This is a repository copy of *Video-based peer assessment of collaborative teamwork in a large-scale interprofessional learning activity*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/219772/

Version: Published Version

BMC Medical Education

Open Access

Video-based peer assessment of collaborative teamwork in a large-scale interprofessional learning activity

Chris Roberts[1*], Christine Jorm[2], Gillian Nisbet[3], Christopher J Gordon[4], Timothy Chen[5], Fabian Held[6] and Inam Haq[7]

Abstract

Background The assessment of team performance within large-scale Interprofessional Learning (IPL) initiatives is an important but underexplored area. It is essential for demonstrating the effectiveness of collaborative learning outcomes in preparing students for professional practice. Using Kane's validity framework, we investigated whether peer assessment of student-produced videos depicting collaborative teamwork in an IPL activity was sufficiently valid for decision-making about team performance, and where the sources of error might lie to optimise future iterations of the assessment.

Methods A large cohort of health professional students ($n = 1218$) of 8 differing professions was divided into teams containing 5–6 students. Each team collaborated on producing a short video to evidence their management of one of 12 complex patient cases. Students from two other teams, who had worked on the same case, individually rated each video using a previously developed assessment scale. A generalisability study quantified sources of error that impacted the reliability of peer assessment of collaborative teamwork. A decision study modeled the impact of differing numbers of raters. A modified Angoff determined the pass/fail mark.

Results Within a large-scale learning activity, peer assessment of collaborative teamwork was reliable (G = 0.71) based on scoring by students from two teams ($n = 10$–12) for each video. The main sources of variation were the stringency and subjectivity of fellow student assessors. Whilst professions marked with differing stringency, and individual student assessors had different views of the quality of a particular video, none of that individual assessor variance was attributable to the assessors' profession. Teams performed similarly across the 12 cases overall, and no particular professions marked differently on any particular case.

Conclusion A peer assessment of a student-produced video depicting interprofessional collaborative teamwork around the management of complex patient cases can be valid for decision-making about student team performance. Further refining marking rubrics and student assessor training could potentially modify assessor subjectivity. The impact of professions on assessing individual peers and the case-specificity of team performances in IPL settings need further exploration. This innovative approach to assessment offers a promising avenue for enhancing the measurement of collaborative learning outcomes in large-scale Interprofessional learning initiatives.

Keywords Peer assessment, Collaborative teamwork, Interprofessional learning, Student-generated videos, Video-based assessment, Generalisability Theory, Teamwork, Collaboration, Professions, Health professional students

---

*Correspondence:
Chris Roberts
chris.roberts@sheffield.ac.uk
Full list of author information is available at the end of the article

Roberts *et al. BMC Medical Education* (2024) 24:1307

Page 2 of 13

Introduction

Interprofessional Learning (IPL) is increasingly important in health professional education, aiming to cultivate collaborative competencies across diverse professions and enhance patient-centred healthcare delivery [1–4]. International frameworks emphasize opportunities for students to work in interprofessional groups, focusing on collaboration and shared decision-making [5–7]. Internationally, although more institutions are adopting IPL [8], activities are often small-scale and voluntary. Several interprofessional competency statements have been published internationally and inform the academic content of IPL assessment rubrics, such as the one used in this study [9–12]. Amidst the drive for promoting IPL, there is a significant and persisting challenge in how to effectively assess the performance of interprofessional teams in authentic large-scale IPL activities [13], which align with the interprofessional learning outcomes critical for healthcare [14]. This highlights a critical gap in the literature; a scarcity of robust assessment methodologies that can comprehensively assess the teamwork, communication, and decision-making skills of students engaged in large scale IPL settings ($n > 300$) [2, 7, 15–17].

Current IPL assessments often fail to capture important teamwork skills, often testing if students can function expertly as competent individuals, or independent practitioners [18–20]. Assessments focus on short-term knowledge acquisition, impact on attitudes to other professions, and student satisfaction [21]. Several studies report self-assessment, which are not optimal given an individual's inability to assess their own learning gains accurately [22, 23]. A growing body of research explores generating valid team scores for large-scale IPL activities [20, 23–26].

In this paper, we extend our research on large scale interprofessional assessment [15, 27–29]. We investigate the validity evidence from a novel large-scale interprofessional activity where student teams created videos showcasing their problem-solving and decision-making for a complex patient case. These videos were then used for peer assessment of teamwork skills. The novelty lies in the use of digital technologies to underpin a methodology designed explicitly for peer assessment of team performance in large scale IPL activities.

We now describe our conceptual framework linking the relationship between our research questions and differing components of the assessment of IPL activities; assessment of teams, peer assessment, video-based assessment, and the validity criteria for the rigour of the assessment process.

Assessment in IPL

Effective assessment is crucial for interprofessional learning (IPL) as it shapes student learning and informs decision-making. Assessing team performance in IPL is complex encompassing aspects like leadership, communication, collaboration, and decision-making. Our challenge was to incorporate peer assessment, embrace video-based assessment, apply appropriate tools for measurement and use a robust validity framework to make the case for using this form of assessment. It is important to distinguish between an assessment designed to measure the collective competence of a team versus the individual competency of team members [30–32]. Factors that impact any team-based assessment include the specific task (e.g. patient case management), team composition, and cultural and organizational contexts [4]. Several tools have been developed to observe and measure student team performance either in real-time (in clinical settings or simulation) or video recordings [7, 20, 24].

An additional consideration is the potential impact of a student's profession on their assessment of teamwork. Established professional perspectives can hinder collaborative learning [33–35], particularly in assessment [2, 7, 19]. While professional hierarchies may lessen after IPL activities, they can resurface later. Furthermore, research suggests professional background might influence how students assess teamwork, but it's unclear if this originates from training or individual student experiences [36].

Peer assessment

Peer assessment in IPL has been recognised as a valuable tool for some time [7, 37–39]. It involves learners from different professions collaborating and then assessing each other's work. In the uniprofessional context peer assessment helps develop student skills in giving and receiving feedback. By reviewing their peers' work, students can reflect on their own performance and identify areas for improvement, thus fostering self-regulated learning [40]. Additionally, students have more opportunities to observe each other compared to instructors, who have limited time to observe each student individually [41, 42]. Students are more likely to see authentic aspects of interprofessional collaboration than tutors when engaged in, for example, interprofessional practice-based learning in the clinical setting [43]. Several studies have measured uniprofessional small group learning effectiveness based on student self-assessment of particular behaviours [44–46]. While the quality of evidence on peer assessment can vary due to methodological differences [47], it is thought that peer ratings are reliable and

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 3 of 13

more valid than self- or tutor-based assessments [39, 48, 49]. Both the giving and receiving of peer feedback have good educational impact for the student assessor [50].

For optimal use of peer assessments, educators should articulate the intended purpose, clarify the significance of teamwork, familiarize students with the assessment tool, evaluate teamwork consistently over time, offer constructive feedback, diminish the weight of grades linked to assessments, and consider partial anonymity in feedback collection [51].

