ORIGINAL ARTICLE

LEGO®-based communication assessment in virtual general surgery residency interviews

Mohamed S. Baloul[1] · Sarah Lund[1,2] · Jonathan D'Angelo[3] · Vicky J.-H. Yeh[1] · Nizamuddin Shaikh[1] · Mariela Rivera[4]

Abstract

Purpose Effective communication skills are a critical quality and skill that is highly sought after for surgeons which largely impacts patient outcomes. Residency programs design their interview processes to select the best candidates. LEGO®-based activities have been frequently used to enhance communication skills and team building. This study investigates the effectiveness and reliability of a novel LEGO®-based communication assessment in interviews for surgical residencies and the feasibility of implementing it in a virtual setting.

Methods This study conducted a retrospective analysis of a LEGO®-based communication assessment at the program's 2020/2021 residency interviews. Each applicant was assessed on a different model. The total scores were analyzed for consistency among raters and correlated to faculty interviews. Furthermore, the impact of the assessment structure, scoring criteria, and range of models' difficulties on the total scores were explored.

Results A total of 54 categorical and 55 preliminary applicants interviewed on 2 days. The assessment on different models and had no impact on applicants' total scores for either categorical and preliminary groups ($p = 0.791$ and 0.709, respectively). The communication components of the assessment showed high consistency between the raters. The two applicant groups displayed a statistically significant difference ($p = 0.004$) in the communication evaluation and model accuracy components. Total scores did not correlate with the faculty interviews of standardized questions in either group.

Conclusion This novel LEGO®-based communication assessment showed high reliability and promising results as a tool to assess communication and problem solving for residency interviews that can be readily implemented in a virtual setting.

Keywords Assessments · Simulation · Interviews · Communication

Abbreviations

CAT	Categorical
PREL	Preliminary
PD	Program director
ACGME	Accreditation council for graduate medical education
NRMP	National resident matching program
MMI	Multiple mini interviews
GS	General surgery
OSCE	Objective structured clinical examination
ICC	Interclass correlation coefficient
ANOVA	Analysis of variance

✉ Mariela Rivera
rivera.mariela@mayo.edu

Mohamed S. Baloul
Baloul.Mohamed@mayo.edu

Sarah Lund
Lund.Sarah@mayo.edu

Jonathan D'Angelo
DAngelo.Jonathan@mayo.edu

Vicky J.-H. Yeh
Yeh.Vicky@mayo.edu

Nizamuddin Shaikh
Shaikh.Nizamuddin@mayo.edu

1. Department of Surgery, Mayo Clinic, Rochester, MN, USA

2. Mayo Clinic Multidisciplinary Simulation Center, Rochester, MN, USA

3. Division of Colon and Rectal Surgery, Mayo Clinic, Rochester, MN, USA

4. Division of Trauma, Critical Care, and General Surgery, Department of Surgery, Mayo Clinic, 200 1st ST SW, Rochester, MN 55905, USA

Background

Interviews play a crucial role in selecting candidates for surgical residencies [1]. This one-on-one time with a candidate can allow program directors (PD) and faculty an opportunity to assess a candidate's residency compatibility, personal characteristics, interest in the program and communication skills [2–4]. This last skill is a particularly important assessment for PDs [5], as strong communication skills are a nearly universally accepted quality of a safe and efficient surgeon [2, 6–10]. Effective physician–patient communication has been repeatedly shown to positively correlate with health outcomes and patients' adherence to treatments [11, 12]. Additionally, communication skills are assessed as one of the Accreditation Council for Graduate Medical Education's (ACGME) milestone evaluations [13]. Furthermore, the 2021 National Resident Matching Program (NRMP) PD Survey listed interpersonal skills as the highest rated personal characteristic in all specialties when ranking the applicants [5]. Applicants' personal characteristics such as attitude, interpersonal relationships, motivation and problem solving have shown to be helpful in predicting residents' clinical performances [14].

For over a decade, LEGO®-based activities have been used for team building and communication assessments. Its most popular activities rely on communication between a team to replicate a model such as LEGO® construction [15] and LEGO® Serious Play [16]. These activities inspire creativity, effective communication, problem solving and decision making. A study by Bethel et al. suggests that the use of LEGO®-based activities in urology interviews allows for a more standardized assessment of ACGME competencies of interpersonal skills and communication where it can identify poor performers [17]. Utilizing LEGO® in a communication activity may not only provide a more structured evaluation tool for candidates but may also give deeper and generalizable insights [17, 18].

While many programs were forced into virtual interviews due to the COVID-19 pandemic, many programs are leaning toward keeping the virtual interview format into the future where it is deemed to be a viable and effective alternative to in-person interviews [19–23]. Moreover, virtual communication has long been demonstrated to allow individuals to communicate more selectively, resulting in more positive interactions [24]. The inclusion of a communication assessment in a virtual interview may be uniquely insightful as a tool that can assess candidates' skills in a standardized manner free of some of the bias and resources that virtual interviews allow [19–21]. However, literature remains scarce on the practicality and effectiveness of LEGO®-based communication assessments for residencies, especially in virtual settings.

