

RESEARCH

Open Access



Assessing the interplay of presentation and competency in online video examinations: a focus on health system science education

Phoomjai Sornsenee^{1*}, Pawita Limsomwong¹, Polatthep Vichitkunakorn¹, Supakorn Sripaew¹, Supinya Sono¹, Kanisorn Sungkaro², Theepat Wongkittithaworn³, Thadakorn Tantisarasart⁴, Pongphon Chuchuen⁵, Katti Sathaporn⁶ and Kwanchanok Chantaramanee⁷

Abstract

Background The integration of Health System Science (HSS) in medical education emphasizes mastery of competencies beyond mere knowledge acquisition. With the shift to online platforms during the COVID-19 pandemic, there is an increased emphasis on Technology Enhanced Assessment (TEA) methods, such as video assessments, to evaluate these competencies. This study investigates the efficacy of online video assessments in evaluating medical students' competency in HSS.

Methods A comprehensive assessment was conducted on first-year medical students ($n = 10$) enrolled in a newly developed curriculum integrating Health System Science (HSS) into the Bachelor of Medicine program in 2021. Students undertook three exams focusing on HSS competency. Their video responses were evaluated by a panel of seven expert assessors using a detailed rubric. Spearman rank correlation and the Interclass Correlation Coefficient (ICC) were utilized to determine correlations and reliability among assessor scores, while a mixed-effects model was employed to assess the relationship between foundational HSS competencies (C) and presentation skills (P).

Results Positive correlations were observed in inter-rater reliability, with ICC values suggesting a range of reliability from poor to moderate. A positive correlation between C and P scores was identified in the mixed-effects model. The study also highlighted variations in reliability and correlation, which might be attributed to differences in content, grading criteria, and the nature of individual exams.

Conclusion Our findings indicate that effective presentation enhances the perceived competency of medical students, emphasizing the need for standardized assessment criteria and consistent assessor training in online environments. This study highlights the critical roles of comprehensive competency assessments and refined presentation skills in online medical education, ensuring accurate and reliable evaluations.

*Correspondence:
Phoomjai Sornsenee
eziinary@gmail.com; poomjai.s@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Keywords Health system science, Online video assessments, Competency-based assessment, Presentation skills, Medical curriculum

Background

In medical education, assessment extends beyond mere knowledge acquisition. It emphasizes the mastery of essential skills and the medical profession. [1] Among various approaches, competency-based assessment stands out, [2] as it encompasses the full range of knowledge, skills, and attitudes. [3] This approach focuses extensively on Health System Science (HSS), which is globally recognized as the third pillar of medical education. [4] HSS integrates systems thinking and the complexities of healthcare systems, equipping medical professionals to adeptly navigate and manage the socio-economic, political, and interpersonal factors influencing healthcare. [5] Furthermore, the cognitive domain is an integral facet of HSS, emphasizing Higher-Order Thinking (HOT) skills, such as critical thinking, problem-solving, and analytical reasoning. [6, 7] These skills are crucial to effectively engaging with the broader constructs of HSS.

The challenges brought by the COVID-19 pandemic have reshaped traditional teaching and prompted a re-evaluation of existing assessment methods. [8–10] Global institutions, faced with the limitations of in-person interactions, have found it essential to transition to digital platforms. [10] This shift has highlighted the need for adopting Technology Enhanced Assessment (TEA) methods, which align well with the emerging online teaching environment. [11, 12] In this new paradigm, it is crucial that TEA methods, especially online video assessments, effectively evaluate competencies for the digital age. However, their efficacy in specific areas like HSS remains under exploration.

Oral presentations, especially in online video formats, may offer unique insights into a student's HOT capabilities, which are essential for HSS assessments. These evaluations could potentially elucidate core areas within HSS, such as systems thinking, health policy, and the social sciences. Previous studies hint at the efficacy of oral assessments in areas like social sciences [13] and professionalism. [14] However, the robustness and reliability of video-based oral presentations as an assessment tool warrant further investigation. Another consideration is the potential bias resulting from varying presentation skills. [15, 16] Students with exemplary presentation skills may overshadow essential content gaps, while those less adept at presenting might not be duly acknowledged for their depth of HSS knowledge. To ensure fair and effective assessment, it is critical to carefully differentiate true competency from mere presentation prowess.

