Impact of video annotation on undergraduate nursing students' communication performance and commenting behaviour during an online peer-assessment activity

Chin-Yuan Lai
National Taichung University of Science and Technology, Taichung, Taiwan

Li-Ju Chen
Technology Education Team, Kaohsiung Compulsory Education Advisory Group, Taiwan

Yung-Chin Yen
Tainan First Senior High School, Tainan, Taiwan

Kai-Yin Lin
National Taichung University of Science and Technology, Taichung, Taiwan

This article reports on the implementation of a web-based video-annotation system that supports online peer-assessment activities in a nursing communication training scenario. A quasi-experimental design was applied to investigate the effects of using video annotation on communication skills and professional attitudes. The participants were fourth-year students from two classes at a nursing college in Taiwan. One class of 50 students served as the experimental group, who used the video-annotation tool we designed to add their comments to videos. The other class of 50 students served as the control group and used YouTube to add comments. Although YouTube also provides video-annotation features, these are not often used. Two rounds of peer-assessment activities indicated that the video-annotation tool notably enhanced nursing students' communication performance. Specifically, the tool was helpful in promoting students' development of communication skills, but not their professional attitudes. The students using the video-annotation tool provided more suggestions in their peer comments than those who did not use it. Moreover, video annotation resulted in closer agreement between peer and expert ratings of students' communication. The use of a video-annotation feature could improve the effectiveness of online peer assessment and thus promote student performance.

Implications for practice or policy:
- Nursing educators could adopt online peer review extensively in the training of nursing communication or other clinical skills.
- Student learning outcomes via online peer review can be improved by a video-annotation tool.

Keywords: nursing communication, nursing education, online peer assessment, peer feedback, video annotation

Introduction

As information technology has greatly progressed in recent decades, peer assessment and peer review have been integrated into web-based learning environments. Many benefits have been derived from the implementation of such tools. For instance, online peer assessment may increase interaction among students (Tseng & Tsai, 2007) and offer teachers the opportunity to monitor and guide students' interaction (Lai, 2016; Wu & Kao, 2008). It also provides options of an anonymous review process through which students may provide critical feedback (Lu & Bol, 2007) and rating scores with greater objectivity (Zhao, 1998). In addition, online peer feedback can strongly motivate students to revise their work more as compared to face-to-face feedback (Hewitt, 2000; Tsai & Chuang, 2013; Tuzi, 2004). However, although studies have claimed that online peer assessment can offer numerous educational benefits compared with traditional paper-based or face-to-face peer assessment, limited research on how to adopt effective methods to train students to provide higher-quality feedback is limited (Lai, 2016; Liou

& Peng, 2009). Multiple studies have indicated that the quality of peer feedback is crucial to the success of learning activities involving peer assessment (Lai, 2016; Smith, Cooper, & Lancaster, 2002; Topping, 1998; Xiao & Lucking, 2008).

A video annotation is a comment or explanation attached to a video. It could enable users to organise and contextualise notes more accurately than the traditional note-taking (Bargeron, Gupta, Grudin, & Sanocki, 1999) and reduce non-learning related cognitive loads (such as video control), thus helping students to focus more on video content and take better notes (Mu, 2010). Some researchers have combined video-annotation tools with online peer-assessment activities in their studies. Wu and Kao (2008) implemented a web-based peer-assessment system by using video-streaming technology to support the training of pre-service teachers. Their results indicated that participants were satisfied with the peer-assessment activities and considered video annotation a useful feature. Hulsman and van der Vloodt (2015) designed a web-based video-annotation system to support the training of medical students in communication skills. They indicated that this was a convenient tool for implementing their reflective practice activities, such as self-evaluation and peer feedback, into the classroom teaching of clinical skills. However, these studies focused on the effect of the peer-assessment process on students who used video annotation rather than on their performance and the quality of peer feedback. The current study was intended to fill this research gap.

Consequently, a web-based video-annotation system was designed to support online peer-assessment learning. It was specifically associated with the activities in a communication training course, the area in which video-based pedagogy and peer-assessment strategy are routinely used in teaching practice (Hammer, Fox, & Hampton, 2014). Although online peer-assessment activities have been conducted in various educational fields, they have seldom been carried out in nursing education or nursing communication courses (Lai, 2016). Moreover, few online peer-assessment studies have adopted a pedagogical strategy or tool to guide students on how to provide their peers with higher-quality feedback. This study aimed to enrich the literature in this field and answer the following research questions:

(1) Will using a video-annotation method to provide peer feedback have a greater impact on nursing students' communication performance than not using video annotation?
(2) What effect does using video annotation to provide peer feedback have on the quality of peer feedback?
(3) What effect does using video annotation to provide peer feedback have on the validity of peer ratings?