In this pilot study, we sought to investigate the feasibility, effectiveness, and reliability of a novel LEGO®-based assessment to evaluate applicants' communication skills in general surgery (GS) residency interviews and the viability to be implemented in a virtual setting.

Methodology

This study was conducted retrospectively on de-identified data of a novel LEGO®-based communication assessment that was part of the GS residency program's 2020/2021 virtual interviews [25]. The Institutional Review Board deemed the study exempt. This communication assessment was developed to evaluate 4 aspects of communication: (1) Clarity of communication, (2) Mutual understanding, (3) Adaptability and flexibility, and (4) Professional attitude. Concurrently, we aimed to briefly assess the problem solving and critical thinking of the applicants.

Interviews were split into two days, one day each for each group of preliminary (PREL, $n = 55$) and categorical (CAT, $n = 54$) positions. Groups were further split into smaller subgroups of up to 16 applicants each that rotated between the different segments of the interview day (e.g., faculty interviews and standardized questions, skills assessments, virtual tours, etc.). All interviews were conducted virtually through a live-video platform.

To ensure effectiveness and test for validity, validation concepts and theories were taken into consideration when designing the assessment [26]. For content validity, the assessment design was carefully planned and developed, testing and trialed several times. For internal structure, construct and consequences validity [27], inter-rater reliability was taken into consideration. Detailed descriptions are further elaborated below.

Applicant preparation

Instructions for the assessment were mailed out as part of the information packet about the interview day. Online access to the instructions was also available. The instructions for this communication assessment included a reference card with official LEGO® terminologies [Appendix A—LEGO® index card]. This provision of standardized terminology to applicants was an attempt to reduce the cognitive burden of word selection, especially for applicants whose first language is not English. Applicants were allowed to use the card during the assessment. Applicants were given further instructions and details on timeline and process of the assessment, setup and expectations of the assessment at the commencement of the interview day. No preparation was mandated.

Assessment setup and design

A moderator, one builder and two assessors were all seated in the same room at our Simulation Center. The two assessors were placed on a table next to each other with one laptop each. One assessor greeted and introduced the assessment to all applicants, while the other assessor managed sharing the screen and moving through the assessment. The builder was placed opposite to the assessors and computers on the same table. That layout prevented the builder from seeing the correct form of the structure to be built, and provided the assessors with visibility to the model while it is being built. A speaker microphone was placed between the assessors and builder and was used for communicating with the applicant (Appendix B—Room Layout & Design). Cost-wise, the setup did not require any special tools or equipment. Since only 15 Lego pieces were used, they were obtained from members of the team. A second replica set of the pieces was kept as backup. The moderator was seated on a separate table to the side of the assessors. The moderator was trained to resolve common technical issues related to video calls and screen sharing. The moderator was on all video calls but entirely off-camera, tasked with strictly keeping track of time, resolving issues that may affect an applicant's assessment, and staying in communication with the interview day tech support team. The room setup allowed for easy communication between the team and facilitated the moderator's task if any issues were to rise. No issues were noted.

During the assessment on the interview day, the applicants were tasked to instruct a builder to create a replica of a LEGO® model in 3 min. At the assessment, the instructions were read to the applicants by one of the assessors, they were then screenshared four pictures of the model—seeing it for the first time and allowed to see it for 20 s right before the start of the 3-min assessment time. The first picture was of the individual pieces comprising the model. The other three pictures were of the completed model, a top-down view and two side views (Fig. 1). Applicants were not able to see the builder assembling the model, but they could request to see the builder's attempt once sometime during the 3-min assessment. A camera of a view of the current build was turned on briefly (10 s) for when the applicant requested to see the built model. This allowed applicants to check their progress and demonstrate problem-solving abilities, as seeing the build often gave the applicants insight into problems with their communication.