In this study, we investigate the effectiveness of online video assessments in capturing medical students' competency in the HSS curriculum, focusing on topics like systems thinking, health policy, medical ethics, and social health determinants. Our research primarily aims to gauge score consistency across assessors. We employ the Interclass Correlation Coefficient (ICC) to measure inter-rater reliability and use Spearman's Rho—a statistical method—to determine correlations among assessors. Alongside, we explore the relationship between foundational HSS competencies (C) and presentation skills (P) through mixed-effects modeling. Elucidating the interplay between C and P offers a nuanced perspective for HSS assessments.

Methods

Study design, setting, and participants

The study involved a comprehensive assessment of all first-year medical students ($n=10$) enrolled in a newly developed curriculum integrating Health System Science (HSS) into the Bachelor of Medicine program in 2021. The curriculum aims primarily at two objectives. First, it seeks to provide comprehensive medical education enriched with HSS concepts. Second, it intends to attract students from rural areas, nurturing them with the hope they will return and serve their home communities, equipped with a deep understanding of the public health system.

Examination process

First-year medical students undertook three distinct exams, each designed to assess various dimensions and complexities of Higher-Order Thinking (HOT) skills within the Health System Science (HSS) curriculum. For a detailed breakdown of exam formats, materials, and timing, refer to Table 1.

- **Exam 1 on Social Health Determinants (SDHs):** This exam assesses foundational HOT skills such as understanding, recall, and application, in addition to critical thinking and problem-solving. It integrates theoretical and practical elements to evaluate how students apply systems thinking to real-world health challenges.
- **Exam 2 on Health Care Policy:** This exam focuses on deeper analytical and synthetic HOT skills, evaluating students' abilities to critically engage with and construct reasoned arguments regarding health policy issues.

Table 1 Examination details for health system science competency assessment

Exam number	Duration	Main competency focus	Related focuses	Material and time management
Exam 1	40 min	Social Health Determinants (SDHs)	System thinking,	Short essay and video clip (21 min); 13 min for preparation, 6 min for response
Exam 2	40 min	Health Care Policy	Multiple Points of View	Video clip (22 min); 12 min for preparation, 6 min for response
Exam 3	40 min	Medical Ethics	System Thinking, Critical Thinking	Video clip (21 min); 13 min for preparation, 6 min for response

- **Exam 3 on Medical Ethics:** Similarly, this exam tests the students' ability to analyze, synthesize, and evaluate ethical dilemmas in medical practice, enhancing their understanding of the ethical frameworks essential to health system science.

Each exam incorporates elements of systems thinking, foundational to mastering HSS. The assessments were developed by an impartial educator and are conducted over a standardized 40-minute session, utilizing a Learning Management System (LMS) developed by the faculty of medicine.

Assessment technique and evaluation

After the three exams, students' response videos were assessed by a panel of seven expert assessors specializing in medical education. The assessment process utilized a detailed rubric, designed to capture the depth and nuance of HOT as intended in each exam. This rubric, informed by the principles of HOT and HSS, combined a numerical scale (ranging from 1 to 10) with explicit descriptors for each score, providing clarity on performance expectations. This ensured evaluations were consistent and adequately represented the depth of student understanding in relation to the complexities integrated into the three exams. The training of our panel of assessors focused on the foundational principles of Health System Science (HSS) and the effective application of the assessment rubric. The initial session introduced the assessors to the rubric, highlighting its alignment with key HSS competencies such as systems thinking and value based care. This included a detailed explanation of the scoring system and performance standards.

The students were assessed based on two primary criteria. **First**, their **competency skills** were evaluated using the rubric formulated by the educational research team

to capture the nuances of HOT. **Second**, their **presentation skills** were assessed, focusing on accuracy, clarity, and the effectiveness of their communication, drawing from criteria established in previous research. [16]

Statistical analysis and reliability assessment

- **Assessor Score Correlation:** We utilized the Spearman rank correlation method to understand the interrelationships among assessor scores. Data for each of the three exams (Exams 1–3) are presented in tables, which display the Spearman's rho coefficients for competency (C) scores and presentation (P) scores. A rho value near 1 signifies a strong correlation; a value between 0.5 and 0.75 indicates a moderate correlation; a value between 0.3 and 0.5 suggests a weak correlation; and a value close to 0 implies negligible correlation.
- **Reliability Estimation:** The level of agreement among assessors was quantified using the Interclass Correlation Coefficient (ICC). Each ICC value was complemented with its respective 95% confidence interval. It is essential to note the interpretative context of the ICC values: those below 0.5 signify poor reliability; values ranging from 0.5 to 0.75 denote moderate reliability; values between 0.75 and 0.9 are indicative of good reliability; and values exceeding 0.9 are demonstrative of excellent reliability.
- **Relationship between Presentation and Competency:** A jitter plot depicted individual student performances, with colors differentiating each assessor. A mixed-effects model was implemented to consider potential variability due to individual assessors. The model's coefficient indicated the anticipated shift in competency relative to presentation skills.
- **Software and Packages:** All analytical processes, from data visualization to heat map creation, were executed using R software, version 4.3.1 with packages. [17–20]