Related research

The following details literature reviews related to this study, including online peer assessment in education, the quality of peer feedback, and video annotation in education.

Use of online peer assessment in education

Numerous researchers have perceived the benefits of online peer assessment and implemented innovative methods in various educational fields. For example, Wu and Kao (2008) designed a web-based peer-assessment system to support the training of pre-service teachers. Xiao and Lucking (2008) used a wiki to provide an online collaborative learning environment that facilitated university students' academic writing performance. Liou and Peng (2009) used weblogs as a peer-assessment platform for English as a foreign language writing instruction. Liang and Tsai (2010) designed an online peer-assessment activity to help college students learn biology through writing. Hulsman and van der Vloodt (2015) implemented an online video-annotation system to enable medical students to review peers' communication videos and mark and annotate positive and negative events. Adwan (2016) used Google Forms for peer-assessment activities to improve group project outcomes among nursing students. Hsia, Huang, and Hwang (2016) proposed a web-based peer-assessment approach for conducting performing arts activities with junior high school students.

However, few of these studies have adopted further effective methods, apart from being online, to guide or train students in how to provide higher-quality feedback to their peers. Many researchers have claimed that the quality of peer feedback plays a crucial role in the success of peer-assessment learning activities.

For example, Xiao and Lucking (2008) found that peer feedback, in addition to ratings, could increase students' performance and their satisfaction with peer assessment. Smith et al. (2002) asserted that brief feedback, in addition to marking, could increase the transparency of the peer review process as well as raise student confidence and enhance their learning outcomes. Topping (1998) claimed that different types of feedback could have different effects on student learning. Liou and Peng (2009) explored the effect of relevant training on students' peer feedback (comments) and their revision quality. They found that the students who received training made more revision-oriented peer comments and had more success in revising their compositions. They concluded that peer review training is crucial to the effectiveness of computer-mediated peer review. Among the above researchers, only Liou and Peng (2009) offered training to their students to be more efficient in providing feedback. Consequently, this study integrated a video-annotation feature into an online video peer-assessment system to guide and train students in providing more beneficial feedback.

Factors related to the quality of peer feedback

The quality of peer feedback is key to effective peer-assessment activities. Researchers have attempted to identify the various types of peer comments that benefit learners (e.g., Cevik, Haslaman, & Celik, 2015; Cheng, Liang, & Tsai, 2015; Gielen, Peeters, Dochy, Onghena, & Struyven, 2010; Lin, Liu, & Yuan, 2001; Nelson & Schunn, 2009; Tseng & Tsai, 2007; Van der Pol et al., 2008). According to Lin et al. (2001), students receiving specific feedback significantly outperformed those who received holistic feedback. Tseng and Tsai (2007) indicated that reinforcing and suggestive peer feedback was more useful in aiding the development of student projects than didactic and corrective feedback. Reinforcing feedback was given by positive or supporting expression, while didactic or corrective feedback was presented in a traditional lecture-like tone. Van der Pol, van der Berg, Admiraal, and Simons (2008) addressed the significantly positive relationship between concrete-suggestion feedback and revising student work. Gielen et al. (2010) revealed that justified feedback (i.e., adding reasons while giving comments) can greatly improve students' writing skills and learning achievements. Furthermore, after analysing the feedback from 47 undergraduate students in a three-round online peer-assessment activity, Cheng et al. (2015) indicated that cognitive feedback (i.e., giving direct correction, personal opinion, or guidance) was more helpful for students' writing abilities than either affective (supporting or opposing comments) or metacognitive (evaluating or reflecting comments) feedback. In summary, these studies have indicated that specific, cognitive, justified, reinforcing, and suggestive feedback are beneficial for learning. We think these types of feedback are more constructive than just a simple positive or negative opinion, but not as critical as didactic or corrective comments. Therefore, they are more beneficial to students' learning.