The builder was blinded to the correct final forms of the models. The builder was not allowed to proactively ask the applicant any questions. However, if the builder did not
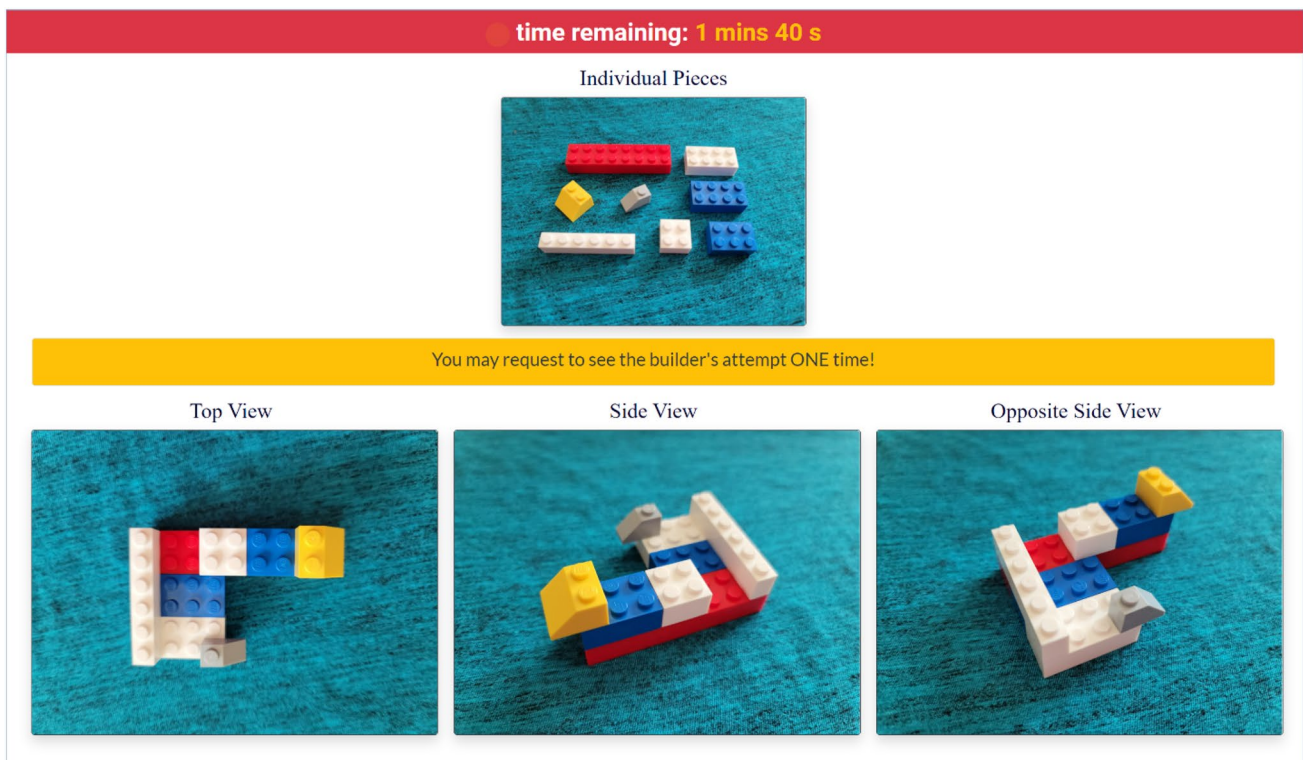
Fig. 1 A depiction of the shared view with the applicant during the assessment. Each applicant was shown the individual pieces used for that model, and three views of the constructed model

understand a certain point, they were encouraged to ask: "Can you clarify?". The builder was instructed to answer all of the applicant's questions. The builder never appeared on screen.

Scoring

The scoring items were selected and constructed based on several factors: (1) Communication skills our institution's program leadership desired in residents, (2) ACGME milestones (Interpersonal and Communication Skills 2: Interprofessional and Team Communication), [13], and, (3) Results from previous studies related to communication Multiple Mini Interviews (MMI) [28]. Two assessors scored applicants' performance using a 5-point agreement scale. Two assessors had visibility to the correct form of the model and were asked to rate the applicants on the four aforementioned aspects of communication and strategy [Appendix C—Scoring rubrics].

The builder was tasked to assess 2 components: (1) Communication effectiveness, and (2) Problem evaluation [Appendix—the scoresheet]. The total score and distribution are shown in Fig. 2. Problem Evaluation was present on the builder's scoring sheet for both interview days and not on the assessors scoring sheet. On the second day (PREL group), Problem Evaluation was added to the assessors' scoring sheet, based on feedback from the moderator and assessors. That change was made to introduce more objectivity by including the observing assessors in assessing that component since the builder had no visibility to the correct form of the model. Problem Evaluation was then calculated as an average of the builder and assessors' problem evaluation scores for PREL, while it was part of the builder's overall assessment for the CAT group.

At the end of each assessment, the assessors would objectively assess the accuracy of the applicant's final build (by counting the number of incorrect pieces used and the number of incorrect placements), i.e., the Model Accuracy Score. The conversion scale for the incorrect pieces was: 0 incorrect = 5 points; 1–2 incorrect = 4 points; 3–4 incorrect = 3 points; 5–6 incorrect = 2 points; 7 incorrect = 1 point.

The builder and the assessors were all first year GS residents who underwent this LEGO® communication assessment in the weeks prior to the interview day as part of their residency simulated skills curriculum. Multiple mock-up runs were conducted in the weeks leading to the interview days to familiarize the moderator, assessors, and builder with the how to conduct and score the assessment. During the trial runs, the assessors were trained on the scoring rubric by skilled raters, who explained the subcomponents of the rubric. Assessors participated in practice scoring sessions during the trial runs for the interview day, at which point any questions or concerns from the assessors were answered and clarified. All builders participated in practice runs prior to the interview day. The builders were trained on the official terminologies as well as different descriptors that may be used by the applicants [Appendix A—LEGO® index card].