Results

Our evaluation offers insights into both the correlation and reliability of assessments. The 10 participants, scores across the three exams were evaluated by seven expert assessors. Spearman's Rho values, as illustrated in Tables 2 and 3, indicate correlations for competency scores (C) and presentation scores (P). All correlations emerged as positive; however, there were discernible variations in strength from weak to strong for both competency and presentation scores across different assessors and exams. The Interclass Correlation Coefficient

Table 2 Spearman's rho for competency scores across exams and assessors

Exam 1	A1	A2	A3	A4	A5	A6	A7
A1	1	0.69	0.74	0.74	0.61	0.7	0.45
A2	0.69	1	0.64	0.64	0.77	0.83	0.59
A3	0.74	0.64	1	1	0.38	0.68	0.49
A4	0.74	0.64	1	1	0.38	0.68	0.49
A5	0.61	0.77	0.38	0.38	1	0.8	0.53
A6	0.7	0.83	0.68	0.68	0.8	1	0.65
A7	0.45	0.59	0.49	0.49	0.53	0.65	1
Interclass correlation (ICC)=0.42 (95% CI=0.18–0.74)							
Exam 2	A1	A2	A3	A4	A5	A6	A7
A1	1	0.37	0.27	0.39	0.31	0.29	0.05
A2	0.37	1	0.47	0.51	0.66	0.7	0.21
A3	0.27	0.47	1	0.73	0.57	0.73	0.06
A4	0.39	0.51	0.73	1	0.56	0.69	0.14
A5	0.31	0.66	0.57	0.56	1	0.74	0.54
A6	0.29	0.7	0.73	0.69	0.74	1	0.3
A7	0.05	0.21	0.06	0.14	0.54	0.3	1
Interclass correlation (ICC)=0.27 (95% CI=0.09–0.61)							
Exam 3	A1	A2	A3	A4	A5	A6	A7
A1	1	0.28	0.84	0.76	0.78	0.69	0.61
A2	0.28	1	0.3	0.54	0.5	0.58	0.63
A3	0.84	0.3	1	0.75	0.58	0.81	0.61
A4	0.76	0.54	0.75	1	0.89	0.78	0.77
A5	0.78	0.5	0.58	0.89	1	0.56	0.62
A6	0.69	0.58	0.81	0.78	0.56	1	0.91
A7	0.61	0.63	0.61	0.77	0.62	0.91	1
Interclass correlation (ICC)=0.35 (95% CI=0.12–0.68)							

A = Assessor ID

(ICC) values for C scores across exams were 0.42 (95% CI=0.18–0.74) for Exam 1, 0.27 (95% CI=0.09–0.61) for Exam 2, and 0.35 (95% CI=0.12–0.68) for Exam 3. P scores for these exams were 0.44 (95% CI=0.20–0.75) for Exam 1, 0.32 (95% CI=0.12–0.66) for Exam 2, and 0.48 (95% CI=0.24–0.78) for Exam 3. When considering the 95% confidence intervals for both competency and presentation scores, these ICC values suggest a range of reliability from poor to moderate.

Further exploration of the correlation between C and P scores across the three exams is depicted in Figs. 1 and 2, and 3, respectively. Each figure presents a jitter plot that highlights individual data points; different colors represent individual assessors. The overall mixed-effects model prediction is captured by a red dotted line, flanked by the 95% confidence interval (CI). For Exam 1, the model indicates a positive correlation between C and P scores with a coefficient of 0.16, but this is not statistically significant ($p=0.073$). In contrast, both Exams 2 and 3 consistently showed statistically significant positive correlations between C and P scores, with all assessors reflecting this trend.

Discussion

Medical education, especially in areas like HSS, has significantly evolved due to global trends. As digital platforms have surged in popularity, exacerbated by the COVID-19 pandemic, the emphasis on online assessments has grown. Our study, deeply rooted in HSS, sheds light on the broader challenges of online evaluations. Specifically, it underscores the balance required between content mastery and effective presentation in an environment where both are paramount.