Another indicator of the quality of peer feedback is the validity and reliability of peer assessment (Gielen et al., 2010). For example, when teacher ratings were used as empirical or external criteria, criterion-related validity (Thorndike & Hagen, 1977) was high if there was a high correlation between peer ratings and teacher or expert ratings. Many peer-assessment studies have adopted this measure to evaluate the quality of peer assessments. After they conducted an online peer-assessment activity involving 708 students across 16 courses, Cho, Schunn, and Wilson (2006) suggested that at least four peer-raters are required to ensure that online peer assessments are reliable and valid. Tseng and Tsai (2007) claimed that the marks awarded by peers mostly corresponded to those awarded by experts, after examining the validity of the peer assessment in a computer course involving 184 students. Bouzidi and Jaillet (2009) found, in a study with participation from 242 students enrolled in three courses, that the validity of peer assessment was equivalent to that of teacher assessment. In brief, most studies found that scores were consistent across peer and teacher assessments. However, after implementing a peer-assessment activity, Falchikov and Goldfinch (2000) found that peers had different views from teachers on items involving professional practice. Considering that the target of our study was peer assessment of communication in a care environment, which is categorised as professional practice, it was a challenge to obtain consistent results from teachers and peers.

Video annotation in education

Studies have shown that video annotation can have numerous pedagogical effects. For example, Bargeron et al. (1999) developed a collaborative video-annotation system called MRAS (Microsoft Research Annotation System), which enabled users to add video annotations, share commentaries, and link them.

They indicated that the MRAS could provide in-context personal notes and enable asynchronous collaboration among groups of users. Therefore, although the users spent extra time on note-taking via the MRAS, they preferred it over handwritten notes. Mu (2010) also developed a video-annotation system, Smartlink, which automatically generated hyperlinked timestamps between video content and associated notes. Mu discovered that students used Smartlink not only for personal notes but also for collaborative group discussion. Students using Smartlink took fewer notes and focused less on video controls and more on video content than those who did not, indicating Smartlink might mitigate cognitive load in video learning. Chiu et al. (2018) developed a video-annotation system to help learning within a cardiopulmonary resuscitation course. They found that video annotation could encourage students to concentrate more on the critical parts of instructional video and thus decrease extraneous cognitive load. In terms of peer-assessment activities, Wu and Kao (2008) integrated video annotation into online peer-assessment activities to support the training of pre-service teachers, who perceived it as a useful feature. Hulsman and van der Vloodt (2015) used an online video-annotation system to enhance medical students' communication skills and concluded it was a convenient tool for online self-evaluation and peer-feedback activities during clinical skills education.

Most of these studies focused on how the students used online video-annotation to take notes to help their learning from instructional video. Although some studies (such as Hulsman & van der Vloodt, 2015, & Wu & Kao, 2008) had integrated video-annotation into the peer-assessment activities, they did not use experimental design method to compare or verify the effect of video annotation on the peer-feedback quality and learning performance. For these reasons, we designed a video-annotation system as a peer-assessment tool and used quasi-experimental design to investigate its effect on student learning outcomes, the quality of peer feedback, and the validity of peer scores. We hypothesised that:

(1) Students who used video annotations to provide peer feedback would show significantly greater improvement in their communication performance.
(2) Students who used video annotation would provide more beneficial peer feedback to their peers.
(3) Students who used video annotation in providing peer feedback would demonstrate closer agreement between peer and expert ratings.

Methodology

In this study, video annotation was used to facilitate the learning process during the peer-assessment activities. From the perspective of cognitive load theory (Sweller, Van Merriënboer, & Paas, 1998), although making additional notes during peer assessment will take more mental effort, it is an effective cognitive strategy in helping learners to make better links between video content and comments, thus providing beneficial feedback. To reduce the mental effort, we designed the video-annotation system to enable the peer-assessment process to be synchronised with the viewing of peers' communication videos, which decreased extraneous cognitive load. A quasi-experimental design was used to investigate the effectiveness of utilising video annotation. The purpose of the experimental design was to determine whether an experimentally manipulated variable – using video annotation to provide peer feedback – would result in an improvement of the learning performance and the level of agreement between peer and expert ratings of communication performance. We also investigated whether the students who learned with video annotation provided more peer feedback, which would be beneficial to their learning.

Participants

The participants in this study were fourth-year student volunteers from two psychiatric nursing classes at a 5-year junior nursing college in Taiwan. The students were all female, and their average age was 19.2. Each class had 50 students. One class was assigned to the experimental group (G_{VA}) and used the video-annotation tool we designed in this study to associate their comments with the video content. The other class served as the control group (G_{non-VA}) and used YouTube to input their comments, without the aid of video annotation. Students from each group were assigned to 12 teams, with four or five students per team. All the learning activities were directed by the same course instructor, who holds a master's degree in nursing education and has taught nursing communication for more than 30 years. Ethical procedures were followed in accordance with the guidelines of the Research Ethics Framework of Society Institute in Taiwan (https://www.tsatw.org.tw/).

