Reliability and Validity of the Multiple Mini Interview in the General Entrance Examination for Medical Schools at a National University

Koji Tsunekawa [*], Michiyo Nakashima [*],
Toshiki Shioiri [*]

Abstract

Introduction: Although MMI is considered an interview technique with good reliability and validity in foreign countries, few universities have implemented it in admissions because of its high cost in Japan. In this study, we examined the reliability and validity of the MMI at a medical education institution that introduced the MMI for general entrance examinations. Methods: Twenty-nine students who took the MMI in 2022 were included in the study. For reliability, G study and D study were conducted based on generalizability theory. For validity, the total score of the MMI and the total score of the achievement test were plotted as a scatter plot and compared with the currently known status. Results: G coefficient was 0.336. There was no significant correlation between MMI and test scores. Unwilling applicants had lower MMI scores. Discussion and Conclusion: Regarding reliability, we found that our MMI needs to be improved. The validity of the MMI needs further investigation, but the results suggest that the MMI may be related to factors important to university administration.

Keywords: Multiple Mini Interview (MMI), Generalizability theory (G-theory), medical education, Institutional Research (IR)

1   Introduction

In recent years, entrance examinations at universities have become more diverse, but the most common method of entrance examination still only measures one aspect of knowledge, and at one point in time, academic outcome. In particular, there are few developed countries other than Japan that use this type of selection method in the selection of medical professionals [1]. In Japan, there are currently 82 medical schools that train medical students to become doctors[2], and all the medical schools use interviews in the general entrance examinations. Interview strategies used in admissions include individual interviews, group interviews, and group discussions, but these interview strategies are generally considered to lack reliability and validity [3-4].

Hence, Multiple mini interviews (MMI) is an interview strategy that has been gaining attention in recent years. MMI was developed at the McMaster University Medical School in 2001 and simply put, it is a 'circulating multiple interview such as Objective Structured Clinical Examination (OSCE)' [5]. In detail, firstly the examinee reads an question in which a prescribed situation is written. Next, the examinee considers the situation in the setting time. Then, they present their answer to the interviewer within another setting time. In some cases, the interviewer may ask additional questions, which the examinee answers further. By having examinees solve

---

[*] Gifu University School of Medicine, Gifu, Japan

multiple tasks, a variety of abilities that cannot be measured at once in a single interview can be assessed.

In Japan, MMI has been used in recent years in entrance examinations at several medical professions training schools [6] and in the admission examination for resident physicians [7-9]. In general entrance examinations of medical schools, according to the authors' research, MMI has been introduced at three schools. There have been several publications on MMI in Japanese entrance examinations, but there are few reports of MMI in entrance examinations at training hospitals in English[7-9], and there are no English papers reporting MMI in entrance examinations of medical school.

Generalizability theory is a method for examining reliability of interviews. It divides the sources of error into multiple sources, utilizes the method of analysis of variance to estimate the magnitude of each variable, and uses the information on the influence of the variance component of each variable estimated to enable the formulation of implementing an appropriate assessment plan [10,11]. Generalizability theory consists of Generalizability Study (G Study) and Decision Study (D Study). G study is used to examine the impact of each variable factor and their interaction on evaluation by focusing on the measurement error that occurs in measurement such as evaluation and estimating the components of the variable factors associated with the measurement and the magnitude of their variation (variance components), which is the cause of the measurement error. On the other hand, D study uses the estimated values of variance components obtained in the G study to calculate a generalizability coefficient, which corresponds to the usual reliability coefficient (alpha coefficient) [12], and simulates how many questions and raters are needed to obtain a sufficient generalizability coefficient to help improve the assessment [13, 14]. Thus, a major advantage of generalizability theory is that it takes into account improvements in valuation.

Therefore, in this study, we examined the MMI administered in the general entrance examinations of a Japanese national university, examining its reliability based on generalizability theory and validity through correlations with other factors such as academic test and early dropout.