One key finding from our study is the role of assessor calibration. Using statistical measures like ICC and Spearman's rho, we observed a positive correlation in inter-rater reliability, accompanied by variations. It is imperative to recognize that our research's primary objective was not solely to evaluate the efficiency of our rubric scores but to scrutinize the efficacy of our video assessment design for HSS. Prior studies support our conclusions, suggesting that rubric-based grading can have varying reliability. [21–23] Such deviations can arise due to factors including the content under examination, the adopted grading criteria, statistical evaluation methods, the nature of the test, and the number of involved assessors or participants. [21, 24]

Table 3 Spearman’s rho for presentation scores across exams and assessors

Exam 1	A1	A2	A3	A4	A5	A6	A7
A1	1	0.36	0.46	0.46	0.66	0.53	0.24
A2	0.36	1	0.77	0.77	0.76	0.75	0.14
A3	0.46	0.77	1	1	0.66	0.8	0.44
A4	0.46	0.77	1	1	0.66	0.8	0.44
A5	0.66	0.76	0.66	0.66	1	0.75	-0.07
A6	0.53	0.75	0.8	0.8	0.75	1	0
A7	0.24	0.14	0.44	0.44	-0.07	0	1
Interclass correlation (ICC)=0.44 (95% CI=0.21–0.75)							
Exam 2	A1	A2	A3	A4	A5	A6	A7
A1	1	0.05	0.48	0.37	0.37	0.59	0.19
A2	0.05	1	0.63	0.65	0.53	0.31	0.71
A3	0.48	0.63	1	0.7	0.47	0.46	0.56
A4	0.37	0.65	0.7	1	0.72	0.79	0.33
A5	0.37	0.53	0.47	0.72	1	0.68	0.02
A6	0.59	0.31	0.46	0.79	0.68	1	0.21
A7	0.19	0.71	0.56	0.33	0.02	0.21	1
Interclass correlation (ICC)=0.32 (95% CI=0.12–0.66)							
Exam 3	A1	A2	A3	A4	A5	A6	A7
A1	1	0.51	0.59	0.63	0.44	0.17	0.4
A2	0.51	1	0.7	0.79	0.9	0.49	0.84
A3	0.59	0.7	1	0.76	0.73	0.32	0.71
A4	0.63	0.79	0.76	1	0.83	0.49	0.71
A5	0.44	0.9	0.73	0.83	1	0.71	0.88
A6	0.17	0.49	0.32	0.49	0.71	1	0.58
A7	0.4	0.84	0.71	0.71	0.88	0.58	1
Interclass correlation (ICC)=0.48 (95% CI=0.24–0.78)							

A= Assessor Number

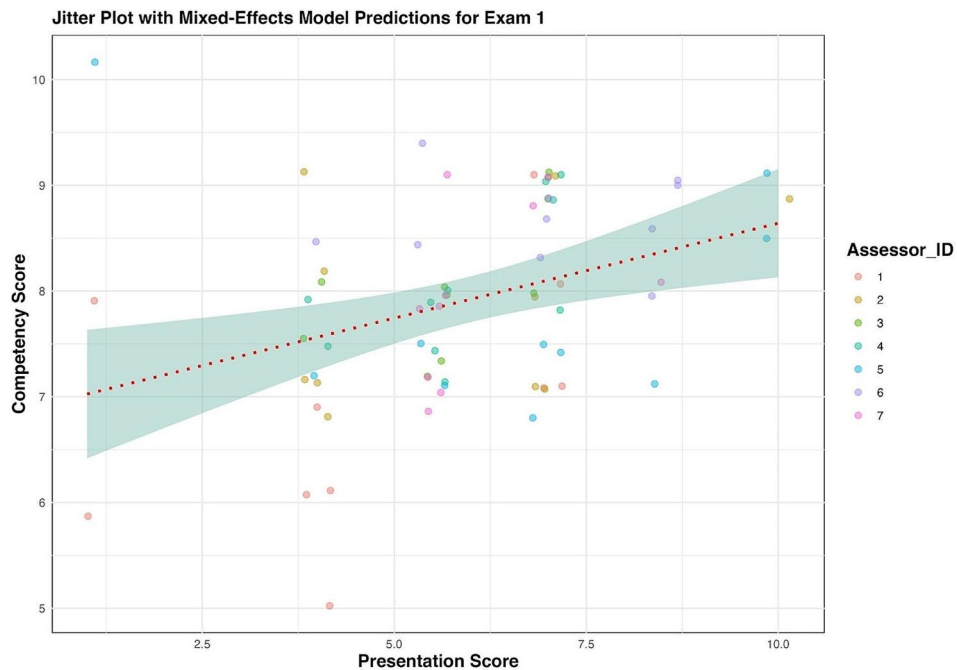


Fig. 1 Jitter plot of exam 1 with individual assessor lines and mixed-effects model correlating competency and presentation scores