Instruments

Rubric for rating students' communication performance
In this study, students' communication performance was evaluated with the communication and consultation skills (CCS) scale developed and validated by Wiskin, Allan, and Skelton (2003). Ryan et al. (2010) used it in a communication training program for Irish medical and nursing students and indicated that it was a universally negotiated and global rating scale. A holistic and negotiated view was adopted to assess communication, but not in a systematic manner. The purpose of this view was to enable the students to develop their ability by reflecting on and adapting their communication styles to suit different individuals, and it was emphasised that communication is a process of meaningful negotiation and cannot be broken down into a series of competencies. With the holistic view, the CCS scale comprised only two dimensions: communication skills and professional attitudes, with a 5-grade rubric (see Appendix A). Communication skills include techniques such as questioning, eye contact, body posture, active listening, empathy, rapport building, and emotional responses. Professional attitudes involve professional, confident, and sincere behaviours. Students were graded on a scale of *A* (5 points) to *E* (1 point) on their communication performance, as evaluated by each criterion shown in Appendix A.

Each student's communication video was rated by three or four peers by using the CCS scale. Therefore, for each item on the CCS scale in each peer-assessment round, the average score awarded by these peers was calculated to represent the student's performance. Cronbach's alpha coefficient was used to assess the interrater reliability of peer scores. For two rounds of peer-assessment activity, the Cronbach's alpha coefficients for G_{VA} were 0.77 and 0.81, and those for G_{non-VA} were 0.72 and 0.85. These results indicated that the peer-assessment scores had adequate reliability.

Coding scheme of peer comments
Peer comments collected in this study were analysed based on the coding scheme developed by Hulsman, Harmsen, and Fabriek (2009). These researchers originally used the coding scheme to measure the level of students' reflection during a communication training program. Hulsman and van der Vloodt (2015) used it to analyse the characteristics of students' self-evaluations and peer-feedback annotations. This coding scheme included both retrospective and prospective dimensions. The retrospective dimension was related to "describing the key event" and was composed of behaviour, motive, and effect categories. The prospective dimension was related to "finding new solutions" and was composed of suggestion and goal categories, which were assumed to be comments that were beneficial for learning. Appendix B presents the description for each category, with examples. For instance, a comment was classified as "behaviour" if it described one specific behaviour. When the comment pointed out the reasons or motivation for a behaviour, it was assigned to the motive category.

When the peer comments were categorised, if a comment referenced more than one behaviour, it was segmented. In addition, some of the comments could be classified into more than one category. For example, the comment "Didn't clearly say one's name; client might confuse them with other nurses" was regarded as being in the behaviour category (Didn't clearly say one's name) as well as the effect category (Client might confuse them with other nurses). In this case, the feedback counted for both behaviour and effect. The categorisation process was conducted by two senior nursing teachers from the college: one with a master's degree in nursing education (i.e., the instructor) and the other with a doctoral degree in nursing informatics. All the comments were categorised independently by these two teachers and then discussed to resolve any potential discrepancies.

Nursing Communication Peer Assessment System
We designed a dedicated web-based system called the Nursing Communication Peer Assessment (NCPA) to enable students to assess their peers' communication efforts online. For these two groups (G_{VA} and G_{non-VA}) of students, the rating process was the same. As shown in Figure 1, the related descriptions of various grades in the CCS were translated into Chinese and shown on the scoring screen. Students could play or pause the video via the embedded player. They were also able to grade their peers' communication efforts according to the CCS scale.

Figure 1. User interface of the CCS rating for two groups (G_{VA} and G_{non-VA})

In addition to the quantitative scores, all students were asked to provide qualitative feedback (i.e., comments) to help their peers improve their communication. The difference between the two groups was that students in G_{VA} used the video-annotation feature to provide comments to their peers, whereas the students in G_{non-VA} did not. That is, in G_{VA}, students had to associate their comment with peers' video clips. When a student watched a video and wanted to comment on it, they had to first click one of the mark buttons (either positive or negative) to specify the type of comment and then write the comment. The system automatically created timestamp links between comments and the associated video clips. Therefore, once the student had received the comments and clicked on the hyperlink (see Figure 2), the video would move to that specific timestamp and start playing from there.

