European Journal of Open Education and E-learning Studies
ISSN: 2501-9120
ISSN-L: 2501-9120
Available on-line at: www.oapub.org/edu

DOI: 10.46827/ejoe.v10i2.6016 Volume 10 | Issue 2 | 2025

TEACHERS' PERSPECTIVES ON STUDENTS' DIFFICULTIES IN USING MULTIPLE-CHOICE ANSWER SHEETS IN THE 2025 NATIONAL HIGH SCHOOL GRADUATION EXAMINATION: A CASE STUDY IN NORTHERN MOUNTAINOUS PROVINCES OF VIETNAM

Tran Quang Hieu[i],
Nguyen Thị Phuong,
Tran Thi Thu
Thai Nguyen University of Education,
Vietnam

Abstract:
In the context of the Ministry of Education's reform of question formats and the use of multiple-choice answer sheets in the National High School Graduation Examination from 2025, research into the difficulties students face will help provide appropriate support strategies and practical advice for students during the test-taking process. This study aims to answer two research questions: (1) What difficulties do students encounter when using the current multiple-choice answer sheets? and (2) Is the application of AI in grading multiple-choice exams feasible, and how should it be implemented? The findings reveal that students in the northern mountainous provinces of Vietnam experience significant challenges with the answer sheets, particularly in misfilling answers, time-consuming corrections, and handling short-answer items. Most teachers support replacing answer shading with direct answer writing and highly value the potential of AI in grading. However, concerns remain regarding the accuracy of handwriting recognition and data security, calling for appropriate technical solutions and communication strategies. Through this study, we emphasize the importance of guiding students on commonly mistaken areas and encourage further research and development of AI-based assessment solutions in education.

Keywords: multiple-choice answer sheets, student, national high school graduation examination

[i] Correspondence: email hieutq@tnue.edu.vn

Copyright © The Author(s). All Rights Reserved.

1. ICT as A Practical Educational Tool

Since the second half of the 20th century, objective multiple-choice testing combined with optical mark recognition (OMR) scanning has become the standard solution in high-stakes examinations such as the SAT, ACT, and A-levels, thanks to its high reliability and rapid processing speed. OMR systems operate reliably, yet their accuracy depends heavily on how precisely candidates fill the bubbles and on a scanner's recognition capability.

In Viet Nam, the Ministry of Education and Training (MOET) first piloted multiple-choice items in the 2006 university-entrance examination, then expanded the format to four subjects (Physics, Chemistry, Biology, and English) in 2007. By the 2017 National High-School Graduation Examination, most subjects had shifted to multiple-choice assessment. On 24 March 2025, MOET issued Circular 1239/BGDĐT-QLCL, introducing a new-generation answer sheet for the 2025 examination: the sheet now includes two True/False bubbles and a short-answer box, in line with Decision 764/QĐ-BGDĐT (8 March 2024) that set the new test structure.

Nevertheless, international research shows that mis-bubbling errors misalignment, double marking, or incomplete erasure occur in 0.3 % to 0.7 % of responses, enough to cause significant score deviations in highly competitive exams. Recent studies on fault-tolerant OMR indicate that geometric distortions (skew, shear) and stray marks can still push overall accuracy below 95 % on real data sets. Test handbooks for major exams (e.g., India's NEET 2025) continue to warn candidates about score losses from excess ink or faint markings.

The rapid advances in artificial intelligence and computer vision have opened more flexible options for automated scoring: deep-learning models such as OMRNet and CheckIt now achieve accuracies of up to 97 % and can even grade images captured by a webcam, trimming infrastructure costs. Handwriting-recognition (OCR) systems are being trialed for short-answer grading, yet concerns remain over reliability, data security, and large-scale deployment where resources vary widely.

Because the 2025 answer sheet will debut in a nationwide examination and empirical data on student difficulties in the mountainous northern provinces as well as teachers' attitudes toward AI-assisted grading are still scarce, a comprehensive reassessment of the current approach, its limitations, and AI's potential applications is urgently needed before the first administration of the new format in June 2025.