2　Methods

A.　Participants and Settings

Twenty-nine examinees who took the general entrance examination (later scheduled entry) of the G University School of Medicine in 2022 were included in this study. The MMI had four stations, and examinees visited each station according to the nested design [7, 15]. Two raters per each station assessed competence on an assignment based on the admission policy. The raters at each station rated the examinees using a 4-point Likert scale, and the analyst calculated the examinees' total score assessed by the raters. The MMI assessments were not used as part of the score distribution for the entrance exam itself, but only used as a source of overall judgment and no examinees were rejected by their MMI scores.

B.　Data analysis

First, to verify the reliability on MMI, we performed multivariate Generalizability (G) and Decision (D) studies. The MMI scores were analyzed with mGENOVA software (Version. 2.1) and G-String IV [11].

The multivariate model for each format was:

e · ×(r · : c ·)

e- examinee, r- raters, c- cases, •- ratings (the fixed facet)

The ratings were considered as a fixed effect, since the rating scales were considered as the universe under consideration, and were used in all stations. Hence, the generalization over ratings was not required.

Next, to test the validity of the MMI, we plotted the calculated total MMI scores and the total scores of the academic achievement tests as a scatter plot and analyzed Pearson's correlation coefficient. We also examined accepted and declined applicants and those who suspend from school immediately after admission (so-called 'masqueraders': students who are trying to get into their preferred university while already enrolled at another one), and matched them with the entrance examination data. SPSS ver. 23.0 Japan for Windows (SPSS Inc., Chicago, IL, USA) was used to perform statistical test. Two-tailed p-values of <0.05 were considered significant.

3 Results

The mean age of the 29 examinees was 19.8 years (range 18-42). Of the 29 examinees, 19 (66%) were male and 10 (34%) were female. The mean ± Standard Deviation (SD) scores of MMI for applicants were 20.31 ± 4.39 , and those of academic test were 1149.95 ± 133.26.

A. Reliability

The variance estimates are presented in Table 1. The variance of examinee-case interaction (see the set of rows for 'ec' in the 'effect' column) was the largest source of variance. And the variance of examinees into examinees within cases and random error (see the set of rows for 'er:c' in the 'effect' column) was the second largest. The variance of cases (see the set of rows for 'c' in the 'effect' column) and the variance of raters within cases (see the set of rows for 'r:c' in the 'effect' column) were relatively small, indicating that there was no substantial case difficulty variation, or inter-rater variability (including the issue of stringency/leniency), achieved by intensive case structuring process comprising: an established competency framework; standardized question types; standardized assessment rubrics with anchored rating scales; two independent raters per examinee; and intensive rater training [7].

Table1: Variance Components

Effect	df	T	SS	MS	VC
e	7	14.5	14.5	2.07143	0.08705
c	3	3.375	3.375	1.125	-0.0067
r:c	4	3.75	0.375	0.09375	-0.01786
ec	21	46.75	28.875	1.375	0.5692
er:c	28	53.75	6.625	0.23661	0.23661

Abbreviations: df- Degrees of Freedom, T- Total variance, SS- Sum of Squares, MS- Mean Squares, VC- Variance Components, e- Examinee, c- Case, r- Rater, r:c- Raters within Cases, ec- Examinees into Cases, er:c- Examinees into Examinees within Cases, and random error.

The multivariate G analyses demonstrated that the G-coefficient was 0.336. The D-study indicated that about 30 stations, each manned by two examiner would provide acceptable reliability (Figure 1).
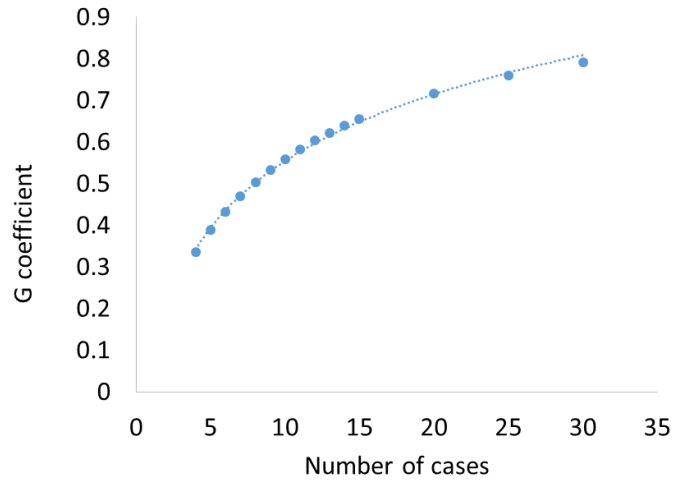