Student-produced video in assessment for learning

There is a long tradition of using video to assess student clinical skills in the laboratory setting [52, 53] and of using video to develop students' knowledge or skills acquisition [54]. There is a developing literature on student-produced video as a format in assessment *for* learning and *of* learning. It is a form of assessment that provides students with important graduate capabilities, to actively apply knowledge in authentic contexts, promote self-reflection and the giving and receiving of feedback from others [55]. Examples include science education [56, 57], health promotion education [58–60], and communication skills training in both dentistry [61], pharmacy [62], and interprofessional settings [63, 64].

In the broader educational literature, studies of student-produced video that include details of the assessment, focus on the adequacy of topic content, communication of the key message, and the technical quality of the video product [15]. A recent review highlighted that projects with unclear purposes and weak pedagogical design can hinder intended learning outcomes in health professions education [65].

Kane's validity framework

To evaluate the evidence for the utility of video based peer assessment of students interprofessional teamworking skills, Kane's argument-based validity framework has value, focusing on four key aspects [66, 67] in Table 1.

As part of the generalisation argument, reliability is important for ensuring that assessment scores are reproducible across different assessments, Reliability, although essential, is not adequate on its own to establish validity. The generalizability co-efficient is a measure used to calculate reliability and estimate multiple sources of error [68].

In summary, assessing Interprofessional Learning (IPL) in medical and health science students involves addressing challenges in assessing whole of team performance. This includes considering task complexity, team composition, and contextual factors. It highlights the use of video-based assessments and peer evaluations as potentially important for the scalability of assessing large cohorts. It also underscores the need to recognize how professional backgrounds can impact collaborative learning and assessment. The application of Kane's validity framework is introduced, emphasizing the importance of credibility, reliability, and validity in the assessment process.

Study aims

In this paper, we posed two research questions.

1) What factors impact the validity of a peer assessment of a student-produced video depicting interprofessional collaborative teamwork in a large-scale IPL activity?
2) How do the student assessor's profession and the specific patient case influence student ratings of collaborative teamwork in the video assessments?

Methods

The research context

While the educational design specifics of the IPL activity are discussed elsewhere [27, 55, 69], we provide a brief overview here. In 2015, the implementation of a large-scale IPL event; "The Healthcare Collaboration Challenge," involved four health faculties and required the mandatory attendance of a number of cohorts of health professional students. At the time of the data collection, there was no systematic preparation of health professional students around IPL, though there were variations in the amount of clinical exposure and thus informal work-based exposure to IPL. Thus, we considered that

Table 1 Kane's validity framework applied to the video based peer assessment of small groups of students in a large scale interprofessional learning activity

• **Scoring:** Translating video observations into one or more scores

• **Generalisation:** Using the observed scores to generate an overall test score representing collaborative teamwork in the IPL setting

• **Extrapolation:** Drawing inferences about what the test score might imply for the collaborative teamwork abilities of students

• **Implications:** Assessing whether the scores are credible and reasonably free from error, thereby making them suitable for the assessment of collaborative teamwork. If the scores demonstrate credibility and reliability, they can be considered a valid basis for making decisions about the collaborative teamwork abilities of students, aligning with the overarching goals of IPL assessment

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 4 of 13

students were naïve to formal IPL prior to this event. However, the HCC was one of several IPE activities to be implemented on an annual basis by the IPL Team as part of an overarching IPL strategy [70].

Students (*n*=1218) from eight different health professions were randomly assigned to one of 208 teams in groups of 5–6 students. Most groups consisted of four or more professions. The larger size of the student cohorts of medicine, nursing, and pharmacy meant most groups had two representatives from these professions. Students groups were randomly assigned to one of 12 case-based learning activities [71]. These had been developed by small multiprofessional groups of clinicians from authentic patient cases and were presented in a structured format to a consistent level of difficulty. The cases involved complex continued patient care based in hospital or in the community, requiring input from multiple health professions. Typical cases included a man living with multiple complications of diabetes and cardiovascular disease, and several palliative care scenarios. (An example of a case is given in Appendix 1) Prior to the learning activity, students were directed to the University learning management system to obtain an overview of the event with information about team allocation, instructions about participation, their allocated case, and guidelines about the assessment task. They were informed that they needed to produce a 5 minute video that depicted interprofessional collaborative teamwork across patient issues, team decision-making negotiations, and their management plan for the case study. Student teams were asked to use their own technology (e.g., smartphones and tablets) to film and edit the video. Exemplars from the previous year were available for students to view. A written one-page management plan formed part of the overall assessment but is not considered in this paper. Each of the eight health professions involved had negotiated differing assessment strategies for their participating students, ranging from pass/fail to a proportion of a unit of study score.

Students attended one of three orientation sessions for a 10-min briefing on IPL's significance and the learning activity requirements. Subsequently, students gathered in teams, introduced themselves, and engaged in icebreaker team-building exercises to foster problem-solving skills (approximately 20 min). Following this, students worked on their cases and producing the video and the management plan.

Teams were given 48 hours after the face-to-face event to complete the assessment and upload videos to the University's Learning Management System. Students then peer-assessed two other teams' videos online on the same case, using the rubric for peer assessment.

Rubric development

The need for the video assessment scale in IPL, it's development and content validation and piloting has been described in detail elsewhere [15]. This was based on an amalgam of several interprofessional competency statements [9–12]. Collaboration in teams is a "complex, voluntary and dynamic process involving several skills" including an appreciation of the patients perspective [4]. In summary, we developed the prototype collaborative teamwork rubric for this purpose by synthesising the contemporary literature on teamwork in IPL, peer assessment, and assessment of student-generated videos. The previous content validation study identified the major domains that were important to staff and students in assessing student-produced videos portraying interprofessional collaborative practice within a large-scale learning activity [15]. The finalised scale was intended to assess four domains.

- **Patient issues:** respect for the patient and family's experiences and ability to view the situation through the lens of the patient and family,
- **Interprofessional negotiation**: the ability to negotiate with other health professionals in problem solving for the patient.
- **The management plan**: the practicalities of interprofessional care including the coordination required for a well-executed interprofessional management plan.
- **The effective use of the video medium** to engage and communicate the key messages of interprofessional working.

Each domain was rated on a Likert scale of 1–4, with 1 indicating 'poor:' and 4, 'excellent.' A global rating on a scale of 1–5 of the overall impression of the video in the context of the IPL activity was included, with 5 being excellent. (See Fig. 1) A modified Angoff standard setting procedure for the whole activity was undertaken for the combined weighted score of 60% video mark plus 40% abstract mark giving a pass mark of 50%.