2. Literature Review

Multiple-choice questions (MCQs) remain a prevalent assessment method in education due to their high objectivity, reliability, and validity when properly constructed (Brady, 2005; Mallick & Ahsan, 2019). The validity and reliability of MCQs can be evaluated using various statistical methods, including Rasch measurement models (Mohd Dzin & Lay,

2021). Developing effective MCQs requires careful planning, in-depth subject knowledge, and adherence to guidelines to minimize flaws (Mallick & Ahsan, 2019; T., 1995). Reliability, a key foundation of assessment quality, is crucial for high-stakes decisions and supports validity (Miller, 2019). However, the utility of MCQs in modern healthcare education depends on question quality, desired clinical competence levels, and addressing confounding variables (Parekh & Bahadoor, 2024). Recent developments in educational measurement incorporate sociocognitive perspectives to address challenges in large-scale testing and new technologies (Mislevy, 2018). Overall, MCQs remain a valuable tool when used appropriately within a comprehensive assessment strategy (Brady, 2005; M. *et al.*, 1994).

Optical Mark Recognition (OMR) is a widely used technique for processing large volumes of hand-filled forms quickly and accurately (Espitia *et al.*, 2019; Patel & Zaid, 2017). Recent advancements in image processing have improved OMR systems, achieving high accuracy rates of up to 99.83% (Espitia *et al.*, 2019; Jingyi *et al.*, 2021). When using OMR for educational assessments, it is crucial to consider the psychometric properties of the tests, particularly validity and reliability (Solomon Chukwu Ohiri & Nnennaya, n.d.). Different scoring methods can affect these properties, with empirically determined scoring systems demonstrating superior reliability and validity compared to other methods (Echternacht, 1976).

Multiple-choice questions (MCQs) have been used in examinations for over a century, with their origins traced back to 1915 (D. Howell *et al.*, 2017). While essay-type exams remained prevalent in Europe, MCQs gained popularity in the United States and Latin America (Kellaghan & Greaney, 2019). The adoption of MCQs in medical examinations in Britain increased steadily, though literature on the subject was limited (Lennox, 1967). MCQs became favored in international certificate exams like TOEFL due to their efficiency in grading (Ammar, 2009). Initially, specialized machines were used for grading MCQ exams, but later, computer-based systems were developed to automate the process(Ammar, 2009; Harris & Buckley-Sharp, 2009). The introduction of MCQs in pathology exams at London University medical schools in the late 1960s led to the implementation of automated scoring systems (STERN *et al.*, 1973). Despite challenges, research at institutions like ETS has been instrumental in overcoming obstacles related to constructed-response formats (Bejar, 2017).

Multiple-choice tests using optical mark recognition (OMR) answer sheets are widely used, but can be prone to errors. Studies have shown that students make frequent marking errors on separate answer sheets, with error rates about three times higher compared to answering in test booklets (Muller *et al.*, 1972). Research comparing multiple-choice (MC), true-false (TF), and short-answer (SA) exam formats has yielded mixed results. Some studies found MC tests to be more reliable than TF tests (Frisbie, 1974), while others showed multiple true-false (MTF) formats to be more reliable than MC (Kreiter & Frisbie, 1989). SA questions were found to be the most valid assessment tool in one study (Rasaeian, 2004).

The 2025 session will be the first National High-School Graduation Examination held under the 2018 General Education Curriculum. Earlier, at the end of December 2023, the Ministry of Education and Training (MOET) released sample papers illustrating the new exam format that will be used from 2025 onward (Ministry of Education and Training, 2023). These prototype papers, based on the revised structure, were piloted in five provinces and municipalities—Hanoi, Hai Phong, Ninh Binh, Gia Lai, and Thai Nguyen—with nearly 5,000 students. The results serve as a foundation for building the item bank and drafting official exam papers for 2025 and beyond. In October 2024, MOET published 18 reference papers for the 2025 examination. For multiple-choice subjects, each paper divides questions into three sections: traditional multiple choice, True/False items, and short-answer multiple choice (Ministry of Education and Training, 2024).

Regarding the answer sheet to be used in the exam, MOET unveiled a draft version in April 2025 featuring two notable additions: separate areas for True/False responses and for short answers. The new sheet retains the overall layout of the earlier draft but expands the candidate-ID bubble grid to eight digits (previously six) and the test-code grid to four digits (previously three) (Ministry of Education and Training, 2025). The reverse side carries detailed instructions to help students complete the sheet accurately.