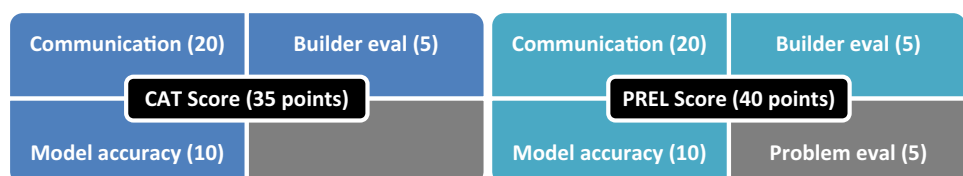## Pieces selection and model generation

To meet the assessment objectives within the 3 min allocated to this test, and after multiple trials by the study team, it was established that the most feasible construction would be to limit the model to 8 pieces composed of 3 vertical layers (base, middle, and top). Unusual pieces, defined as pieces with an irregular shape that may require significant time to describe, were excluded. To aid in color identification, contrasting colors were chosen, and any pieces in similar colors but varying shades were excluded. To accommodate for applicants with color-deficiency, all green pieces were excluded (the inability to differentiate between green and red is the most common color blindness—Protanopia and Deuteranopia) [29]. Based on the above criteria, a total of 15 pieces were selected—from which 8 would only be used at a time. The 7 extra pieces were to require the applicant to select the correct pieces and to allow a larger set of combinations.

A total of 40 LEGO® models were generated through custom-made software. The software was set to randomly select 8 out of the 15 pre-selected pieces for each model. The piece selection for each layer involved 3–4 pieces for the base, 3–4 for the middle layer, and a maximum of 2 on the top layer. Larger pieces always appeared on the lower layers.

Model difficulty selection

Out of the 40 generated LEGO® models, we selected a pool of 20 models, meaning that on a given assessment day, every model was used twice but no model was used more than 3 times. This was to balance several considerations: (1) Builder's memory: to reduce recall bias—the chance of the builder recognizing repeated models; (2) Similar difficulty level: the

Fig. 2 Score distribution and weights for each of the components. To the left is for Day 1—categorical (CAT). Day 2—preliminary (PREL) on the right

Communication (20)	Builder eval (5)	Communication (20)	Builder eval (5)
CAT Score (35 points)		**PREL Score (40 points)**	
Model accuracy (10)		Model accuracy (10)	Problem eval (5)

lower the number of different models, the lower the variability in difficulties, hence, higher consistency. This ensures fairness among applicants; (3) Model-reliability testing: To test the reliability of models by introducing them more than once, we would be able to better understand if lower/higher score is likely due to applicant's actual performance or due to the model itself (e.g., if the majority of applicants on that model perform well, low performers are likely true low performers; but if all applicants perform differently from each other on a particular model, that may indicate a higher difficulty of the model or individual variability).

In a modified Delphi process, an initial set of 40 models was sent to a team of 8, comprised of surgical educators and GS first year residents who had previously participated in the LEGO® communication assessment. For the first round, the team was asked to independently rank the models based on difficulty on a 5-point scale (1 = easy, 3 = medium, 5 = hard). For each model, the mean of the ratings was set to be the difficulty level. The closest 20 ratings to the midpoint (equidistant lower and upper range) were selected (total of 22 models). The second round consisted of the tied models at the lower (easier) and upper (harder) ends to eliminate the easiest and the hardest ones, respectively. In the end, 20 models (difficulty ranged from 2.8 to 3.2) were selected for the interviews.

Each applicant received a model randomly selected from the final pool of 20. For the assessment, the sequence at which the models appeared for the applicants was randomized. Additionally, it was ensured that no models would repeat until all 20 were exhausted. Models that appear at the ends of the sequences were set be spaced at least 8 positions from their last appearance.

Statistical analysis

Descriptive and statistical analyses were done using Microsoft Excel 365 (Microsoft Corporation—Redmond, Washington). Interclass Correlation Coefficient (ICC) estimates to test rater agreement (mean-rating: $k = 2$ per assessment day, absolute agreement, 2-way random-effects model), and Spearman's Rank-Order Correlations were calculated using BlueSky Statistics v7.2 (BlueSky Statistics LLC—Chicago, Illinois). Analysis of Variance (ANOVA) was conducted to test the relationship between the model difficulty (independent variable) and total score (dependent variable) for each applicant group. A two-sample t test was used to test the differences in "Communication" component scores between the CAT and PREL groups. A Wilcoxon–Mann–Whitney test was used to test differences between male and female applicants.

Results

The GS interviews consisted of 54 categorical and 55 preliminary applicants. The range of the total LEGO® communication assessment scores for the CAT group was 19.5–34.5 (total score = 35), and 18.5–40 (total score = 40) in the PREL group. Figure 3 shows the distributions of the total scores for each group. Although the PREL group resulted in a normal distribution, the CAT group was negatively skewed. The skewness (Pearson's moment coefficient) for the CAT and PREL groups were − 0.48 and 0.15, respectively. While the p-values for the Shapiro–Wilk Normality test were 0.003 and 0.356, respectively.