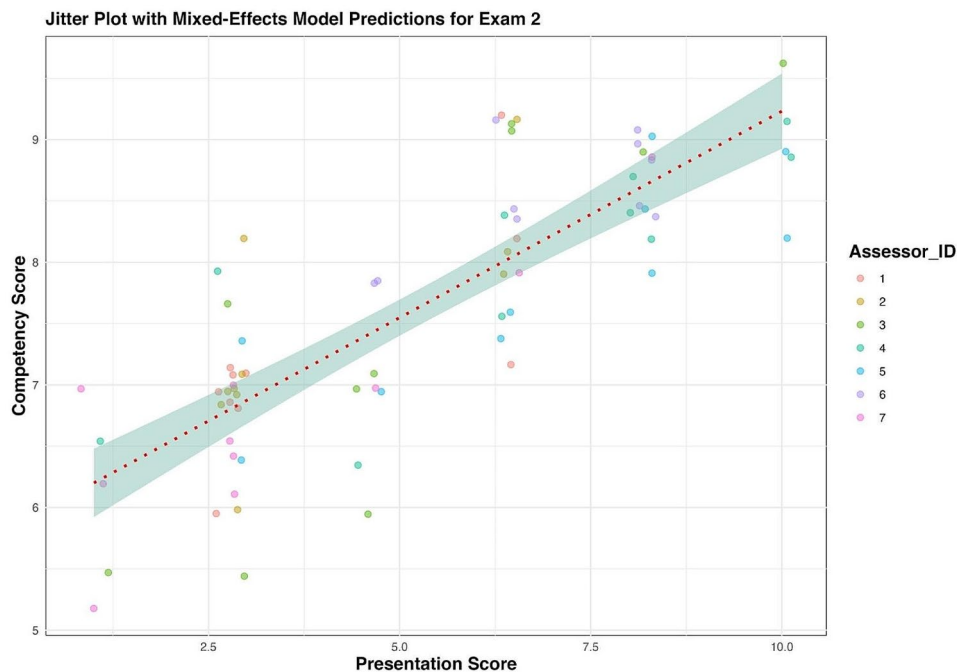


Fig. 2 Jitter plot of exam 2 with individual assessor lines and mixed-effects model correlating competency and presentation scores

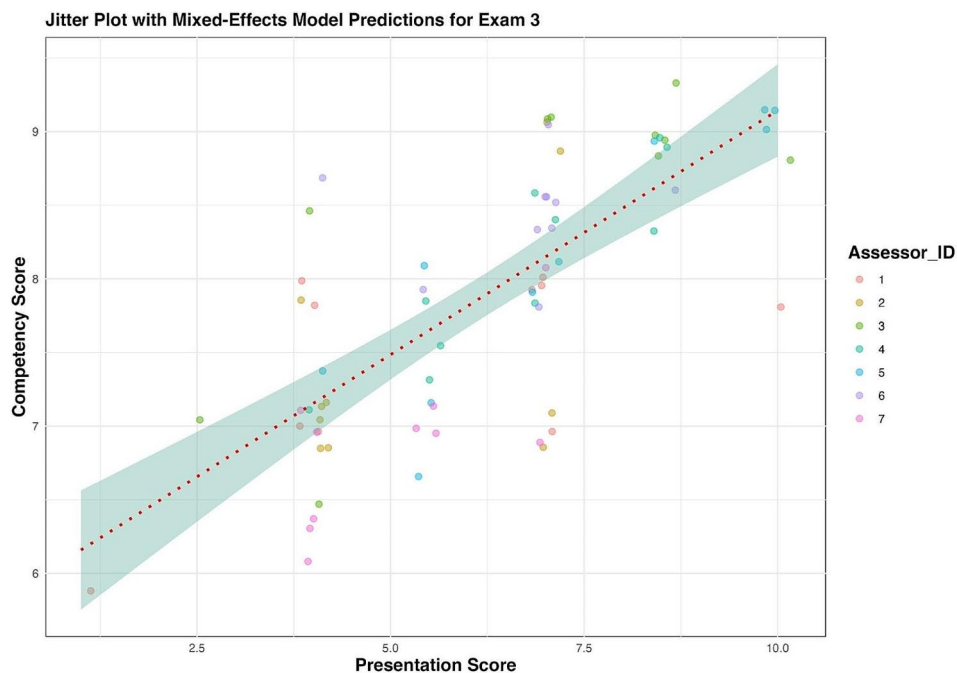


Fig. 3 Jitter plot of exam 3 with individual assessor lines and mixed-effects model correlating competency and presentation scores