Recent studies have explored the potential of AI systems for grading handwritten short-answer responses in various academic disciplines. These systems have demonstrated high accuracy and consistency compared to human graders, with some outperforming human raters in terms of reliability (Gobrecht *et al.*, 2024).

Therefore, if AI is to be employed in grading the National High-School Graduation Examination, research must first assess the difficulties and issues students face. Such studies will clarify current trends, confirm the need for AI, and indicate how best to deploy it to support learners.

3. Material and Methods

To gain a deep and comprehensive understanding of the research problem, we adopted a multidimensional approach that combined three analytic methods: document content analysis, expert content analysis, and mathematical statistical analysis.

To fully grasp the regulations governing the use of answer sheets in the 2025 National High-School Graduation Examination, we conducted an in-depth review of relevant sources. We examined legal documents and guidelines issued by the Ministry of Education and Training to clarify the rules on test structure, question formats, multiple-choice answer-sheet design, and the specific answer-sheet template mandated for the 2025 exam. This analysis enabled us to understand more clearly how the 2025 answer sheets will be used and provided a solid foundation for formulating recommendations to minimize student errors in the examination.

The study was designed to address two main questions: (1) What difficulties do students face when using the new answer-sheet template from 2025 onward? (2) How confident are teachers in employing artificial intelligence (AI) for grading? To explore these issues, we created a survey and distributed it to teachers to gather their evaluations on both research questions. Feedback from these experts offers deeper insight into teachers' attitudes and perspectives toward adopting the 2025 multiple-choice answer-sheet format and AI-based grading in their teaching practice.

The survey targeted teachers at upper-secondary schools in northern provinces (Bac Ninh, Bac Giang, Thai Nguyen, Ha Giang, Hanoi, Cao Bang, Bac Kan, and Quang Ninh). Data were collected via an electronic questionnaire hosted on Google Forms (link: https://forms.gle/oCgXGeF5Tt89banq9). After the pedagogical shifts prompted by the COVID-19 pandemic in Viet Nam, completing Google Forms questionnaires has become routine and presented no difficulty for the respondents. Data were gathered from 21 March 2025 to 31 March 2025, yielding 38 valid responses that met the requirements for the subsequent data-analysis phase (Table 1).

Table 1: Survey sample

Category	**Teaching experience**		
	< 5 years	**5-10 years**	**> 10 years**
Number of teachers	8	8	22
Percentage (%)	21,1	21,1	57,9

The research sample shows a diverse distribution of teaching experience, with most respondents having more than 10 years in the classroom (57.9 %), a factor that directly enhances the reliability of the information gathered. In addition, the participation of younger teachers and those with medium-length careers (each group 21.1 %) brings complementary perspectives to the study.

Table 2: Key interview questions

I. General Observations on Students' Multiple-Choice Work		
Q1	In your opinion, do students experience difficulties when shading multiple-choice answers?	• ○ Very often • ○ Occasionally • ○ Rarely • ○ Never
Q2	In your opinion, what are the most common difficulties students face when using the multiple-choice answer sheet? *(You may select more than one option)*	• ☐ Shading the wrong option • ☐ Difficulty correcting an error • ☐ Wasting time when shading short answers • ☐ Other (please specify))
Q3	Which option do students most commonly shade incorrectly, based on your observation? *(You may select more than one option)*	• ☐ A • ☐ B • ☐ C • ☐ D • ☐ True/False • ☐ Short answer box