Data collection

The data cleaning process for the peer assessments of collaborative teamwork videos involved several steps to ensure data integrity. First, all instances of missing data were removed to maintain completeness. Twenty-five instances of self-assessment, where the 'team' value equalled the 'own team' value, were identified, and removed to avoid bias. Verification of categorical variables included confirming the levels and ranges

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 5 of 13

HCC Video Assessment Rubric

	Poor 1	Satisfactory 2	Good 3	Excellent 4
Patient issues	Issues faced by the patient and family not evident.	Describes the major issues faced by the patient and family.	Depicts appreciation of depth and/or breadth of issues faced by the patient and family.	Depicts considerable appreciation of depth and breadth of issues faced by the patient and family.
Interprofessional negotiation	Does not display negotiation, shared goal setting and shared decision making.	Shows limited appreciation of negotiation, shared goal setting and shared decision making.	Shows appreciation of negotiation, shared goal setting and shared decision making.	Sophisticated approach to negotiation, shared goal setting and shared decision making.
Interprofessional management plan in action	Management plan not evident in video.	Video depicts limited evidence of management plan in action.	Video depicts good evidence of management plan in action.	Strong depiction of co-ordinated and well executed interprofessional care.
Effective use of video medium to engage audience	Poor use of video medium.	Appropriate use of video medium.	Engaging.	Highly engaging and memorable.

Fig. 1 Rubric for the peer assessment of a student produced video depicting collaborative teamwork in a large-scale IPL activity

for anonymised Student Identity number, Team, Item, Own Team, Case, Discipline, and Value.

Data analysis
Descriptive statistics were generated for the number of students including missing data, the number of ratings, the mean score, and the standard deviation of the mean scores. In addition, marginal mean score and standard error of the mean was calculated for each discipline along with significances of distance between mean scores.

Exploratory factor analysis
Principle axis factoring with an Eigenvalue threshold of 1 and Varimax rotation was employed for Exploratory Factor Analysis (EFA) to identify positively correlated factors. The Kaiser–Meyer–Olkin measure and Bartlett's test of sphericity assessed sampling adequacy (accepted if > 0.50) and sphericity (considered sufficient if $p < 0.05$), respectively. Factor loadings exceeding 0.4 were deemed indicative of a good fit for the items.

Generalisability theory
In Generalisability theory [72], the G–study provides a multifactorial perspective of the peer rated collaborative teamwork scores by quantifying sources of measurement error that arise from potentially controllable factors. The characteristic of interest in this study was the quality of collaboration displayed in the team video. We were interested to derive the sources of error, which did not

relate to this construct of interest. This included assessor stringency/leniency, which is a first-order effect and is defined as the consistent tendency of student assessors to use either the top or the bottom end of the rating scale. Assessor subjectivity refers to variable student assessor preferences in relation to the collaborative teamwork as displayed in the video, and includes how different assessors favour different examples of videos differently over and above their baseline stringency/leniency [73]. We were also interested to know if the students' profession contributed to error and if the case that students rated, impacted their assessment—either as individuals or by their profession.

We used the General Linear Model within SPSS (version 24) to estimate the influence of all these factors (facets in G-theory)—both first-order effects (e.g., profession and case) and second-order effects (e.g., the interaction between profession and case). A reverse stepwise regression analysis [74] started with the most comprehensive possible model and then excluded redundant variables, and variables contributing less than 3% of the overall variance, one at a time, to derive the simplest fully-explanatory regression model.

Having determined the best fit model for our data in this way, variance estimates were combined [73] to provide an index of reliability (the G coefficient). This allows future iterations of the assessment program to be modified to address the main sources of error identified in the initial study. The G-study was based on a partially crossed

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 6 of 13

design with students as assessors (student assessor) partially crossed with the rated student teams (team). Once redundant factors were removed, the reliability of student assessors' ratings of collaborative teamwork in the fully nested situation (worst case scenario) was calculated in a D-study as:

$$G = Var_{\text{student assessor}}/(Var_{\text{team}} + (Var_{\text{student assessor}}/n) + (Var_{\text{student assessor}*\text{team}}/n))$$

where n is the number of student assessors rating a team video.

This assumes that there was a mix of similar professions in each of the teams, but minimal crossing of the student assessors. The actual situation was slightly more favourable since each student assessed the videos of two other teams – providing some crossover. This would yield slightly more favourable reliability but is difficult to model.

Ethics

All research method were conducted in accordance with relevant guidelines and regulations. The University of Sydney Human Research Ethics Committee approved the research. (Protocol number: 2015/320). The learning activity itself was mandatory. There were prizes awarded to student groups for the best video production participation in the research was entirely voluntary. Students having any difficulties with the activity were referred to their tutors. No financial incentive was offered to the students to allow their data to be included in this study. Informed consent of the students to participate in the study was obtained by the following method. All students were provided with information about the study and their rights as participants. They were given the opportunity to indicate their consent for their data to be used in evaluative research by clicking on an ethical statement within the Learning Management System.

Results

Impacts on the validity of the assessment

Data on peer assessment rating of collaborative teamwork were available for 1218 of the 1220 students scheduled to undertake the activity in 208 teams, using a marking rubric with four checklist items and a global rating, to rate performances on one of 12 patient cases, by teams made up of 5–6 members drawn from 8 health professions. (See Table 2) Medical and nursing students provided scores very close to the mean (3.18 and 3.19) students, a lower mean score (2.97). No teams failed the expected standard.

The Exploratory Factor Analysis revealed a single factor with positive correlations, explaining 54.7% of the variance. This aligns strongly with the primary assessment construct tied to their respective domains of patient issues, team decision-making negotiations, their management plan for the case study, and effective use of the video. All factor loadings surpassed 0.4 (Table 3).

During the reverse stepwise regression process, several facets were excluded as redundant based on their insignificant contribution to score variance. Case related factors (Var_{case}) and its interactions with item and profession

Table 2 Students with complete data (*n* = 1218) ranked by cohort size rating, showing profession, year of study, program duration, number of students, number of ratings, mean collaborative teamwork rating and standard deviation

Student profession	Cohort			Number of students (missing data)	Number of ratings	Mean rating (SD)
	Year of Study	Program				
		Duration	Type			
Medicine	1	4	Graduate Entry	317 (8)	3085	3.18 (0.76)
Nursing	3	3	Undergraduate	312 (11)	3010	3.19 (0.77)
	2	2	Graduate Entry Masters			
Pharmacy	4	4	Undergraduate	235 (4)	2295	3.29 (0.74)
	2	2	Graduate Entry Masters			
Speech Pathology	3	4	Undergraduate	136 (1)	1345	3.08 (0.76)
	2	2	Graduate Entry Masters			
Diagnostic Radiography	2	2	Graduate Entry Masters	74 (1)	730	3.56 (0.67)
Occupational Therapy	2	2	Graduate Entry Masters	73 (0)	725	3.09 (0.69)
Physiotherapy	3	4	Undergraduate	70 (0)	700	2.97 (0.68)
Exercise Physiology	2	2	Graduate Entry Masters	26 (0)	260	3.04 (0.73)
Total				1218 (25)	12,150	3.18 (0.73)

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 7 of 13

Table 3 Exploratory factor analysis using principal axis rotation factor matrix revealing single factor structure, (which cannot be rotated) of the video assessment rubric across all five items

Component	Factor 1
Overall	0.80
Plan	0.78
Negotiation	0.74
Video	0.72
Patient	0.66

($Var_{case*item}$ & $Var_{case*profession}$) did not make a significant independent contribution to score variance (supplementary tables S1, S2A and S2B). This implies that no specific case was perceived as easier or more difficult than others, no item or element received 'uncharacteristic' ratings on a particular case, and no professional group demonstrated varying stringency or leniency across cases.