A t test of the communication evaluation scores for the CAT and PREL groups (means of 17.3 and 15.8, respectively) displayed a statistically significant difference between the groups (p value: 0.004). A breakdown of the averages in performances in each component of the score sheet is summarized in Table 1.
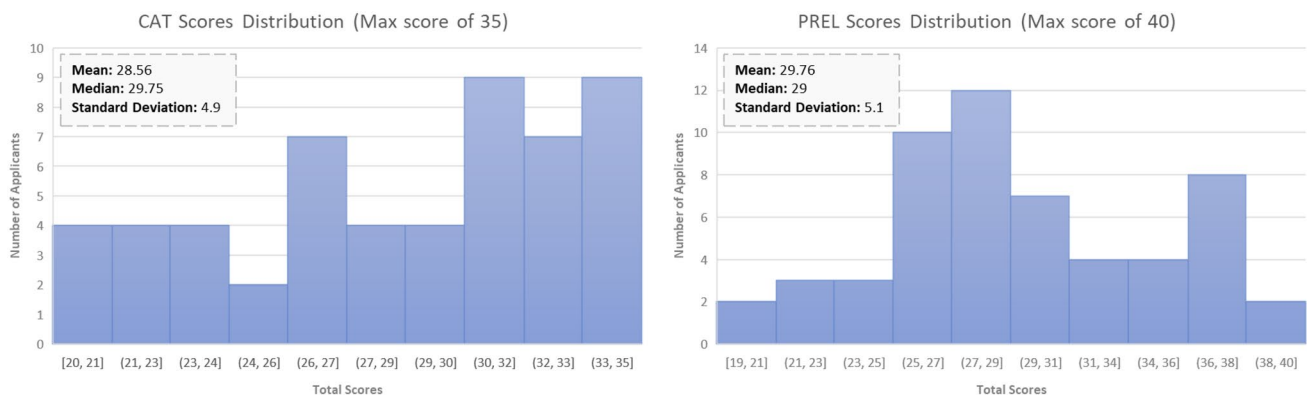
Fig. 3 Distribution of the total scores of the CAT and PREL applicants. Note that the max scores are different for each group. (*CAT* categorical, *PREL* preliminary)

Table 1 The distribution of applicants' scores on the communication assessment per component and per group

	CAT Mean ± SD	PREL Mean ± SD	*p* value
Communication evaluation (20)	17.3 ± 2.9	15.8 ± 2.56	0.004
Builder's evaluation (5)	3.7 ± 0.6	3.6 ± 0.83	†
Model accuracy score (10)	7.6 ± 1.4	6.9 ± 1.58	0.016
Problem evaluation (5)	–	3.5 ± 0.85	†

CAT categorical, *PREL* preliminary

†Not calculated due to scoring differences between CAT and PREL

Gender performance differences (Wilcoxon–Mann–Whitney test)

In the CAT group, 32 were females, and 22 males. For the PREL applicants 35 were males, and 20 females. No statistically significant difference was found between males and females in both the CAT and PREL groups (*p* values: 0.60 and 0.34, respectively).

Raters' interclass correlation of communication components

The results for the communication aspect of the assessment are summarized in Table 2. The majority of the communication components show a high ICC. Professional attitude ICC was low in both the CAT and PREL groups (0.33 and 0.53, respectively).

ANOVA analysis of model difficulty and the total score

When plotting applicants' total scores against the difficulties of the models they encountered, no statistical significance was found (*p* values: 0.791 and 0.709; R^2—Coefficient of determination: 0.001 and 0.003) for the CAT or PREL groups, respectively.

Correlations with faculty interviews and standardized questions

We calculated Spearman's rank-order correlation between the totals scores of the communication assessment to two interview metrics, namely, the faculty interviews and the standardized questions. The analysis demonstrated weak to no correlation of faculty interviews and standardized questions to the communication assessment scores for the CAT (− 0.09 and 0.07) and PREL groups (0.01 and 0.15), respectively.

The assessment did not display any statistically significant correlation with demographic information, nor with any other metrics such as Step 1 or 2 CK scores, number of publications, or honors.

Discussion

The aim of this study was to analyze the feasibility, effectiveness, and reliability of a newly designed LEGO®-based assessment of communication that can be conducted virtually. Overall, this communication assessment showed reliable results in the design of the assessment (model difficulties and selections), applicants' performance distributions, and a high consistency between the raters. Our findings shed light on the feasibility and viability of using this LEGO®-based method in assessing communication for interviews. Our tool may help residency programs further consider interpersonal communication skills in an efficient and standardized manner.

The assessment exhibited strong inter-rater reliability for the different communication components. The weakest ICC estimate was the "Professional Attitude" component, which assesses the applicants use of respectful and/or professional tone of voice, attitude, or body language. The aggregation of multiple elements into a single items may have caused a range restriction affecting the reliability [30, 31]. Requiring raters to assess multiple constructs with a single rating might have led to various foci for judgment and a more subjective rating [32]. Comparable conclusions were stated by Mazor

Table 2 The interclass correlation estimates of the two assessors on the communication component of the assessment (mean ratings, absolute agreement, 2-way random-effects model)

	CAT (*n* = 54)			PREL (*n* = 55)		
	ICC	CI	*p* value	ICC	CI	*p* value
Clear communication	0.94	(0.89–0.96)	< 0.00	0.82	(0.70–0.90)	< 0.00
Mutual understanding	0.91	(0.84–0.95)	< 0.00	0.76	(0.57–0.86)	< 0.00
Adaptability and flexibility	0.81	(0.67–0.89)	< 0.00	0.68	(0.21–0.85)	< 0.00
Professional attitude	0.33	(− 0.12–0.60)	0.064	0.53	(0.20–0.72)	0.001
Overall	0.92	(0.87–0.96)	< 0.00	0.82	(0.61–0.90)	< 0.00

CAT categorical, *PREL* preliminary, *ICC* interclass correlation, *CI* confidence interval (Lower bound–Upper bound)

et al. on their in-depth analysis of professionalism in objective structured clinical examinations (OSCE) [33]. Consequently, due to the complexity of rating this component, splitting "Professional Attitude" into separate components may yield improved reliability results. A systemic review by Knorr et al. raised the question if it is needed and "possible to measure several distinct constructs at one station"—when an overall assessment score could be sufficient [34].