Q4	For short-answer items, which of the following difficulties do students encounter most often? *(You may select more than one option)*	• ☐ Taking too much time to answer • ☐ Easily making mistakes while shading the response • ☐ Finding it hard to correct an incorrect answer • ☐ Other difficulty (please specify) ____
Q5	How do you rate the proposal to replace shading with writing the answer directly (A, B, C, D, Đ, S or a number)?	• ○ Very effective • ○ Possibly effective • ○ Ineffective • ○ No need to change
Q6	If students write answers instead of shading, what difficulties do you foresee? *(You may select more than one option)*	• ☐ Illegible handwriting • ☐ Takes more student time • ☐ Software struggles to recognise answers • ☐ Other difficulty (please specify) ____
II. Use of Artificial Intelligence (AI) In Automated Grading		
Q7	How do you evaluate the use of AI to support exam grading?	• ○ Strongly support • ○ Support but have some concerns (please specify): __ • ○ Do not agree (please state the reason): _____
Q8	What are your main concerns regarding the use of AI in automated exam grading? *(You may select more than one option)*	• ☐ Low accuracy • ☐ Errors in recognizing handwriting • ☐ Data security issues • ☐ Legal and transparency issues • ☐ Other difficulties (please specify): _____
Q9	Have you heard of the Mathpix OCR tool for handwriting recognition?	• ☐ Aware and have used it • ☐ Aware but have not used it • ☐ Not aware of it
Q10	Based on your experience, what should be done to improve the accuracy of AI-based answer recognition? *(You may select more than one option)*	• ☐ Instruct students to write clearly and properly • ☐ Upgrade the answer recognition algorithm • ☐ Redesign the answer sheet format • ☐ Other (please specify): _____
Q11	In your opinion, which subject should be prioritized for the initial application of AI-based automated grading? *(You may select more than one option)*	• ☐ Mathematics • ☐ Physics • ☐ Chemistry • ☐ Biology • ☐ Geography • ☐ Other subject (please specify): _____
Q12	Do you have any additional recommendations for improving the answer-sheet template or the technical solutions proposed by the research team?	• Suggestion: _____

The collected data was statistically analyzed to provide observations and evaluations, offering an overview of students' difficulties and the use of AI in exam grading.

4. Results and Discussion

4.1. Assessing Students' Difficulties with the Multiple-Choice Shading Format

Survey data reveals that shading the answer sheet is a common technical bottleneck: 65.8% of teachers observed that students "frequently" or "very frequently" encountered difficulties, while only 34.2% believed such issues were rare. Three prominent issues characterize these difficulties:

1) correcting shading errors (65.8%),
2) shading the wrong bubble (57.9%), and
3) wasting time on short-answer formats (39.5%).

Analysis suggests that all three challenges stem from the mechanical design of the shading grid, rather than from cognitive limitations of the test-takers.

Errors in shading are not randomly distributed but tend to occur in central or irregularly formatted bubbles. Specifically, most mistakes were found in option B and the True/False cells (both 34.2%), followed by option C (28.9%), while option A accounted for only 18.4%. This pattern suggests that central placement and inconsistent labeling reduce positional accuracy. In short-answer responses—where students must shade multiple digits—the risk increases significantly: 76.3% of teachers reported that students often make mistakes during the transition from writing to locating and shading the correct boxes. This explains why some students lose points despite mastering the content.

The root cause lies in the design: answer bubbles are positioned too closely, forcing eye-hand coordination to constantly switch between the test booklet and the answer sheet. When mistakes occur, the erasing–reshading process is time-consuming, smudges the paper, and impairs scanner recognition. As a result, time pressure and psychological stress rise, hindering students' ability to concentrate on content.

Given these issues, the top priority should be redesigning the answer sheet: increasing spacing between bubbles, using more contrasting symbols, and adding safe correction zones. In parallel, error-checking skills should be taught: students should be instructed to double-check rows and columns before and after shading, and be guided in how to properly correct mistakes to avoid faint marks. Ultimately, the success of advanced solutions such as AI-assisted grading depends on optimizing the "input" phase. Thus, shading-related issues are not individual problems but systemic challenges in design and process that can be mitigated with proper intervention.

4.2. Teachers' Views on Replacing Shading with Direct Answer Writing

When asked about allowing students to write answers directly (e.g., A, B, C, D, T, F, or numbers), 79% of teachers responded positively (7.9% "very effective," 71.1% "potentially effective"), while only 21.1% felt that such change was unnecessary. No respondent rated the method as "ineffective." The high level of support reflects the alignment between this proposed solution and the technical difficulties described earlier:

teachers expect that eliminating shading will free students from mechanical errors and allow greater focus on academic competence.