Figure 2. G_{VA} user interface for providing and receiving peer comments

Students in G$_{non-VA}$ did not need to correlate their comments with the video clips, and therefore used YouTube to directly provide positive or negative comments for their peers, using the "leave message" feature. Figure 3 is a screenshot of the messages posted by three students in G$_{non-VA}$. Although YouTube also provides a video-annotation feature for users to leave comments, we found that none of the students in G$_{non-VA}$ used YouTube's video-annotation tool.

Figure 3. G$_{non-VA}$ user interface for providing and receiving peer comments

Peer-assessment procedure

In this study, students were required to participate in two rounds of peer-assessment activities. Each round lasted 2 weeks. The first round was scheduled for Week 8 of the semester, and the second was held in Week 17. Each round comprised four steps: interview, peer rating and commenting, expert rating, and viewing and reflection (see Figure 4). Prior to the first round of peer assessment, the students participated in a 50-minute instructive course to understand the peer-assessment process, CCS rating scales, and detailed criteria for both quantitative and qualitative feedback. To avoid potential bias, each group was informed that the peer-assessed scores would not be included in their final course grades.
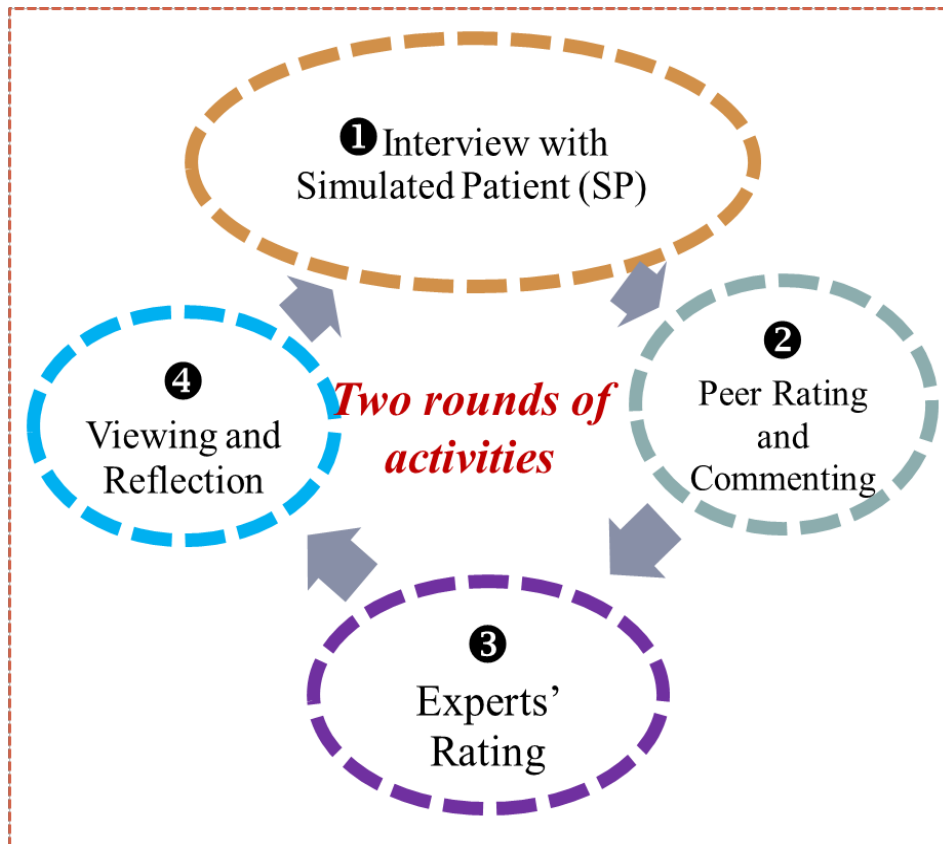
Figure 4. Process of the peer-assessment activity

First, during the interview step, each student was required to have a 4-minute discussion with the same simulated patient. Each student's communication scenario was video recorded and uploaded to YouTube. Then, students logged onto the NCPA and posted the URL of the individual video clip to the system. The students were briefed on how to perform these procedures.

Secondly, in the peer rating and commenting step, each student in the two groups was assigned three to four peers' video clips to rate their communication performance according to the CCS scale and provide qualitative comments. The rating process was the same for G_{VA} and G_{non-VA}. The only difference between these two groups was that the students in G_{VA} used the video-annotation tools (see Figure 2) to mark both positive and negative feedback when providing comments, whereas the students in G_{non-VA} used the comment tools provided by YouTube, as shown in Figure 3.