The rating item (Var_{item}) contributed minor score variance (3% of overall variance), and rating item interactions with student assessor, team, and profession ($Var_{item*student\ assessor}$, $Var_{item*team}$ & $Var_{item*profession}$) made only very small contributions to score variance ($\leq 3\%$ of all variance) (Supplementary tables S3A, S3B, and S3C). This suggests that while some items were slightly easier than others, no item or element received 'uncharacteristic' ratings from a particular student assessor, profession, or in relation to a specific team.

In the next step (supplementary table S4), whilst professions marked with differing stringency/leniency, ($Var_{profession}=10\%$) and individual student assessors had different views of the quality of a particular video ($Var_{student\ assessor*team}=32\%$), none of that individual assessor variance was attributable to the rater's profession ($Var_{team*profession}=0\%$). This indicates that there is variability in stringency/leniency among professions and among individual student assessors assessing the quality of a particular video. However, the last part ($Var_{team*profession}=0\%$) suggests that the variability in individual assessor subjectivity is not influenced by the

profession of the rater. S4). In other words, the profession of the rater does not explain the individual subjectivity in their ratings; individuals from the same profession do not have more similar tastes than individuals from different professions.

The final model included only the variance components shown in Table 4 – all the facets which explain or cause significant independent score variance. The results of the G study with our best-fit model, show both wanted and unwanted facets contributing to the variation in collaborative teamwork score (Var_{team}). The largest contributor to error variance was student assessor stringency/leniency ($Var_{student\ assessor}$), followed by assessor subjectivity ($Var_{team*student\ assessor}$), and profession differences ($Var_{profession}$). This highlights the key contributors to error variance in the collaborative teamwork scores, with student assessor stringency/leniency being the most significant, followed by student assessor subjectivity and profession differences making a smaller contribution.

In the fully nested D study, we represent the worst-case scenario in terms of sources of unwanted variance that contribute to the reliability of the assessment. The dependability estimates combining the variance components according to the formulae above are given in Table 5. Peer assessment of collaborative teamwork was reliable (G=0.71), where each team video is rated by

Table 5 D study (fully nested) modeling changes in reliability estimates for groups of 5–6 students, assuming a matching of similar professions but no crossing of student assessors

No. of Student Assessors	G coefficient (fully nested)
4	0.45
6	0.55
8	0.62
10	0.67
12	0.71
14	0.74
16	0.76
18	0.78
20	0.80

Table 4 Variance components of the peer assessment of a video depicting interprofessional collaborative teamwork around management of one of 12 complex clinical cases in health professional students ($n=1218$) of 8 differing professions

Component	Meaning	Variance Estimate	Proportion of total Variance
$Var_{student\ assessor}$	Student Assessor stringency/leniency	0.164	43%
Var_{rated_team}	Team performance as rated by the video	0.056	15%
$Var_{profession}$	Profession Stringency/leniency	0.037	10%
$V_{rated\_team*student\ assessor}$	Student Assessor subjectivity	0.120	32%

twelve student assessors. In our IPL activity, this represents two other teams. Using three teams to rate one video would have provided higher reliability (G=0.78). However, due to practical constraints, this approach was not feasible. Despite facing challenges in the study setting, the peer assessment of collaborative teamwork demonstrated acceptable reliability with an achievable rater configuration.

Exploring differences in professions

To further explore differences in marking by profession, we looked at the marginal mean score and standard error of the mean for each discipline along with significances of distance between mean scores. (Table 6).

The shaded squares show where no significant distance exists. Our analysis revealed significant differences ($p < 0.05$) in the mean scores between several disciplines on the assessment. Post-hoc tests (e.g., Tukey's HSD) further illuminated these differences, identifying four distinct score groups.

Diagnostic Radiography students achieved the highest mean score (3.740, SE=0.024), which was statistically significantly different from all other disciplines ($p < 0.05$). Pharmacy students followed closely with a second-highest average score (3.454, SE=0.014), statistically indistinguishable from the scores of Medicine (3.342, SE=0.012) and Nursing (3.347, SE=0.012).

A third group emerged, comprising Occupational Therapy (3.239, SE=0.024), Speech Pathology (3.234, SE=0.018), and Exercise Physiology (3.188, SE=0.040) students. There were no statistically significant differences in mean scores between these disciplines.

Interestingly, Physiotherapy students (3.124, SE=0.025) scored the lowest on average. This score, however, was not statistically different from the scores in the third group, suggesting Physiotherapy might be an outlier within that cluster.

These findings highlight potential variations in performance across disciplines on this assessment. Further investigation is warranted to understand the reasons behind these differences.

Discussion
Summary of key findings

This study concerns students from different professions working in teams and assessing their peers' inter-professional collaborations on complex cases. We sought evidence of the validity of a video-based assessment and whether the profession of the student had influence on their marking. The findings estimated the reliability of a peer assessment of a health professional student-produced video of interprofessional collaborative teamwork. Peer assessment in this context was sufficiently reliable, with modest assessor sampling, for decision-making about the quality of the student team performance as depicted by the video. Most of the error variance was attributable to student assessor stringency/leniency and subjectivity. Therefore, strategies to reduce unwanted variance should focus on both elements of student assessor judgement.

We also explored the influence of students' professions on assessors' judgements about collaborative teamwork during an IPL activity. Some professions rated more stringently than others across the board. However, the professional designations were similar in which videos they thought were better and which were worse (same preferences, but differing standards). That is differences between professions were evident, but these

Table 6 Showing the marginal mean score and standard error of the mean for each discipline along with significances of distance between mean scores

(I) Discipline	mean	s.e. mean	Significance of distance between row and column mean score							
			Physio	Exercise - Physio	Speech Path	Occup-Therapy	Medicine	Nursing	Pharm	Diag-Radio
Physiotherapy	3.124	0.025		0.994	0.007	0.023	0.000	0.000	0.000	0.000
Exercise Physiology	3.188	0.040	0.994		1.000	1.000	0.007	0.004	0.000	0.000
Speech Pathology	3.234	0.018	0.007	1.000		1.000	0.000	0.000	0.000	0.000
Occupational Therapy	3.239	0.024	0.023	1.000	1.000		0.003	0.001	0.000	0.000
Medicine	3.342	0.012	0.000	0.007	0.000	0.003		1.000	0.000	0.000
Nursing	3.347	0.012	0.000	0.004	0.000	0.001	1.000		0.000	0.000
Pharmacy	3.454	0.014	0.000	0.000	0.000	0.000	0.000	0.000		0.000
Diagnostic Radiography	3.740	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000	

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 9 of 13

did not significantly affect the overall reliability of the assessment.