Figure 1: D-study of MMI

B. Validity

Figure 2 shows a scatter plot with the total score of the academic achievement test on the horizontal axis and the total score of the MMI on the vertical axis. The Pearson's correlation coefficient between the total score of academic achievement test and the total score of MMI was $r = -0.172$, $p = 0.266$, showing no significant correlation. The 11 students with the highest total score of the achievement test were treated as accepted (▲), and two of them declined enrolment or suspend from school immediately after admission (Arrow).
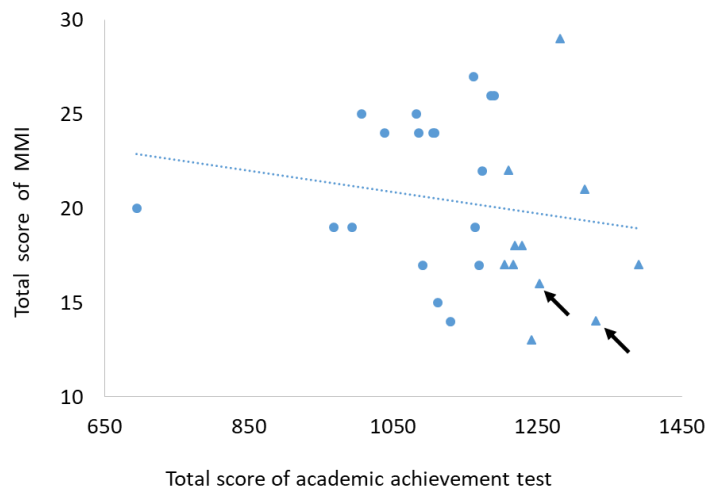

Figure 2: Scatter plot with the total score on the academic achievement test and MMI

4   Discussion

MMI is considered to be more reliable and valid than traditional interviewing methods such as individual interviews, group interviews, and group discussions; however [4, 16], the number of evaluators, the room to be used, and the time required are more than those of these methods [17]. Generally, the number of stations required to design a reliable MMI with one rater per station is 7 or more [18], but the MMI at the University of G in this study has only 4 stations and a low G-coefficient   (0.336), even though there are 2 raters in each station. This is due to the lack of rooms acceptability of the raters to having one rater per station, and the time limitations.

Furthermore, in the D-study, it showed that a G-coefficient of more than 0.8 requires more than 30 stations when there are two raters in a station and even more than 35 stations even when there is one rater in a station (data not shown). These results suggest that there are issues in the coordination of the rater's assessments before MMI and in the assessment sheet during MMI. and that faculty development, improvement of the evaluation form, and creation of an evaluation manual for MMI evaluation are needed in the future. Future work includes faculty development of the assessment of MMI for raters, improvement of the assessment sheet, and the creation of an evaluation manual and FAQs.

On the other hand, with regard to validity, as shown in Figure 2, no significant correlation was obtained with the scores of the academic achievement tests. This result is consistent with that of a previous studies [18, 19], suggesting that the MMI measures competencies which is different from knowledge. Furthermore, the low scores on the MMI for involuntary enrollees, such as those who declined enrolment and suspend from school immediately after admission, reveal that incorporating the MMI into total scores of entrance examination may prevent the admission of involuntary enrollees. Previous studies have reported that the MMI is useful as a predictor of OSCE performance and national medical examinations [19, 20].

As a whole, these researches show interesting preliminary results. However, due to the lack of enough data such as subsequent outcomes, further study is needed to lead to robust conclusions.

5   Conclusions

The MMI at G University School of Medicine needs further improvement in reliability, while the validity of the MMI is tentatively considered good.

Notes

1. OSCE is a proficiency test to measure skill and attitude at third or fourth grade medical students. It comprises six to nine stations. Raters assess student performance at each station, and the judgment is based on a 6-point Likert-type scale. In Japan, This test is controlled by Common Achievement Tests Organization (CATO).