After all students finished the rating and commenting activity, two experts in psychiatric nursing cooperatively rated the students' communication scenarios in both groups, using the CCS scale. One expert, a senior nursing staff member with more than 25 years of practical experience in a psychiatric hospital, served as the SP in this study. The other expert was a retired nurse who had been in charge of a psychiatric nursing department for more than 10 years. To cooperatively rate each student's performance, these two experts observed the students' communication scenarios together, examined and discussed in detail the students' behaviour, and then issued their final scores.

Finally, in step 4, each student could view their own CCS scores (i.e., the average of the scores issued by three or four peers, not the individual score from each peer) and the comments from their peers. Students also received the CCS scores from the experts' rating at this time. Although students could instantly see the quantitative and qualitative feedback online, the whole process was not considered complete until this stage, because some students might not finish their peer evaluations until the last minute. Considering that all comments added to videos on YouTube are public and may be viewed by any other participant in G_{non-VA}, the interface for G_{VA} was designed to ensure that participants could view comments made by other participants in the same group. This design would not only maintain the consistency of user interfaces for

G_{VA} and G_{non-VA}, but also help participants learn from comments. Furthermore, each comment included the reviewer's name to strengthen students' sense of responsibility. The students agreed with these practices, after being told by the instructor. After viewing the feedback, students were asked to reflect on their communication skills by recording their thoughts, summarising their peers' comments, analysing the causes of any poor performance, and noting how they could improve in the future. The purpose of reflection was to promote students' self-awareness in communication learning and thus we did not assess their self-reflection.

Results

Effect of video annotation on student performance

The students' communication performances were rated by their peers and the two experts according to the CCS scale (1 to 5 points). Table 1 shows the descriptive statistics and *t* test results for the scores on communication skills and professional attitudes in the two rounds of peer-assessment activity. In the first round, the average CCS scores provided by experts ($t(98) = 1.15$, $p > 0.05$) and peers ($t(98) = -1.00$, $p > 0.05$) showed no significant difference between the two groups of students. However, in the second round, the expert CCS scores showed that both groups had improved, and the scores of G_{VA} were significantly higher than those of G_{non-VA}, $t(98) = 2.57$, $p < 0.05$. The results indicated that the video-annotation peer-feedback method greatly improved student communication performance, supporting the first hypothesis.

With respect to the *t* test of CCS subscales, the scores on communication skills in G_{VA} were also significantly higher than those in G_{non-VA} according to the expert ratings ($t(98) = 3.66$, $p < 0.001$). However, although the expert scores on professional attitudes for the G_{VA} group were higher than those for the G_{non-VA} group, there was no significant difference between the two groups ($t(98) = 0.55$, $p = 0.58$). The results revealed that using video annotation to provide peer feedback was especially helpful in promoting the development of communication skills.

Table 1
Students' CCS scores as rated by experts and peers (n = 100)

	Group	Score$_{expert}$			Score$_{peer}$		
		Mean	**SD**	***t***	**Mean**	**SD**	***t***
First round							
Communication skills	G_{VA}	2.2	0.5	1.56	3.7	0.4	-1.09
	G_{non-VA}	2.1	0.4		3.8	0.6	
Professional attitudes	G_{VA}	2.3	0.5	0.64	3.9	0.3	-1.10
	G_{non-VA}	2.2	0.4		4.0	0.5	
Total CCS	G_{VA}	4.5	1.0	1.15	7.6	0.6	-1.00
	G_{non-VA}	4.3	0.8		7.8	0.9	
Second round							
Communication skills	G_{VA}	3.1	0.7	3.66[**]	3.2	0.4	-6.47[**]
	G_{non-VA}	2.6	0.5		3.8	0.5	
Professional attitudes	G_{VA}	3.1	0.6	0.55	3.3	0.3	-9.47[**]
	G_{non-VA}	3.0	0.5		4.1	0.5	
Total CCS	G_{VA}	6.1	1.1	2.57[*]	6.5	0.7	-8.46[**]
	G_{non-VA}	5.7	0.7		7.8	0.9	

[*]$p < 0.05$, [**]$p < 0.001$

Although the expert scores showed that the two groups' communication performances improved in the second round, the peer scores showed conflicting results. Only the mean score for professional attitude in G_{non-VA} showed an increase in the second round. The mean scores for communication skills in the two groups and for professional attitude in G_{VA} were lower in the second round than in the first. We

speculated that this may have been related to the students' cognition deficiencies in therapeutic communication. In the beginning, the students seemed unaware that they had a particular deficiency in their communication with the psychiatric patients. Therefore, in the first round, for both G_{VA} and G_{non-VA}, the differences between student and expert scores were large. In the second round, although the differences between peer and expert scores in G_{VA} decreased, they still existed in G_{non-VA}. This explains why in the second round, student scores for G_{non-VA} were significantly higher than those for G_{VA} ($t(98) = -8.46$, $p < 0.001$), contrasting with the expert scores.