Further, the validity of the assessment was investigated by examining the key claims, assumptions, and inferences that link scores derived from peer ratings of interprofessional team working with their intended interpretations using Kane's validity framework.

Scoring

The peer assessment involved a rubric with four checklist items and a global rating, contributing to the scoring component. Exploratory Factor Analysis (EFA) revealed it to be assessing a single construct with positive correlations, aligning with the primary assessment constructs tied to patient issues, interprofessional negotiation, interprofessional management planning and the effective use of video to depict effective collaborative teamwork.

Generalisation

Reverse stepwise regression excluded facets like case, item, and their interactions that didn't significantly contribute to score variance. This implies that the findings can be generalized, suggesting no specific case, item, or element significantly impacted the assessment. Factor structure remained flat, indicating that including or excluding the global item resulted in a single factor explaining a significant portion of the variance, further supporting generalisation.

Extrapolation

The study provides insights into how different professions marked with differing stringency/leniency, indicating potential variations in expectations. While individual student assessors had different views, the results show no influence of the rater's profession on individual subjectivity suggesting that the profession of the rater doesn't explain individual subjectivity. Notably, no team failed the assessment, indicating a consistent level of performance across teams.

Implications

The final model, derived from G study results, identifies key contributors to error variance, with student assessor stringency/leniency being the most significant, followed by assessor subjectivity and profession differences. These findings have implications for training assessors and refining the assessment process. Dependability estimates from the D study highlighted the reliability of peer assessment considering practical constraints in the study setting [67].

Comparison with existing literature

Our findings add to the literature around assessment *for* the learning of interprofessional teamwork for whole student cohorts [2]. In previous work, we have detailed the educational design of this interprofessional learning activity that engaged all students in authentic team working, whilst at the same time enabling reflection, learning, and demonstrating the creation of new collective knowledge [27]. This study extends the validity evidence from the existing content validation of the video scale [15, 27] by providing psychometric evidence on aspects of a video-based assessment of interprofessional collaborative teamwork. Our findings support the argument that peer assessment might overcome some of the challenges of large scale applications of IPL [7].

Not only does peer assessment offer feasibility and learning benefits for large cohorts of students, but it can also provide reliable scores. Given the purpose of the collaborative video assessment task was to determine whether a team of students have demonstrated the expected standards of interprofessional collaboration, our data support the use of peer assessment for this purpose. It also extends research demonstrating the utility of the case based approach for IPL activities for large numbers of students [75].

Our work provides further insight into the sources of measurement error in using Interprofessional assessment tools [76] and addressing student assessors' subjectivity and stringency/leniency. Students' reflection on the purpose and likely outcomes of peer assessment may influence the way in which they make their judgements. They will need to understand how reflection and goal setting can influence their own professional behaviours [77–79] in IPL, as well as their peer rating behaviour of interprofessional collaboration in others.

The impact of profession on rating team performance is an interesting finding. In post qualification settings where ratings involve doctors and nurses around interprofessional collaboration, it appears that professional designation is a significant factor [36]. Our findings that medical and nursing students appear to have similar preferences and standards for team performance suggest that the change may happen post qualification in the workplace. However, this finding is likely to be context specific to the particular assessment task [22]. That is this observation of a non-hierarchical approach by students might be because team performance, rather than individual performance, is being assessed. Further, the IPL activity included first-year graduate entry medical students and more senior medical students may have different perceptions of IPL team performance in this context.

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 10 of 13

That teams performed similarly across the 12 patient cases overall, and no profession marked differently on any particular case is an interesting finding. There has been a call from some quarters that there is value for all students in team-based learning to undertake the same case, often for pragmatic reasons [80]. Our finding that providing multiple diverse cases albeit of similar complexity, does not negatively impact learning, is reassuring. This is important because there are many educational benefits of a large-scale IPL event where students engage with a range of authentic clinical scenarios [27, 71].

We have demonstrated reasonable characteristics for an assessment process that engaged a large number of students with minimal imposition on faculty. We agree with Boud et al. [14] that explicitly focusing on improving assessment practice in IPL may have a large impact on the quality of the collaborative learning agenda.

Methodological strengths and challenges

This exploratory study is one of few studies to identify the factors that influence student assessor judgement in a peer assessment of collaborative teamwork within a large-scale interprofessional learning activity. We acknowledge the potential impact of three important issues influencing the generalisability of our findings to other settings: the generalisability design, the robustness of the peer assessment tool, and the IPL research context for student learning. Conceptually, G theory treats data from a multivariate random effects perspective [81]. The generalisability study makes use of naturalistic data to fully explore all possible effects or facets and their interactions. However, this is a different paradigm from hypothesis testing which typically fixes and isolates variables for effect testing. We consider that the exploratory paradigm was more appropriate to our research questions. [81]. We acknowledge that for performance assessments, rater error is often outweighed by task sampling error [82]. One way to average out subjectivity in future iterations of the assessment would be to increase the number of tasks (team producing more than 1 video) – this would average out "subjectivity." However, this was not feasible in our setting. No existing previously validated peer assessment tool of a video of collaborative teamwork was available at the time, therefore we developed a tool to best suit our context. We acknowledge from a construct validity perspective that our data only provides preliminary evidence for a robust validity argument (scoring, generalisation, extrapolation, and implications) for the video-based assessment tool.

We don't have the data to understand what drives student assessor error in these ratings but plan to investigate these questions using qualitative data. Unfortunately, we did not categorise the data for the study to record the differences between graduate and undergraduate responses. We acknowledge that this is a measure of a collective competency [30] in collaborative teamwork. It does not separate out the individual contribution to collaborative learning. In considering the generalisability to other IPL settings, we have sampled widely across students, differing types of cases, and differing health professions.

Implications for educators

Educators can draw valuable insights from this study to enhance their Interprofessional Learning (IPL) approaches. This study shows that an assessment of the performance of interprofessional teams provided scores that can be trusted, and second members of one professional group can be trusted to judge the performance of members in other professional groups in dealing with particular types of patient cases. This study provides some preliminary validity evidence for integrating peer assessment into IPL activities.. Therefore, assessment can be leveraged as a driving force in implementing IPL. By shifting the focus to assessment *for* learning, educators can provide robust evidence of the quality of interprofessional collaborative learning. It seems possible that a focus on assessment in IPL reframes the 'problem' of assessing large cohorts' in IPL [2] to one of opportunity and provides a possible way forward in progressing the IPL agenda. Student assessment often drives learning.