In addressing the challenges of evaluating complex topics such as healthcare politics and policies, our assessor training was tailored to enhance the use of a comprehensive rubric and objective assessment skills. The training included detailed reviews of HSS case studies and scenario-based exercises, emphasizing strategies to mitigate personal biases and maintain consistency across

varied content. This approach helped ensure that all assessors were equipped to objectively assess responses, even in subjects prone to subjective interpretation. Upon deeper observation, we found that Exam 2 showed a lower correlation among scorers compared to Exams 1 and 3. This could be attributed to the content of Exam 2, which focused on healthcare politics and policies. Such

policy-oriented topics can lead to varied interpretations among assessors, making it challenging to determine a universally acceptable answer. [25]

Our research underscores the crucial balance between competency and presentation skills. With the rise of digital platforms, the ability to articulate and communicate effectively has become indispensable for professional competency. [26] A related study, utilizing platforms like Zoom for assessments, aligns with our findings. [27] It emphasizes that when evaluating online, it is essential to distinguish domains like metacognition and creativity. Many evaluation methodologies prioritize oral presentation capabilities, relegating content knowledge to a secondary role. [14, 15, 28] In disciplines demanding HOT skills, such as HSS, merging these domains could potentially obscure the importance of specific competencies. Therefore, to ensure fair and effective assessments, it is vital to carefully differentiate between true competency and the enhancement effect of presentation prowess. This vigilance is crucial because exemplary presentation skills can sometimes overshadow gaps in content knowledge, while less polished presentation skills might lead to underestimation of a student's understanding and competence. In response to this challenge, educational institutions should incorporate robust communication and presentation training modules into their curricula. [28] This would ensure that graduates are not only well-versed in their fields but also adept at articulating their knowledge.

Reflecting on our video assessment approach in HSS, the novel method we employed to understand the nexus between competency and presentation skills emerges as a strength. With the prevailing trend of transitioning to digital platforms in medical education, our insights stress the essence of both thorough content comprehension and proficient communication.

In our study, we employed a numeric scoring scale to quantify student performances, which provides precision and facilitates straightforward comparison. This method offers the advantage of detailed quantification, allowing for fine distinctions between levels of competency. However, it may introduce inconsistencies, especially when assessing complex competencies that require nuanced judgment. Alternatively, ordinal rating systems, which categorize performances into descriptive levels such as excellent, good, fair, or poor, can simplify assessments and potentially enhance consistency by more clearly delineating broad performance categories. Each system has its merits, with numeric scales offering granularity and ordinal scales providing clearer benchmarks. [29] This distinction underscores a key aspect of our current methodology and presents a valuable area for future research to explore the optimal balance between detailed quantification and categorical assessment in evaluating

both competency and presentation skills in medical education.

In this study, we intentionally employed a diverse panel of seven expert assessors rather than the conventional use of two raters. This decision was guided by the aim to enhance the reliability and depth of our evaluations, particularly given the complex competencies involved in HSS. A larger panel allows for a more comprehensive range of perspectives on student performances, which is critical in a field where subjective judgment can significantly influence scoring. However, this approach can also introduce variability in scoring due to differences in each rater's interpretation and emphasis, which might not be as pronounced with a smaller, more uniform panel. By using seven raters, we aimed to capture a broader spectrum of interpretations, which, while enriching the assessment, could also lead to increased score dispersion and affect the overall consistency of the results. This aspect of our methodology could have influenced the outcomes by either mitigating or exaggerating individual biases, thus impacting the inter-rater reliability as reported. To address these challenges, integrating AI technology could provide a valuable tool. Artificial intelligence can assist in standardizing evaluations by consistently applying predefined criteria, potentially reducing the variability introduced by multiple human assessors. [30, 31] This hybrid approach, blending human insight with AI precision, represents a promising direction for future research, aiming to balance depth and reliability in complex competency assessments.