Nevertheless, three key concerns accompanied this support. First and most prominent is the accuracy of handwriting recognition software: 86.8% of teachers were worried that AI would misinterpret handwritten answers. Second, 57.9% cited the variation in students' handwriting styles as a challenge, indicating that "input quality" remains a major barrier. Third, 23.7% feared that response time may increase, although this was a less critical concern.

This gap between support and hesitation suggests a three-phase implementation roadmap:

1) Pilot testing with structured question types to build a real-world training dataset for AI and demonstrate effectiveness.
2) Handwriting standardization: 71.1% of teachers suggested training students to write neatly prior to exams, combined with optimized answer sheet design for writing.
3) Algorithm improvement: Besides enhancing recognition accuracy, the system should provide warnings when confidence is low so that human reviewers can verify results and reduce misgrading risk.

Communication strategies are also crucial: most teachers are "cautiously optimistic"—they prefer empirical evidence over theoretical promises. Therefore, publishing trial data and enabling user feedback will strengthen trust and help persuade the 21.1% who remain skeptical. Once AI infrastructure is reliable and handwriting is standardized, the direct-writing format will not only solve traditional shading problems but also pave the way for a digital grading ecosystem that is more transparent and equitable in the current era of assessment reform.

4.3. Evaluating the Potential, Challenges, and Roadmap for AI-Based Exam Grading

Most teachers view AI as a breakthrough solution to overcome the mechanical bottlenecks of traditional exam grading. In the entire sample, 100% of teachers supported integrating AI into grading, with 68.4% strongly in favor and 31.6% supporting with reservations—no respondents opposed the idea outright. This support rate far exceeds the 7.9% who rated direct-answer writing as "very effective," reflecting strong belief that technology can fundamentally solve technical issues that procedural changes alone cannot. Teachers expect AI to:

1) automate grading with high accuracy,
2) reduce manual workload, and
3) establish fast and transparent feedback mechanisms for both students and reviewers.

However, optimism comes with three major categories of concern. The first is technical: 73.7% of teachers worry about errors in handwriting recognition—the most prominent concern. Relatedly, 86.8% believe that current software struggles to identify

handwritten answers, and 57.9% emphasized the challenge posed by varied handwriting styles. The second is non-technical: concerns about data security (60.5%) and legal/transparency issues (34.2%) indicate growing attention to privacy and algorithm accountability. Lastly, the technological literacy gap remains wide: 71.1% of teachers had never heard of specialized OCR tools like Mathpix, and only 2.6% had used them. This awareness gap suggests that training and communication must accompany technical development.

The data points to a strategic triangle—**people – process – technology**—as the optimal approach. Regarding people, 71.1% of teachers recommend instructing students to write more clearly to improve AI input; for process, 57.9% propose redesigning the answer sheet; and for technology, 55.3% emphasize upgrading recognition algorithms. Initial deployment should focus on subjects with clearly structured answers such as Mathematics (86.8%) as the pilot area, followed by Physics (65.8%) and Chemistry (55.3%). Subjects with more semi-structured responses like Biology (31.6%) and Geography (15.8%) should be included once the system is more refined. Small-scale pilots should combine AI with human evaluation to calibrate models and build trust, gradually expanding once accuracy and reliability exceed acceptable thresholds.

In summary, teachers view AI as a strategic lever to shift multiple-choice testing from mechanical scoring to intelligent assessment in a transparent and effective manner. The success of this roadmap depends on aligning three pillars: user training, process optimization, and technical capacity—especially in handwriting recognition and data security. When these factors are synchronized, AI will not only eliminate procedural errors but also lay the foundation for a smart, fair, and adaptive grading ecosystem in Vietnam's digital education transformation.

5. Conclusion

The study affirms the pivotal role of design and procedural factors in ensuring the reliability of multiple-choice assessments. When the process of presenting answers involves complex mechanical actions, assessment results can be influenced more by shading skills than by students' academic competence. This phenomenon highlights the need for a comprehensive rethinking: rather than merely optimizing question content, equal attention must be given to the ergonomics of the answer sheet, the test-taker's procedural skills, and the underlying grading infrastructure.