Furthermore, educators should emphasise the preparation of students for peer assessment. By cultivating a deep understanding of peer assessment's purpose and implications, students can better assess their peers. Educators should consider interventions that enhance student assessor judgment. Recognizing that biases in peer assessment primarily stem from student assessor leniency and subjectivity, educators can proactively address these influences by encouraging students to critically reflect on their judgement of an IPL performance in the light of the intended criteria.

Incorporating diverse cases of equal complexity in large-scale IPL events need not compromise learning outcomes. Educators can confidently expand the scope of scenarios, enriching students' exposure to authentic clinical situations.

Implications for further research

First, the transformative role of assessment *fo*r learning within the IPL landscape emerges as a focal point of this study. The study illustrates assessment's capacity to catalyse learning experiences. The shift in perspective from viewing assessment challenges as obstacles to recognizing them as avenues for advancing IPL prompts

researchers to consider innovative ways of integrating assessment *for* learning methods that resonate with the goals of IPL.

Second, the study underscores the need for more validity evidence of peer assessment integration within the IPL paradigm. This observation encourages researchers to investigate the potential of peer assessment as a dependable assessment mechanism across diverse IPL contexts. Given the role of student assessor leniency and subjectivity in shaping assessment outcomes, researchers could explore the underpinning causes of these biases, with a view to conceptualizing strategies that can temper their impact. In our study, most students had had little formal preparation for IPL activities, which accords with the findings of others [13]. In settings where there is formal preparation for IPL, we anticipate a reduction in student assessor subjectivity and the need for fewer student ratings. Further empirical studies on IPL learning activities that use video-based peer assessment are a rich area for further research.

Conclusion

A peer assessment of a student-produced video depicting interprofessional collaborative teamwork around management of complex patient cases can be valid for decision-making about student team performance. Each student profession had the same view of good or poor collaborative teamwork but marked to differing standards. The overall standard of collaborative teamwork was similar across a range of cases and the judgements of different professions did not diverge by case. It's possible that further refining marking rubrics and assessor training could modify student assessor subjectivity. Similarly, the impact of profession on assessing individual peers and the case-specificity of team performances in IPL settings need further exploration. This innovative approach to assessment offers a promising avenue for enhancing the measurement of collaborative learning outcomes in large-scale Interprofessional Learning initiatives.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12909-024-06124-4.

Supplementary Material 1.

Supplementary Material 2.

Data availability
Findings from the datasets supporting the conclusions of this article are included within the article. The datasets generated and analysed during the current study are not publicly available due to confidentiality agreements approved by the Human Research Ethics Committee but are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
All research method were conducted in accordance with relevant guidelines and regulations. The University of Sydney Human Research Ethics Committee approved the research. (Protocol number: 2015/320). The learning activity itself was mandatory. There were prizes awarded to student group for the best video production participation in the research was entirely voluntary. Students having any difficulties with the activity were referred to their tutors. No financial incentive was offered to the students to allow their data to be included in this study. Informed consent of the students to participate in the study was obtained by the following method. All students were provided with information about the study and their rights as participants. They were given the opportunity to indicate their consent for their data to be used in evaluative research by clicking on an ethical statement within the Learning Management System.

Consent for publication
Not applicable.

Competing interests
The following authors have a conflict to declare. CR is a Senior Editorial Board member and IH is an editorial board member of BMC Medical Education. CJ, GN, CG, TC and FH have no conflicts of interest to declare.

Author details
¹Division of Medicine, School of Medicine and Population Health, The University of Sheffield, Sheffield, UK. ²Independent Scholar, Sydney, NSW, Australia. ³Faculty of Medicine and Health, The University of Sydney, Sydney School of Health Sciences, Sydney, NSW, Australia. ⁴Department of Health Sciences, Faculty of Health and Human Sciences, Macquarie University, Sydney, NSW, Australia. ⁵Faculty of Medicine and Health, School of Pharmacy, The University of Sydney, Sydney, NSW, Australia. ⁶Office of the Deputy-Vice Chancellor Education (Enterprise and Engagement), The University of Sydney, Sydney, NSW, Australia. ⁷Education, Learning and Assessment, Royal Australian College of Physicians, Executive General Manager, Sydney, NSW, Australia.

References
1. Bogossian F, Craven D. A review of the requirements for interprofessional education and interprofessional collaboration in accreditation and practice standards for health professionals in Australia. J Interprof Care. 2021;35(5):691–700.
2. Simmons BS, Wagner SJ, Reeves S: Assessment of Interprofessional Education: Key Issues, Ideas, Challenges, and Opportunities. In: Assessing

Competence in Professional Performance across Disciplines and Professions. edn. AG Switzerland: Springer; 2016. p. 237–252.

3. Thistlethwaite JE. Values-based interprofessional collaborative practice: Working together in health care. Cambridge: Cambridge University Press; 2012.

4. D'Amour D, Ferrada-Videla M, Rodriguez LSM, Beaulieu M-D. The conceptual basis for interprofessional collaboration: core concepts and theoretical frameworks. J Interprof Care. 2005;19(sup1):116–31.

5.   Collaborative HPA. Guidance on developing quality interprofessional education for the health professions. In: Chicago: Health Professions Accreditors Collaborative; 2019.

6. Thistlethwaite JE, Forman D, Matthews LR, Rogers GD, Steketee C, Yassine T. Competencies and frameworks in interprofessional education: a comparative analysis. Acad Med. 2014;89(6):869–75.

7. Rogers GD, Thistlethwaite JE, Anderson ES, Abrandt Dahlgren M, Grymonpre RE, Moran M, Samarasekera DD. International consensus statement on the assessment of interprofessional learning outcomes. Med Teach. 2017;39(4):347–59.

8. Khalili H, Lackie K, Langlois S, da Silva Souza CM, Wetzlmair L-C. The status of interprofessional education (IPE) at regional and global levels – update from 2022 global IPE situational analysis. J Interprof Care. 38(2):388–93.

9. Thistlethwaite J, Moran M, World Health Organization Study Group on Interprofessional E, Collaborative P. Learning outcomes for interprofessional education (IPE): literature review and synthesis. J Interprof Care. 2010;24(5):503–13.

10. O'Keefe M. Collaborating across boundaries - A framework for an integrated interprofessiona curriculum. In: Australian Government Office for Learning and Teaching. 2015.

11. Orchard C, Bainbridge L, Bassendowski S, Stevenson K, Wagner SJ, Weinberg L, Curran V, Di Loreto L, Sawatsky-Girling B. A national interprofessional competency framework. 2010.

12. Collaborative IE. Core competencies for interprofessional collaborative practice: 2016 update. Washington, DC: Interprofessional Education Collaborative; 2016. p. 1–19.

13. Thistlethwaite J, Dallest K, Moran M, Dunston R, Roberts C, Eley D, Bogossian F, Forman D, Bainbridge L, Drynan D, et al. Introducing the individual Teamwork Observation and Feedback Tool (iTOFT): development and description of a new interprofessional teamwork measure. J Interprof Care. 2016;30(4):526–8.