Conclusion

Considering the observed shift towards online assessments in medical education, this study sought to evaluate the efficacy of video assessments in measuring both competency and presentation skills within the Health System Science (HSS) framework. Our findings reveal positive correlations between these skills, indicating that effective presentation can enhance the perception of competency. However, variations in inter-rater reliability highlight the necessity for standardized assessment criteria and consistent assessor training. By implementing these strategies, educational institutions can ensure greater accuracy and reliability, which is crucial for accurately assessing student competencies in a digital learning environment. Thus, our study substantiates the importance of both thorough competency assessment and the enhancement of presentation skills in online medical education.

Acknowledgements

We wish to thank the Department of Family Medicine and Preventive Medicine at Prince of Songkla University for supporting our research endeavors. Our gratitude is also extended to Assistant Professor Kanyika Chamniprasas, Vice Dean for Education, and Assistant Professor Supaporn Dissaneevate for their roles in creating and backing the health system reform curriculum for medical students, which contributed to this study.

Author contributions

All authors made a significant contribution to the reported work. P.S., P.L. S.So., T.T., P.C., K.S., K.C. designed the study and obtained research ethics approval; P.S., P.V., S.Sr., K.S., and T.W. analysed and visualized the data; all authors interpreted the results, drafted, and revised the manuscript, and read and approved the final version of the manuscript.

Funding

No funding was received for this research.

Data availability

The dataset analysed in the current study is available upon request from the corresponding author.

Declarations

Ethics approval and consent to participate

This study was approved by the Human Research Ethics Committee (HREC), Faculty of Medicine, Prince of Songkla University, on March 12, 2022. (Approval Number: REC REC.65-114-9-1). HREC is official institute responsible for the ethical review and supervision of human research. The study was conducted in accordance with the Declaration of Helsinki. All participants and/or their legal guardian(s) provided electronic informed consent before participating in the study, in accordance with established procedures. All questionnaires were fully computerised by the researchers and anonymised to ensure confidentiality.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Family Medicine and Preventive Medicine, Faculty of Medicine, Prince of Songkla University, 15 Karnjanavanich Rd., Hat Yai, 90110 Songkhla, Thailand

²Division of Neurosurgery, Department of surgery, Faculty of medicine, Prince of Songkhla University, Kho Hong, Hat Yai District, 90110 Songkhla, Thailand

³Division of Head and neck surgery, Department of otolaryngology, Faculty of medicine, Prince of Songkhla University, Kho Hong, Hat Yai District, Songkhla, Thailand

⁴Department of Anesthesiology, Faculty of medicine, Prince of Songkhla University, Kho Hong, Hat Yai District, 90110 Songkhla, Thailand

⁵Department of Rehabilitation Medicine, Faculty of medicine, Prince of Songkhla University, Kho Hong, Hat Yai District, 90110 Songkhla, Thailand

⁶Department of Psychiatry, Faculty of medicine, Prince of Songkhla University, Kho Hong, Hat Yai District, 90110 Songkhla, Thailand

⁷Department of Emergency Medicine, Faculty of medicine, Prince of Songkhla University, Kho Hong, Hat Yai District, 90110 Songkhla, Thailand