Effects of video annotation on the quality of peer feedback

Two rounds of comments were combined to compare the overall difference between the two groups' comments. A total of 1851 and 453 comments were posted for G_{VA} and $G_{non-VA,}$ respectively for the two rounds. Because some comments might reference more than one behaviour or could be assigned to more than one category, the total counts for G_{VA} and G_{non-VA} were 1875 and 877 respectively. The total frequency of peers' comments was much higher in G_{VA} than in G_{non-VA}.

Comments from peers were analysed using the five-category coding scheme of Hulsman and van der Vloodt (2015). Table 2 shows the descriptive statistics for the frequency and percentage of peers' comments in each category. In G_{VA}, the frequencies of the comments for both rounds in the five categories were, from highest to lowest, behaviour, suggestion, motive, effect, and goal. For G_{non-VA}, they were behaviour, motive, suggestion, effect, and goal. A χ^2 test of homogeneity of proportions was performed to examine whether the frequency pattern of the five categories exhibited any differences between the two groups. The result showed that the frequencies among the categories were significantly different ($\chi^2(4, 2752) = 39.98$, $p < 0.001$); therefore, the type of peer comment appeared to be affected by the use of video annotation.

Table 2
Five categories of peer comments

Category of peer comments	G_{VA}		G_{non-VA}	
	Frequency	Percentage within group	Frequency	Percentage within group
Behaviour	1373	73.2%	642	73.2%
Motive	159	8.5%***	118	13.5%***
Effect	63	3.4%*	44	5.0%*
Suggestion	228	12.2%***	63	7.2%***
Goal	52	2.8%**	10	1.1%**
Total	1875	100%	877	100%

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Post hoc analysis further showed that the percentage of peers' comments in the behaviour category was not significantly different between the two groups. The percentages of motive and effect comments in G_{VA} were significantly lower than those in G_{non-VA}. However, the percentages of peers' comments in the suggestion and goal categories in G_{VA} were significantly higher than those in G_{non-VA}.

Effect of video annotation on agreement between peer and expert scores

In this study, Pearson correlations of the CCS scores between the peer and expert ratings were conducted to determine how video annotation affects the validity of peer rating. Table 3 displays the correlation coefficients between peer and expert scores for communication skills and professional attitudes in each round. In the first round, there was no significant correlation between the two scores for G_{VA} ($r(48) = -0.06$, $p = 0.679$) and G_{non-VA} ($r(48) = 0.07$, $p = 0.629$). However, in the second round, there was a significant correlation between the two scores for G_{VA} ($r(48) = 0.49$, $p < 0.001$) and G_{non-VA} ($r(48) = 0.38$, $p < 0.01$), with the correlation coefficient in G_{VA} being higher than that of G_{non-VA}.

Table 3
Correlation between expert and peer CCS scores

Peer-assessment activity	Group	Communication performances		
		Communication skills	Professional attitudes	Total
First round	G_{VA}	-0.15	0.06	-0.06
	G_{non-VA}	0.10	0.03	0.07
Second round	G_{VA}	0.54***	0.22	0.49***
	G_{non-VA}	0.30*	0.36*	0.38**

*$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

With regard to the correlations of CCS subscales, in G_{non-VA}, the scores for communication skills and professional attitudes awarded by peers significantly correlated with those awarded by experts (communication skills, $r(48) = 0.30$, $p < 0.05$; professional attitudes, $r(48) = 0.36$, $p < 0.05$). However, in G_{VA}, the "professional attitude" item scores showed no significant correlation ($r(48) = 0.22$, $p = 0.124$), as peer and expert scores were only significantly correlated in the "communication skills" item ($r(48) = 0.54$, $p < 0.001$). This result was consistent with the findings that using video annotation to provide peer comments seemed especially useful in enhancing students' communication skills. Therefore, the correlation was particularly high.