In this context, artificial intelligence (AI) emerges as a strategic enabler, but its effectiveness can only be fully realized under three conditions:

1) test materials must be user-friendly in terms of operation,
2) teachers must possess adequate digital competencies to monitor and calibrate algorithms, and
3) data processing procedures must be transparent, secure, and accountable.

Once these conditions are met, AI not only eliminates mechanical errors but also unlocks the potential for system-level learning analytics that inform curriculum and instructional improvements.

More broadly, the shift from "manual scoring" to "data-driven assessment" suggests a learner-centered evaluation framework where presentation mechanics no longer serve as barriers. This transformation requires close collaboration among schools, educational researchers, and technology developers to ensure that pedagogical benefits are prioritized over technical sophistication. From a research perspective, the next step is to examine the long-term impact of new answer sheet formats and recognition algorithms on student motivation, exam-related anxiety, and educational equity—thereby advancing the vision of a fully digitized assessment landscape in Vietnam.

Conflict of Interest Statement

The authors declare no conflicts of interest.

About the Author(s)

Dr. Tran Quang Hieu is a lecturer at the Faculty of Physics, Thai Nguyen University of Education, Vietnam. He holds a Ph.D. in Physics and is currently involved in teaching and conducting research in the field of physics education and related interdisciplinary approaches. His academic interests include innovative teaching methodologies, integration of technology in science education, and the development of critical thinking skills in students.
ORCID: https://orcid.org/0009-0008-5618-6517
Email: hieutq@tnue.edu.vn

References

Ammar, A. (2009). *An Automatic System to Grade Multiple Choice Questions paper based exams*. Retrieved from https://www.researchgate.net/publication/312136231_An_Automatic_System_to_Grade_Multiple_Choice_Questions_paper_based_exams

Bejar, I. I. (2017). A Historical Survey of Research Regarding Constructed-Response Formats. In *Methodology of Educational Measurement and Assessment* (pp. 565–633). Springer International Publishing. https://doi.org/10.1007/978-3-319-58689-2_18

Brady, A.-M. (2005). Assessment of learning with multiple-choice questions. *Nurse Education in Practice*, *5*(4), 238–242. https://doi.org/10.1016/j.nepr.2004.12.005

D. Howell, D., Tseng, D. C., & T. Colorado-Resa, J. (2017). Fast Assessments with Digital Tools Using Multiple-Choice Questions. *College Teaching*, *65*(3), 145–147. https://doi.org/10.1080/87567555.2017.1291489

Echternacht, G. (1976). Reliability and validity of item option weighting schemes. *Educational and Psychological Measurement*, *36*(2), 301–309. https://doi.org/10.1177/001316447603600208

Espitia, O., Paez, A., Mejia, Y., Carrasco, M., & Gonzalez, N. (2019). Optical mark recognition based on image processing techniques for the answer sheets of the Colombian high-stakes tests. *Applied Computer Sciences in Engineering: 6th Workshop on Engineering Applications, WEA 2019, Santa Marta, Colombia, October 16–18, 2019, Proceedings 6*, 167–176.

Frisbie, D. A. (1974). The Effect of Item Format on Reliability and Validity: a Study of Multiple Choice and True-False Achievement Tests. *Educational and Psychological Measurement*, *34*(4), 885–892. https://doi.org/10.1177/001316447403400417

Gobrecht, A., Tuma, F., Möller, M., Zöller, T., Zakhvatkin, M., Wuttig, A., Sommerfeldt, H., & Schütt, S. (2024). Beyond human subjectivity and error: a novel AI grading system. *ArXiv Preprint ArXiv:2405.04323*. http://dx.doi.org/10.48550/arXiv.2405.04323

Harris, F. T. C., & Buckley-Sharp, M. (2009). Automation of Multiple-choice Examination Marking. *Medical Education*, *2*(1), 48–54. https://doi.org/10.1111/j.1365-2923.1968.tb02004.x

Jingyi, T., Hooi, Y. K., & Bin, O. K. (2021). Image processing for enhanced OMR answer matching precision. *2021 International Conference on Computer & Information Sciences (ICCOINS)*, 322–327. https://doi.org/10.1109/ICCOINS49721.2021.9497172