Received: 21 August 2023 / Accepted: 22 July 2024

Published online: 06 August 2024

References

1. Astin AW. Assessment for excellence: the philosophy and practice of assessment and evaluation in higher education. Rowman & Littlefield; 2012.
2. Frank JR, Mungroo R, Ahmad Y, Wang M, De Rossi S, Horsley T. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. 2010;32(8):631–7.
3. Ng IK, Mok SF, Teo D. Competency in medical training: current concepts, assessment modalities, and practical challenges. *Postgrad Med J*. 2024;cgae023.
4. Skochelak SE. Health systems science. Elsevier Health Sciences; 2020.
5. Gonzalo JD, Dekhtyar M, Starr SR, Borkan J, Brunett P, Fancher T, et al. Health systems science curricula in undergraduate medical education: identifying and defining a potential curricular framework. *Acad Med*. 2017;92(1):123–31.
6. Lewis A, Smith D. Defining higher order thinking. *Theory into Pract*. 1993;32(3):131–7.
7. Schraw G, Robinson DH. Assessment of higher order thinking skills. 2011.
8. Adedoyin OB, Soykan E. Covid-19 pandemic and online learning: the challenges and opportunities. *Interact Learn Environ*. 2020;1–13.
9. García-Peñalvo FJ, Corell A, Abella-García V, Grande-de-Prado M. Recommendations for mandatory online assessment in higher education during the COVID-19 pandemic. *Radical solutions for education in a crisis context*. Springer; 2021. pp. 85–98.
10. Pokhrel S, Chhetri R. A literature review on impact of COVID-19 pandemic on teaching and learning. *High Educ Future*. 2021;8(1):133–41.
11. Khan RA, Jawaid M. Technology enhanced assessment (TEA) in COVID 19 pandemic. *Pakistan J Med Sci*. 2020;36(COVID19–S4):S108.
12. Fuller R, Joynes V, Cooper J, Boursicot K, Roberts T. Could COVID-19 be our 'There is no alternative'(TINA) opportunity to enhance assessment? *Med Teach*. 2020;42(7):781–6.
13. Hazen H. Use of oral examinations to assess student learning in the social sciences. *J Geogr High Educ*. 2020;44(4):592–607.
14. Huxham M, Campbell F, Westwood J. Oral versus written assessments: a test of student performance and attitudes. *Assess Evaluation High Educ*. 2012;37(1):125–36.
15. Memon MA, Joughin GR, Memon B. Oral assessment and postgraduate medical examinations: establishing conditions for validity, reliability and fairness. *Adv Health Sci Educ*. 2010;15:277–89.
16. Chiang YC, Lee HC, Chu TL, Wu CL, Hsiao YC. Development and validation of the oral presentation evaluation scale (OPES) for nursing students. *BMC Med Educ*. 2022;22(1):318.
17. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:14065823*. 2014.
18. Lüdtke D. Ggeffects: tidy data frames of marginal effects from regression models. *J Open Source Softw*. 2018;3(26):772.
19. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J. Package 'corplot' *Stat*. 2017;56(316):e24.
20. Gamer M, Lemon J, Gamer MM, Robinson A. Kendall's W. Package 'irr'. Various coefficients of interrater reliability and agreement. 2012;22:1–32.
21. Jonsson A, Svingsby G. The use of scoring rubrics: reliability, validity and educational consequences. *Educational Res Rev*. 2007;2(2):130–44.
22. Reddy YM, Andrade H. A review of rubric use in higher education. *Assess Evaluation High Educ*. 2010;35(4):435–48.
23. Brookhart SM, Chen F. The quality and effectiveness of descriptive rubrics. *Educational Rev*. 2015;67(3):343–68.
24. Dawson P. Assessment rubrics: towards clearer and more replicable design, research and practice. *Assess Evaluation High Educ*. 2017;42(3):347–60.
25. Musick DW. Policy analysis in medical education: a structured approach. *Med Educ Online*. 1998;3(1):4296.
26. Akimov A, Malin M. When old becomes new: a case study of oral examination as an online assessment tool. *Assess Evaluation High Educ*. 2020;45(8):1205–21.
27. Maor R, Levi R, Mevarech Z, Paz-Baruch N, Grinshpan N, Milman A et al. Difference between zoom-based online versus classroom lesson plan performances in creativity and metacognition during COVID-19 pandemic. *Learning Environ Res [Internet]*. 2023 Feb 21 [cited 2023 Aug 22]; <https://link.springer.com/https://doi.org/10.1007/s10984-023-09455-z>
28. Murillo-Zamorano LR, Montanero M. Oral presentations in higher education: a comparison of the impact of peer and teacher feedback. *Assess Evaluation High Educ*. 2018;43(1):138–50.
29. Yen WM, THE CHOICE OF SCALE FOR EDUCATIONAL MEASUREMENT: AN IRT PERSPECTIVE. *J Educational Meas*. 1986;23(4):299–325.
30. Ndiaye Y, Lim KH, Blessing L. Eye tracking and artificial intelligence for competency assessment in engineering education: a review. *Front Educ [Internet]*. 2023 Nov 3 [cited 2024 Jul 14];8. <https://www.frontiersin.org/journals/education/articles/https://doi.org/10.3389/educ.2023.1170348/full>

31. Khan S, Blessing L, Ndiaye Y. Artificial Intelligence for Competency Assessment in Design Education: A Review of Literature. In: Chakrabarti A, Singh V, editors. *Design in the Era of Industry 4.0, Volume 3* [Internet]. Singapore: Springer Nature Singapore; 2023 [cited 2024 Jul 14]. pp. 1047–58. (Smart Innovation, Systems and Technologies; vol. 346). https://link.springer.com/https://doi.org/10.1007/978-981-99-0428-0_85

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.