14. Boud D, Sadler R, Joughin G, James R, Freeman M, Kift S, Webb G. Assessment 2020: Seven propositions for assessment reform in higher education. Sydney, Australia: Australian Learning and Teaching Council. In; 2010.

15. Nisbet G, Jorm C, Roberts C, Gordon C, Chen T. Content validation of an interprofessional learning video peer assessment tool. BMC Med Educ. 2017;17:258.

16. Skinner K, Robson K, Vien K. Interprofessional education: a unique approach to addressing the challenges of student assessment. J Interprof Care. 2021;35(4):564–73.

17. Smeets HWH, Sluijsmans DM, Moser A, van Merriënboer JJ. Design guidelines for assessing students' interprofessional competencies in healthcare education: a consensus study. Perspect Med Educ. 2022;11(6):316–24.

18. Roberts SD, Lindsey P, Limon J. Assessing students' and health professionals' competency learning from interprofessional education collaborative workshops. J Interprof Care. 2018;33(1):1–9.

19. Frost JS, Hammer DP, Nunez LM, Adams JL, Chesluk B, Grus C, Harvison N, McGuinn K, Mortensen L, Nishimoto JH, et al. The intersection of professionalism and interprofessional care: development and initial testing of the interprofessional professionalism assessment (IPA). J Interprof Care. 2019;33(1):102–15.

20. Almoghirah H, Nazar H, Illing J. Assessment tools in pre-licensure interprofessional education: A systematic review, quality appraisal and narrative synthesis. Med Educ. 2021;55(7):795–807.

21. Dunston R, Forman D, Rogers GD, Thistlethwaite JE, Yassine T, Hager J, Manidis M, Rossiter C, Alliex S, Brewer ML, Buys N, Carr S, Gray J, Jones S, Kumar K, Matthews L, Moran M, Nicol P, Nicol P, ..., White J. Curriculum renewal for interprofessional education in health: Final report 2014. Commonwealth of Australia, Office for Learning and Teaching; 2014.

22. Reeves S, Fletcher S, Barr H, Birch I, Boet S, Davies N, McFadyen A, Rivera J, Kitto S. A BEME systematic review of the effects of interprofessional education: BEME guide No 39. Med Teach. 2016;38(7):656–68.

23. Hinyard L, Toomey E, Eliot K, Breitbach A. Student perceptions of collaboration skills in an interprofessional context: Development and initial

24. Hayes CA, Carzoli JA, LeRoy RJ. Development and implementation of an interprofessional team-based care rubric to measure student learning in interprofessional education experiences: a pilot study. J Interprofessional Educ Pract. 2018;11:26–31.

25. Hill AE, Bartle E, Copley JA, Olson R, Dunwoodie R, Barnett T, Zuber A. The VOTIS, part 1: development and pilot trial of a tool to assess students' interprofessional skill development using video-reflexive ethnography. J Interprof Care. 2023;37(2):223–31.

26. Earnest M, Madigosky WS, Yamashita T, Hanson JL. Validity evidence for using an online peer-assessment tool (CATME) to assess individual contributions to interprofessional student teamwork in a longitudinal team-based learning course. J Interprof Care. 2022;36(6):923–31.

27. Jorm C, Nisbet G, Roberts C, Gordon C, Gentilcore S, Chen T. Using complexity theory to develop a student-directed interprofessional learning activity for 1220 healthcare students. BMC Med Educ. 2016;16:199.

28. van Diggele C, Roberts C, Haq I. Optimising student-led interprofessional learning across eleven health disciplines. BMC Med Educ. 2021;21(1):1–18.

29. van Diggele C, Roberts C, Lane S. Leadership behaviours in interprofessional student teamwork. BMC Med Educ. 2022;22(1):1–7.

30. Lingard L. What we see and don't see when we look at 'competence': notes on a god term. Adv Health Sci Educ. 2009;14(5):625–8.

31. Sebok-Syer SS, Chahine S, Watling CJ, Goldszmidt M, Cristancho S, Lingard L. Considering the interdependence of clinical performance: implications for assessment and entrustment. Med Educ. 2018;52(9):970–80.

32. Roberts C, Wilkinson TJ, Norcini J, Patterson F, Hodges BD: The intersection of assessment, selection and professionalism in the service of patient care. Med Teach. 2018:1–6.

33. Roberts C, Howe A, Winterburn S, Fox N. Not so easy as it sounds: a qualitative study of a shared learning project between medical and nursing undergraduate students. Med Teach. 2000;22(4):386–91.

34. Pirrie V, Wilson RM, Harden JA. AMEE Guide No. 12: multiprofessional education: part 2 - promoting cohesive practice in health care. Med Teach. 2009;20(5):409–16.

35. Hall p. Interprofessional teamwork: professional cultures as barriers. J Interprof Care. 2005;19 Suppl 1(sup1):188–96.

36. Crossley J, Jolly B. Making sense of work-based assessment: ask the right questions, in the right way, about the right things, of the right people. Med Educ. 2012;46(1):28–37.

37. Thistlethwaite J. Interprofessional education: a review of context, learning and the research agenda. Med Educ. 2012;46(1):58–70.

38. Simmons B, Wagner S. Assessment of continuing interprofessional education: lessons learned. J Contin Educ Health Prof. 2009;29(3):168–71.

39. Chhabria K, Black E, Giordano C, Blue A. Measuring health professions students' teamwork behavior using peer assessment: validation of an online tool. J Interprofessional Educ Pract. 2019;16:100271.

40. Nicol D. Guiding principles for peer review: unlocking learners evaluative skills. In: Kreber C, Entwistle N, McArthur J, editors. Advances and Innovations in University Assessment and Feedback. Edinburgh: Edinburgh University Press; 2014.

41. Nofziger A, Naumburg E, Davis B, Mooney C, Epstein R. Impact of peer assessment on the professional development of medical students: a qualitative study. Acad Med. 2010;85:140–7.

42. Arnold L, Shue C, Kalishman S, Prislin M, Pohl C, Pohl H, Stern D. Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study. Acad Med. 2007;82(6):578–86.

43. Anderson ES, Ford J, Kinnair DJ. Interprofessional education and practice guide no. 6: developing practice-based interprofessional learning using a short placement model. J Interprof Care. 2016;30(4):433–40.

44. Visschers-Pleijers ASF, Dolmans DJM, Wolfhagen IAP, Vleuten CMV. Student perspectives on learning-oriented interactions in the tutorial group. Adv Health Sci Educ. 2005;10(1):23–35.

45. de Grave WS, Dolmans DH, van Der Vleuten CP. Student perceptions about the occurrence of critical incidents in tutorial groups. Med Teach. 2001;23(1):49–54.