To gauge the effectiveness of using this LEGO® communication assessment in evaluating the communications skills of residency applicants, we analyzed the scores distributions [35]. The PREL group displayed a normal distribution while the CAT group was negatively skewed, with more applicants scoring higher. This could be due to the CAT group being predominately composed of American Medical Graduates (AMG) [36, 37]. While reference cards for standard LEGO® descriptions and names were sent ahead of time to all applicants, AMGs might have had the advantage of being more familiar with LEGO®s and the descriptions than International Medical Graduates. Hence, AMGs may have found it easier to describe the pieces. Nonetheless, applicants were free to use any terms and were not bound by the index card. Furthermore, another important factor to consider was the use of different raters for scoring that may have influenced the scores.

Interestingly, the communication assessment did not correlate with faculty interviews and standardized questions. This may suggest that this assessment accesses a different domain of communication from those assessed in traditional in-person interviews. Such differences and poor correlations are discussed extensively in the literature being attributed to context specificity [38–40]. Expressly, expected similar individual traits are not static and are heavily dependent on the context within which the performance was tested [18]. Selection methods remain essential to predict the best candidates for medical training and future doctors [4]. Literature reveals that the adoption of multi-sample approaches like multi-station OSCEs or MMIs along with academic records and aptitude tests provide a more generalizable portrayal of applicants' abilities [18, 28, 41, 42]. Hay et al. goes further to emphasize that academic potential alone does always imply a higher achieving applicant more than the ability to succeed academically [42, 43]. This LEGO assessment was one of several that aimed to assess non-cognitive traits. Ultimately, aiming to increase the validity and fairness of the selection process [44]. Applicant scores on this assessment were ranked to quartiles, which accounted for one metric among others for ranking [42]. The program uses discretion on ranking low performers on any of the interview assessments.

Limitations of this study include that this assessment was developed and conducted at one residency program for a limited number of applicants ($n = 109$). The modification of

the test between the two days (CAT and PREL) as well as the replacement of the assessors have hindered the ability to compare the two groups' total performances. However, this change was made knowing that it would not affect the fairness of the assessment as the applicants in each group were competing for a residency position only with applicants from their own group. Furthermore, while we controlled the frequency and number of times a model appeared on a given assessment day, it is not entirely known if the builder's memory recollection on probable combinations of the models affected the results. Nevertheless, it was not revealed to the builder that models could appear more than once nor exposed to how accurate any applicant's attempt was. Lastly, given the short assessment time of the task, other factors like unfamiliarity with LEGO blocks or restrictions to the questions the builder could ask may have affected the end results positively or negatively.

In conclusion, communication is paramount in the medical field, and more so in surgical specialties [45]. Poor communication in the OR may jeopardize patient safety and affect cognitive loads [46, 47]. Fostering mutual understandings through communication and open dialog can promote a culture of safety and camaraderie [48]. This LEGO®-based communication assessment showed high reliability and promising results in differentiating between applicants as a tool to assess communication and problem solving that can be easily replicated in virtual settings. Future studies should be done to investigate the impact of the inclusion of communication tests in the residency selection process and their effects on clinical practice and performance.

Declarations

References

1. Kim RH, Gilbert T, Suh S, Miller JK, Eggerstedt JM. General surgery residency interviews: are we following best practices? Am J Surg. 2016;211(2):476-81.e3. https://doi.org/10.1016/j.amjsurg.2015.10.003.

2. Vaughan LA, Quick JA. Evidence-based selection of surgical residents. Surg Clin North Am. 2021;101(4):667–77. https://doi.org/10.1016/j.suc.2021.05.012.

3. Stephenson-Famy A, Houmard BS, Oberoi S, Manyak A, Chiang S, Kim S. Use of the interview in resident candidate selection:

a review of the literature. J Grad Med Educ. 2015;7(4):539–48. https://doi.org/10.4300/jgme-d-14-00236.1.

4. Cleland J, Dowell J, McLachlan J, Nicholson S, Patterson F. Identifying best practice in the selection of medical students. London: General Medical Council; 2012.

5. Program TNRM. Results of the 2021 NRMP Program Director Survey. The National Resident Matching Program; 2021. p. 149–55.

6. Hutul OA, Carpenter RO, Tarpley JL, Lomis KD. Missed opportunities: a descriptive assessment of teaching and attitudes regarding communication skills in a surgical residency. Curr Surg. 2006;63(6):401–9. https://doi.org/10.1016/j.cursur.2006.06.016.