Kellaghan, T., & Greaney, V. (2019). A Brief History of Written Examinations. In *Public Examinations Examined* (pp. 43–74). The World Bank. Retrieved from https://doi.org/10.1596/978-1-4648-1418-1_ch3

Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of multiple true-false items. *Applied Measurement in Education*, *2*(3), 207–216. http://dx.doi.org/10.1207/s15324818ame0203_2

Lennox, B. (1967). Marking Multiple-choice Examinations. *Medical Education*, *1*(3), 203–211. https://doi.org/10.1111/j.1365-2923.1967.tb01698.x

Linn, R. & Gronlund, N. E. (1994). *Measurement and Assessment in Teaching*. Retrieved from https://www.researchgate.net/publication/242393428_Measurement_and_Assessment_In_Teaching

Mallick, A. K., & Ahsan, M. (2019). Multiple Choice Questions: A Potent Tool for Assessment of Medical Stu dents. In *Indian Journal of Medical & Health Sciences*. Red Flower Publication Private, Ltd. https://doi.org/10.21088/ijmhs.2347.9981.6119.1

Miller, M. D. (2019). Reliability in Educational Assessments. *Education*. https://doi.org/10.1093/obo/9780199756810-0228

Ministry of Education and Training. (2023). *The Exam Format Structure for the National High School Graduation Examination from 2025.* Electronic Information Portal of the Ministry of Education and Training of Vietnam. Retrieved from https://moet.gov.vn/tintuc/Pages/tin tong hop.aspx?ItemID=8979

Ministry of Education and Training. (2024). *Decision No. 764/QĐ-BGDĐT: Regulation on the Exam Format Structure for the National High School Graduation Examination from 2025.*

Ministry of Education and Training. (2025). *Official Dispatch No. 1239/BGDĐT-QLCL: Guidelines on Certain Aspects of Organizing the 2025 National High School Graduation Examination.*

Mislevy, R. J. (2018). *Sociocognitive Foundations of Educational Measurement*. Routledge. https://doi.org/10.4324/9781315871691

Mohd Dzin, N. H., & Lay, Y. F. (2021). Assessing the Validity and Reliability of Science Multiple Choice Test Using Rasch Dichotomous Measurement Model. *Journal of Baltic Science Education*, *20*(6), 924–941. m https://doi.org/10.33225/jbse/21.20.927

Muller, D., Calhoun, E., & Orling, R. (1972). Test Reliability as a Function of Answer Sheet Mode. *Journal of Educational Measurement*, *9*(4), 321–324. https://doi.org/10.1111/j.1745-3984.1972.tb00964.x

Parekh, P., & Bahadoor, V. (2024). The Utility of Multiple-Choice Assessment in Current Medical Education: A Critical Review. *Cureus*. https://doi.org/10.7759/cureus.59778

Patel, D., & Zaid, S. (2017). Efficient System for Evaluation of Omr Sheet-A Survey. *International Journal of Advanced Research in Engineering, Science & Management*, *3*(7).

Rasaeian, N. (2004). *A new methodology for comparison of three-test exam techniques in medical students*. Retrieved from https://brieflands.com/articles/jme-105084.pdf

Solomon Chukwu Ohiri, D. I., & Nnennaya, C. (n.d.). *Psychometric Properties of a Test: An Overview*. http://dx.doi.org/10.55248/gengpi.5.0224.0539

Stern, M. J., Harris, F. T. C., & Buckley-Sharp, M. D. (1973). Multiple-choice question paper scores in pathology: A review of scores obtained in a pretest and in four final MB examinations of London University. *Medical Education*, *7*(1), 10–15. https://doi.org/10.1111/j.1365-2923.1973.tb02203.x

T., H. (1995). Developing and validating multiple-choice test items. *Choice Reviews Online*, *32*(06), 32–3425. https://doi.org/10.5860/choice.32-3425

Creative Commons licensing terms
Authors will retain the copyright of their published articles, agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Open Education and E-learning Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflict of interests, copyright violations and inappropriate or inaccurate use of any kind of content related or integrated on the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a Creative Commons Attribution 4.0 International License (CC BY 4.0).