46. Khoiriyah U, Roberts C, Jorm C, Van der Vleuten C. Enhancing students' learning in problem based learning: validation of a self-assessment scale for active learning and critical thinking. BMC Med Educ. 2015;15(1):140.

validation of the self-assessed collaboration skills instrument. Eval Health Prof. 2019;42(4):450–72.

Roberts *et al. BMC Medical Education*   (2024) 24:1307

Page 13 of 13

47. Kamp RJA, Dolmans DHJM, Van Berkel HJM, Schmidt HG. Can students adequately evaluate the activities of their peers in PBL? Med Teach. 2011;33(2):145–50.

48. Eva KW. Assessing tutorial-based assessment. Adv Health Sci Educ. 2001;6(3):243–57.

49. Sullivan ME, Hitchcock MA, Dunnington GL. Peer and self assessment during problem-based tutorials. Am J Surg. 1999;177(3):266–9.

50. Roberts C, Jorm C, Gentilcore S, Crossley J. Peer assessment of professional behaviours in problem-based learning groups. Med Educ. 2017;51:390–400.

51. Yang A, Brown A, Gilmore R, Persky AM. A practical review for implementing peer assessments within teams. Am J Pharm Educ. 2022;86(7):8795.

52. Cant RP, Cooper SJ. Simulation-based learning in nurse education: systematic review. J Adv Nurs. 2010;66(1):3–15.

53. Bussard ME. Self-reflection of video-recorded high-fidelity simulations and development of clinical judgment. J Nurs Educ. 2016;55(9):522–7.

54. Kelly M, Lyng C, McGrath M, Cannon G. A multi-method study to determine the effectiveness of, and student attitudes to, online instructional videos for teaching clinical nursing skills. Nurse Educ Today. 2009;29(3):292–300.

55. Jorm C, Roberts C, Gordon C, Nisbet G, Roper L. Time for university educators to embrace student videography. Camb J Educ. 2019;49(6):673–93.

56. Hoban G, Nielsen W, Shepherd A. Student-generated Digital Media in Science Education: Learning. Explaining and Communicating Content. London: Routledge; 2015.

57. Nielsen W, Georgiou H, Jones P, Turney A. Digital explanation as assessment in university science. Res Sci Educ. 2020;50:2391–418.

58. Shuldman M, Tajik M. The role of media/video production in non-media disciplines: the case of health promotion. Learn Media Technol. 2010;35(3):357–62.

59. Krull H. Video project brings new life to community engagement. J Nurs Educ. 2013;52(8):480–480.

60. Warren CM, Dyer A, Blumenstock J, Gupta RS. Leveraging Mobile Technology in a School-based participatory asthma intervention: findings from the Student Media-Based Asthma Research Team (SMART) study. Am J Health Educ. 2016;47(2):59–70.

61. Omar H, Khan SA, Toh CG. Structured student-generated videos for first-year students at a dental school in Malaysia. J Dent Educ. 2013;77(5):640–7.

62. Frenzel JE, Skoy ET, Eukel HN. Using student produced videos to increase knowledge of self-care topics and nonprescription medications. Curr Pharm Teach Learn. 2013;5(1):44–8.

63. Olson RE, Copley JA, Bartle E, Hill AE, Barnett T, Dunwoodie R, Zuber A. The VOTIS, part 2: using a video-reflexive assessment activity to foster dispositional learning in interprofessional education. J Interprof Care. 2023;37(2):232–9.

64. Lyons KJ, Giordano C, Speakman E, Smith K, Horowitz JA. Jefferson Teamwork Observation Guide (JTOG): an instrument to observe teamwork behaviors. J Allied Health. 2016;45(1):49-53C.

65. Liu Q, Geertshuis S, Gladman T, Grainger R. Student video production within health professions education: a scoping review. Med Educ Online. 2022;27(1):2040349.

66. Kane MT. Validating the interpretations and uses of test scores. J Educ Meas. 2013;50(1):1–73.

67. Cook DA, Brydges R, Ginsburg S, Hatala R. A contemporary approach to validity arguments: a practical guide to Kane's framework. Med Educ. 2015;49(6):560–75.

68. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166 e167–116.

69. Nisbet G, Gordon CJ, Jorm C, Chen T. Influencing student attitudes through a student-directed interprofessional learning activity: a pilot study. Int J Pract-based Learn Health Soc Care. 2016;4(1):1–15.

70. van Diggele C, Roberts C, Bloomfield J, Lane S. Interprofessional education: building social capital among faculty. Are we there yet? Focus Health Prof Educ: Multi-Prof J. 2024;25(1):92–109.

71. Thistlethwaite JE, Davies D, Ekeocha S, Kidd JM, MacDougall C, Matthews P, Purkis J, Clay D. The effectiveness of case-based learning in health professional education. A BEME systematic review: BEME guide no. 23. Med Teach. 2012;34(6):e421-444.

72. Cronbach LJ, Glaser GC, Nanda H, Rajaratnam N. The dependability of behavioural measurements: the theory of generalisability for scores and profiles. New York: John Wiley; 1972.

73. Crossley J, Russell J, Jolly B, Ricketts C, Roberts C, Schuwirth L, Norcini J. "I'm pickin" up good regressions': the governance of generalisability analyses. Med Educ. 2007;41(10):926–34.

74. Cookson J, Crossley J, Fagan G, McKendree J, Mohsen A. A final clinical examination using a sequential design to improve cost-effectiveness. Med Educ. 2011;45(7):741–7.

75. Connaughton J, Edgar S, Waldron H, Adams C, Courtney J, Katavatis M, Ales A. Health professional student attitudes towards teamwork, roles and values in interprofessional practice: the influence of an interprofessional activity. Focus Health Prof Educ: Multi-Discip J. 2019;20(1):8–18.

76. Lie DA, Richter-Lagha R, Forest CP, Walsh A, Lohenry K. When less is more: validating a brief scale to rate interprofessional team competencies. Med Educ Online. 2017;22(1):1314751.

77. Kamp RA, van Berkel HM, Popeijus H, Leppink J, Schmidt H, Dolmans DJM. Midterm peer feedback in problem-based learning groups: the effect on individual contributions and achievement. Adv Health Sci Educ. 2014;19(1):53–69.

78. van Zundert M, Sluijsmans D, van Merriënboer J. Effective peer assessment processes: research findings and future directions. Learn Instr. 2010;20(4):270–9.

79. Sluijsmans DMA, Brand-Gruwel S, van Merriënboer JJG. Peer assessment training in teacher education: effects on performance and perceptions. Assess Eval High Educ. 2002;27(5):443–54.

80. Parmelee D, Michaelsen LK, Cook S, Hudes PD. Team-based learning: a practical guide: AMEE guide no. 65. Med Teach. 2012;34(5):e275–87.

81. Brennan RL. (Mis) conception about generalizability theory. Educ Meas Issues Pract. 2005;19(1):5–10.

82. Shavelson RJ, Baxter GP, Gao X. Sampling variability of performance assessments. J Educ Meas. 1993;30(3):215–32.

Publisher's Note