 (13.6%)
2 **VSA**	0.89	66 (89.2%)	8 (10.8%)
3 **MCQ**	0.91	68 (92%)	6 (8%)
3 **VSA**	0.82	61 (82.4 %)	13 (17.6%)
4 **MCQ**	0.91	68 (92%)	6 (8%)
4 **VSA**	0.87	65 (87.84 %)	9 (12.16%)
5 **MCQ**	0.59	44 (60%)	30 (40%)
5 **VSA**	0.60	45 (60.81 %)	29 (39.19%)

Table 2 Dental Development: Correct and Incorrect Responses by Difficulty Level and Test Type			
Question	**Difficulty (p)**	**Correct Responses (N=74)**	**Incorrect Responses (N=74)**
1 **MCQ**	0.90	67 (90.5%)	7 (9.5%)
1 **VSA**	0.74	55 (74.3 %)	19 (25.7%)
2 **MCQ**	0.87	65 (87.84%)	9 (12.16%)
2 **VSA**	0.81	60 (81.08%)	14 (18.91%)
3 **MCQ**	0.94	70 (94.6%)	4 (5.4%)
3 **VSA**	0.87	65 (87.84 %)	9 (12.16%)
4 **MCQ**	0.91	68 (91.9%)	6 (8.1%)
4 **VSA**	0.87	65 (87.8 %)	9 (12.2%)
5 **MCQ**	0.59	44 (59.45%)	30 (40.55%)
5 **VSA**	0.69	51 (68.9 %)	23 (31.1%)

Table 3. Chi-Square Test Comparing Results for the Five Oral Mucosa Questions (MCQ vs. VSA)

Question	χ^2 (Chi-square)	p-value	Statistical Interpretation
1	13.67	0.0002	Significant (MCQ > VSA), $p < 0.05$
2	0.17	0.680	Not significant
3	2.94	0.086	Not significant
4	0.61	0.434	Not significant
5	0.03	0.0869	Not significant

DISCUSSION

According to Bloom's taxonomy, which classifies educational objectives from basic knowledge to higher-order skills such as analysis and synthesis, different question formats may be more suitable for evaluating different cognitive levels (9). This specific study evaluated only the basic level of knowledge, such as specific data allowing the student to recognize or differentiate evolutionary stages or morphological characteristics of oral cavity tissues.

Based on the results, we can deduce that both Multiple Choice and Very Short Answer questions are useful for this purpose. In both cases, they serve as a platform to reach the applied knowledge that dentistry students will need in the clinical learning stage.

McCoubrie (2004) and Jabeen et al. (2023) state that MCQs are widely used due to their high reliability and efficiency in automated grading, allowing for a wide range of content coverage in a short time (4, 10). However, Guzmán-Valdivia (2016) points out limitations, such as the possibility of guessing the correct answer and the "cueing effect," where options provide hints to the student (11). Furthermore, literature suggests that MCQs often stimulate recognition-based learning and may not adequately assess higher-order thinking skills or the ability to construct a response (12).

Conversely, VSAs mitigate guessing and cueing by requiring constructed responses (typically, one word/phrase) (12). Descriptive results and t-test suggest superior MCQ mean performance (85.08% vs. 79.48%), aligning with the literature on lower scores in constructed-response formats (12, 13).

The only exception was the higher-difficulty item (p=0.59-0.69), with more correct VSA responses (nonsignificant). Presutti et al. (2019) reported that VSAs offer greater reliability and discrimination than MCQs, despite lower means (10,11).

We agree with researchers who assert that VSA questions tend to better reflect real-life clinical situations and may encourage deeper preparation for clinical practice. However, their grading requires more time and may introduce examiner bias if not rigorously standardized (10, 13). It is possible that MCQs are more suitable for formative assessment stages and VSAs for summative assessment, though this is not conclusive.