Acknowledgement

References

[1] Moradi, N., Mohammadi, R., & Goldasteh, A. (2019). A Comparative Study of Higher Education Entrance Examinations in Iran with Some Selected Countries. Iranian Journal of Comparative Education, 2(4), 518-532.

[2] Nogami, T., Arai, M., Ishigami, T., Nakada, Y., Matsuda, T., Odaguchi, H., & Hanawa, T. (2021). Comparison of the 2011 and 2019 Kampo Medicine Curricula Across All Japanese Medical Schools. The Tokai Journal of Experimental and Clinical Medicine, 46(2), 75-82.

[3] Benbassat, J., & Baumal, R. (2007). Uncertainties in the selection of applicants for medical school. Advances in Health Sciences Education, 12(4), 509-521.

[4] Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. Medical education, 50(1), 36-60.

[5] Eva, K. W., Rosenfeld, J., Reiter, H. I., & Norman, G. R. (2004). An admissions OSCE: the multiple mini-interview. Medical education, 38(3), 314-326.

[6] Kishi T. (2017). The relationship between Multiple Mini Interview and academic performance of medical students. Tne 81st annual convention of the Japanese Psychological Association pp. 2D-080 (written in Japansese).

[7] Yoshimura, H., Kitazono, H., Fujitani, S., Machi, J., Saiki, T., Suzuki, Y., & Ponnamperuma, G. (2015). Past-behavioural versus situational questions in a postgraduate admissions multiple mini-interview: a reliability and acceptability comparison. BMC medical education, 15(1), 1-9.

[8] Yamada, T., Sato, J., Yoshimura, H., Okubo, T., Hiraoka, E., Shiga, T., ... & Ban, N. (2017). Reliability and acceptability of six station multiple mini-interviews: past-behavioural versus situational questions in postgraduate medical admission. BMC Medical Education, 17(1), 1-7.

[9] Yamada, T., Hiraoka, E., & Takemura, Y. (2021). Predictive Validity of Past-Behavioral Versus Situational Questions in Postgraduate Medical Admissions. Academic Medicine, 96(2), 164-165.

[10] Brennan, R. L. (2000). Performance assessments from the perspective of generalizability theory. Applied Psychological Measurement, 24(4), 339-353.

[11] Streiner, D. L., Norman, G. R., & Cairney, J. (2015). Health measurement scales: a practical guide to their development and use. Oxford University Press, USA. 200-225.

[12] Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International journal of medical education, 2, 53.

[13] Shavelson, R.（1991）. Webb, N. Generalizability Theory ; A Primer. Sage Publications, Thousand Oaks, CA.

[14] Brennan, R.L. （2005）. Generalizability theory, New York : Springer Germany.

[15] Oliver, T., Hecker, K., Hausdorf, P. A., & Conlon, P. (2014). Validating MMI scores: are we measuring multiple attributes?. Advances in Health Sciences Education, 19(3), 379-392.

[16] Salvatori, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. Advances in Health Sciences Education, 6(2), 159-175.

[17] Rosenfeld, J. M., Reiter, H. I., Trinh, K., & Eva, K. W. (2008). A cost efficiency comparison between the multiple mini-interview and traditional admissions interviews. Advances in Health Sciences Education, 13(1), 43-58.

[18] Rees, E. L., Hawarden, A. W., Dent, G., Hays, R., Bates, J., & Hassell, A. B. (2016). Evidence regarding the utility of multiple mini-interview (MMI) for selection to undergraduate health programs: A BEME systematic review: BEME Guide No. 37. Medical Teacher, 38(5), 443-455.

[19] Pau, A., Jeevaratnam, K., Chen, Y. S., Fall, A. A., Khoo, C., & Nadarajah, V. D. (2013). The multiple mini-interview (MMI) for student selection in health professions training–a systematic review. Medical teacher, 35(12), 1027-1041.

[20] Kelly, M. E., Dowell, J., Husbands, A., Newell, J., O'Flynn, S., Kropmans, T., ... & Murphy, A. W. (2014). The fairness, predictive validity and acceptability of multiple mini interview in an internationally diverse student population-a mixed methods study. BMC Medical Education, 14(1), 1-13.