7. Bonvicini KA, Perlin MJ, Bylund CL, Carroll G, Rouse RA, Goldstein MG. Impact of communication training on physician expression of empathy in patient encounters. Patient Educ Couns. 2009;75(1):3–10. https://doi.org/10.1016/j.pec.2008.09.007.

8. Gillespie BM, Chaboyer W, Longbottom P, Wallis M. The impact of organisational and individual factors on team communication in surgery: a qualitative study. Int J Nurs Stud. 2010;47(6):732–41. https://doi.org/10.1016/j.ijnurstu.2009.11.001.

9. Nakagawa S, Fischkoff K, Berlin A, Arnell TD, Blinderman CD. Communication skills training for general surgery residents. J Surg Educ. 2019;76(5):1223–30. https://doi.org/10.1016/j.jsurg.2019.04.001.

10. Chandawarkar RY, Ruscher KA, Krajewski A, Garg M, Pfeiffer C, Singh R, et al. Pretraining and posttraining assessment of residents' performance in the fourth accreditation council for graduate medical education competency: patient communication skills. Arch Surg. 2011;146(8):916–21. https://doi.org/10.1001/archsurg.2011.167.

11. Stewart MA. Effective physician-patient communication and health outcomes: a review. CMAJ. 1995;152(9):1423–33.

12. Zolnierek KBH, Dimatteo MR. Physician communication and patient adherence to treatment: a meta-analysis. Med Care. 2009;47(8):826–34. https://doi.org/10.1097/MLR.0b013e31819a5acc.

13. Education TACfGM: Surgery milestones. 2019. https://www.acgme.org/Portals/0/PDFs/Milestones/SurgeryMilestones.pdf?ver=2020-09-01-152718-110. Accessed 11 May 2021.

14. Brothers TE, Wetherholt S. Importance of the faculty interview during the resident application process. J Surg Educ. 2007;64(6):378–85. https://doi.org/10.1016/j.jsurg.2007.05.003.

15. Harding SR, D'Eon MF. Using a Lego-based communications simulation to introduce medical students to patient-centered interviewing. Teach Learn Med. 2001;13(2):130–5. https://doi.org/10.1207/S15328015TLM1302_8.

16. Dann S. Facilitating co-creation experience in the classroom with Lego serious play. Australas Marketing J (AMJ). 2018;26(2):121–31. https://doi.org/10.1016/j.ausmj.2018.05.013.

17. Bethel EC, Marchetti KA, Hecklinski TM, Daignault-Newton S, Kraft KH, Hamilton BD, et al. The LEGO™ exercise: an assessment of core competencies in urology residency interviews. J Surg Educ. 2021. https://doi.org/10.1016/j.jsurg.2021.05.011.

18. Eva KW, Rosenfeld J, Reiter HI, Norman GR. An admissions OSCE: the multiple mini-interview. Med Educ. 2004;38(3):314–26. https://doi.org/10.1046/j.1365-2923.2004.01776.x.

19. Asaad M, Rajesh A, Kambhampati PV, Rohrich RJ, Maricevich R. Virtual interviews during COVID-19: the new norm for residency applicants. Ann Plast Surg. 2021;86(4):367–70. https://doi.org/10.1097/sap.0000000000002662.

20. D'Angelo JD, D'Angelo A-LD, Mathis KL, Dozois EJ, Kelley SR. Program director opinions of virtual interviews: whatever makes my partners happy. J Surg Educ. 2021. https://doi.org/10.1016/j.jsurg.2021.04.008.

21. Vadi MG, Malkin MR, Lenart J, Stier GR, Gatling JW, Applegate RL 2nd. Comparison of web-based and face-to-face interviews for application to an anesthesiology training program: a pilot study. Int J Med Educ. 2016;7:102–8. https://doi.org/10.5116/ijme.56e5.491a.

22. Daram SR, Wu R, Tang S-j. Interview from anywhere: feasibility and utility of web-based videoconference interviews in the gastroenterology fellowship selection process. Off J Am Coll Gastroenterol | ACG. 2014;109(2):155–9. https://doi.org/10.1038/ajg.2013.278.

23. D'Angelo A-LD, D'Angelo JD, Beaty JS, Cleary RK, Hoedema RE, Mathis KL, et al. Virtual interviews – utilizing technological affordances as a predictor of applicant confidence. Am J Surg. 2021;222(6):1085–92. https://doi.org/10.1016/j.amjsurg.2021.10.003.

24. Walther JB. Selective self-presentation in computer-mediated communication: hyperpersonal dimensions of technology, language, and cognition. Comput Hum Behav. 2007;23(5):2538–57. https://doi.org/10.1016/j.chb.2006.05.002.

25. Lund S, Shaikh N, Yeh VJ, Baloul M, de Azevedo R, Pena A, et al. Conducting virtual simulated Skills multiple mini-interviews for general surgery residency interviews. J Surg Educ. 2021;78(6):1786–90. https://doi.org/10.1016/j.jsurg.2021.04.013.

26. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Adv Simul. 2016;1(1):31. https://doi.org/10.1186/s41077-016-0033-y.

27. Cook DA, Lineberry M. Consequences validity evidence: evaluating the impact of educational assessments. Acad Med. 2016;91(6):785–95. https://doi.org/10.1097/acm.0000000000001114.

28. Oliver T, Hecker K, Hausdorf PA, Conlon P. Validating MMI scores: are we measuring multiple attributes? Adv Health Sci Educ. 2014;19(3):379–92. https://doi.org/10.1007/s10459-013-9480-6.

29. Institute NE: Types of Color Blindness. 2019. https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness/types-color-blindness. Accessed 5 Mar 2021.

30. Ilgen JS, Ma IWY, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ. 2015;49(2):161–73. https://doi.org/10.1111/medu.12621.

31. Sackett PR, Laczo RM, Arvey RD. The effects of range restriction on estimates of criterion interrater reliability: implications for validation research. Pers Psychol. 2002;55(4):807–25. https://doi.org/10.1111/j.1744-6570.2002.tb00130.x.

32. Naumann FL, Marshall S, Shulruf B, Jones PD. Exploring examiner judgement of professional competence in rater based assessment. Adv Health Sci Educ. 2016;21(4):775–88. https://doi.org/10.1007/s10459-016-9665-x.

33. Mazor KM, Zanetti ML, Alper EJ, Hatem D, Barrett SV, Meterko V, et al. Assessing professionalism in the context of an objective structured clinical examination: an in-depth study of the rating process. Med Educ. 2007;41(4):331–40. https://doi.org/10.1111/j.1365-2929.2006.02692.x.

34. Knorr M, Hissbach J. Multiple mini-interviews: same concept, different approaches. Med Educ. 2014;48(12):1157–75. https://doi.org/10.1111/medu.12535.

35. Sudan R, Lynch TG, Risucci DA, Blair PG, Sachdeva AK. American college of surgeons resident objective structured clinical examination: a national program to assess clinical readiness of entering postgraduate year 1 surgery residents. Ann Surg. 2014;260(1):65–71. https://doi.org/10.1097/sla.0000000000000325.

36. Mittal VK, Lax EA. Hurdles in US surgical training for international medical graduates. Indian J Surg. 2016;78(4):257–8. https://doi.org/10.1007/s12262-016-1517-7.

37. Results and Data - 2021 Main Residency Match. The National Resident Matching Program (NRMP); 2021. p. 7.

38. Eva KW. On the generality of specificity. Med Educ. 2003;37(7):587–8. https://doi.org/10.1046/j.1365-2923.2003.01563.x.

39. Eva KW, Neville AJ, Norman GR. Exploring the etiology of content specificity: factors influencing analogic transfer and problem solving. Acad Med. 1998;73(10):S1–5. https://doi.org/10.1097/00001888-199810000-00028.

40. Lin DT, Kannappan A, Lau JN. The assessment of emotional intelligence among candidates interviewing for general surgery residency. J Surg Educ. 2013;70(4):514–21. https://doi.org/10.1016/j.jsurg.2013.03.010.

41. Lemay J-F, Lockyer JM, Collin VT, Brownell AKW. Assessment of non-cognitive traits through the admissions multiple mini-interview. Med Educ. 2007;41(6):573–9. https://doi.org/10.1111/j.1365-2923.2007.02767.x.

42. Patterson F, Knight A, Dowell J, Nicholson S, Cousans F, Cleland J. How effective are selection methods in medical education? A systematic review. Med Educ. 2016;50(1):36–60. https://doi.org/10.1111/medu.12817.

43. Hay M, Mercer AM, Lichtwark I, Tran S, Hodgson WC, Aretz HT, et al. Selecting for a sustainable workforce to meet the future healthcare needs of rural communities in Australia. Adv Health Sci Educ. 2017;22(2):533–51. https://doi.org/10.1007/s10459-016-9727-0.

44. Patterson F, Ferguson E. Testing non-cognitive attributes in selection centres: how to avoid being reliably wrong. Med Educ. 2012;46(3):240–2. https://doi.org/10.1111/j.1365-2923.2011.04193.x.

45. Nagpal K, Vats A, Lamb B, Ashrafian H, Sevdalis N, Vincent C, et al. Information transfer and communication in surgery: a systematic review. Ann Surg. 2010;252(2):225–39. https://doi.org/10.1097/SLA.0b013e3181e495c2.

46. Lingard L, Espin S, Whyte S, Regehr G, Baker GR, Reznick R, et al. Communication failures in the operating room: an observational classification of recurrent types and effects. Qual Saf Health Care. 2004;13(5):330–4. https://doi.org/10.1136/qshc.2003.008425.

47. Sutcliffe KM, Lewton E, Rosenthal MM. Communication failures: an insidious contributor to medical mishaps. Acad Med. 2004;79(2):186–94.

48. Gillespie BM, Gwinner K, Chaboyer W, Fairweather N. Team communications in surgery – creating a culture of safety. J Interprof Care. 2013;27(5):387–93. https://doi.org/10.3109/13561820.2013.784243.