Limitations

Despite identical learning environments, non-random cohort-based format assignments risk unmeasured confounders (e.g., motivation). The sample size (n=74) detected large differences but limited the power for smaller effects and generalizability; multicenter studies are needed. Scope limited to basic Bloom level in one institution's basic dentistry cycle, non-extrapolable elsewhere. Further studies in the field should consider conducting research on higher order levels of thinking and also use multi-institutional samples in order to make it generalizable

CONCLUSION

This study provides relevant information on the use of two common assessment methods in health sciences. While the VSA exam requires more time for correction, faculty subjectively noted the students' level of concentration and dedication to formulating precise answers. However, VSA requires the teaching team to carefully draft each question to avoid answer dispersion and interpretation needs. MCQs, while simpler and faster to grade, require correct diagramming of distractors to reduce guessing.

This work did not address higher cognitive levels (comprehension, application, analysis). We conclude that for the exploration of basic knowledge, both systems are adequate if employed with a clear understanding of their strengths and weaknesses and the teaching-learning objectives.

Conflict of Interest The authors declare no conflicts of interest regarding the publication of this article.

REFERENCES

1. Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. Medical Teacher, 29(8), 746-762. https://doi.org/10.1080/01421590701775453
2. Dudley, R. A. (2012). The washback effect: considering how we assess clinical practice. Nurse Education Today, 32(7), 727-732. https://doi.org/10.1016/j.nedt.2012.05.012
3. Newble, D. I., & Jaeger, K. (1983). The effect of assessments and examinations on the learning of medical students. Medical Education, 17(3), 165-171. https://doi.org/10.1111/j.13652923.1983.tb00657.x

4. McCoubrie, P. (2004). Improving the fairness of multiple-choice questions: a literature review. Medical Teacher, 26(8), 709-712. https://doi.org/10.1080/01421590400013442
5. van Wijk, E. V., Janse, R. J., Ruijter, B. N., Rohling, J. H. T., van der Kraan, J., Crobach, S., et al. (2023). Use of very short answer questions compared to multiple choice questions in undergraduate medical students: An external validation study. PLoS ONE, 18(7), e0288558. https://doi.org/10.1371/journal.pone.0288558
6. Gupta, P., Verma, M., & Sudesh, G. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment of medical students. Journal of Family Medicine and Primary Care, 10(1), 724-728. https://doi.org/10.4103/jfmpc.jfmpc_1073_20
7. Downing, S. M. (2003). Validity: on the meaningful interpretation of assessment data. Medical Education, 37(9), 830-837. https://doi.org/10.1046/j.1365-2923.2003.01594.x
8. Martínez-Ramos, M. (2016). Evaluación de reactivos de opción múltiple en medicina: evidencia de validez de un instrumento [Evaluation of multiple choice items in medicine: evidence of validity of an instrument]. Investigación en Educación Médica, 5(19), 198-206. https://doi.org/10.1016/j.riem.2016.05.002
9. Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. New York: David McKay Company.
10. Jabeen, M., Syeda, K., Akhter, R., & Saeed, R. (2023). Validity of multiple choice questions for the assessment of medical post graduate students-a retrospective observational study. Journal of the Pakistan Medical Association, 73(6), 1345-1350. https://doi.org/10.47391/JPMA.6878
11. Guzmán-Valdivia, C. H. (2016). Reflexiones sobre adivinar en preguntas de opción múltiple y sus implicaciones en la evaluación [Reflections on guessing in multiple choice questions and its implications in assessment]. Investigación en Educación Médica, 5(19), 55-60. https://doi.org/10.1016/j.riem.2016.04.005
12. Presutti, R., Lucero, E., Garfunkel, V., Garófalo, M. E., Garzon, C., Nucifora, E., et al. (2019). Impacto de diferentes modalidades de evaluación en el promedio de notas de los estudiantes de Farmacología [Impact of different assessment modalities on the average grades of Pharmacology students]. Revista del Hospital Italiano de Buenos Aires, 39(2), 64-68. https://doi.org/10.51952/1853-0605-39-2-64
13. Leenen, I. (2014). Virtudes y limitaciones de la teoría de respuesta al ítem para la evaluación educativa en las ciencias médicas [Virtues and limitations of item response theory for educational assessment in medical sciences]. Investigación en Educación Médica, 3(9), 40-55. https://doi.org/10.1016/S20075057(14)72729-9