In the second round, even though there were correlations between the peer and expert scores in the two groups, the peer scores of G_{VA} were closer to the expert scores than those of G_{non-VA}. Table 4 presents the descriptive statistics and *t*-test results of the average total CCS scores provided by peers and experts for both rounds. The results showed that the score for communication performance awarded by the students in G_{VA} was very close to that marked by the experts in the second round ($t(98) = 1.9$, $p = 0.066$). However, the students in G_{non-VA} significantly overestimated their communication performance in the two rounds (first round: $t(98) = 19.6$, $p < 0.001$; second round: $t(98) = 13.1$, $p < 0.001$). This finding revealed that the agreement between peer and expert scores in G_{VA} was closer than that for G_{non-VA}, which also supports the third hypothesis.

Table 4
T test of CCS scores rated by peers and experts

Peer-assessment activity	Group	$Score_{peer}$		$Score_{expert}$		
		Mean	SD	Mean	SD	*t*
First round	G_{VA}	7.6	0.6	4.5	1.0	17.7*
	G_{non-VA}	7.8	0.9	4.3	0.8	19.6*
Second round	G_{VA}	6.5	0.7	6.1	1.1	1.9
	G_{non-VA}	7.8	0.9	5.7	0.7	13.1*

*$p < 0.001$

Discussion and implications

In this study, an online video peer-assessment system integrating a video-annotation function was employed to support communication training for junior nursing college students. The purpose of this study was to facilitate peer feedback and improve student communication performance. We therefore investigated the effects of video annotation on the students' communication performance, the quality of peer comments, and the agreement between peer and expert scores in this online peer-assessment environment.

Unsurprisingly, the data showed that G_{VA} students exhibited greater improvement in their communication than G_{non-VA} students. The video-annotation tool was especially helpful in promoting the students' communication skills. We surmised that compared to professional attitudes, communication skills are easier for students to learn from each other because they involve external behaviours. When students used the video-annotation method to provide peer feedback, they had to be more deliberate because they

needed to provide evidence (i.e., based on the video clip) to support their comments. Therefore, the students could learn more from their peers in terms of their communication skills when they used the video-annotation tool.

Supporting our second hypothesis, the data indicated that the G_{VA} students provided more "suggestion" and "goal" categories of comments to their peers. According to the coding scheme of Hulsman et al. (2009), these two categories were related to "finding new solutions," as suggestive feedback, which was deemed beneficial for learning. This result was similar to that of Van der Pol et al. (2008), who found that more concrete suggestions were provided in an annotation system than in a Blackboard discussion forum. This result also concurred with Mu (2010)'s claim that the video-annotation system enabled the users to take more contextual notes. In our study, when the students used video annotation as a tool to provide peer feedback, they seemed to need to constantly rewatch the video because they associated peer feedback with the video's contents. Hence, when students repeatedly watched the videos, they were able to suggest alternatives to their peers, just like their instructors and the experts. Because they received more suggestive peer feedback, G_{VA} students showed greater improvement in their communication behaviour than their G_{non-VA} counterparts did. This is consistent with the findings of Tseng and Tsai (2007) and Van der Pol et al. (2008) that suggestive feedback is helpful for the improvement and revision of peer work. In addition, we found that video annotation appeared to motivate students to provide more comments. This result was similar to that of Bargeron et al. (1999), who found that students preferred the video-annotation feature to handwritten notes, although more time was required when using video annotation. We also speculated that because students were asked to associate their comments with the video content, they needed to be more deliberate in their commenting, which resulted in an increase in comments.

Likewise, data analysis revealed that peer scores in G_{VA} agreed more closely with expert scores. In the first round, for both G_{VA} and G_{non-VA}, the differences between student and expert scores were large. This finding is consistent with that of Falchikov and Goldfinch (2000), who found a difference between peer and expert scores for professional practice. In the second round, although the difference in scores between peers and experts in G_{non-VA} was still large, the difference in G_{VA} had decreased. The finding revealed that G_{VA} students seemed to more thoroughly understand their lack of communication competence with the psychiatric patients. That is, the results indicated that consistency between the scores provided by experts and peers could be enhanced through the use of video annotation. The third hypothesis was thus supported. Bransford, Brown, and Cocking (2000) believed that students had the most favourable outcomes when building on their own assessment skills while working with their peers. Sluijsmans, Brand-Gruwel, and Van Merrienboer (2002) supported this argument and indicated that peer-assessment tasks can be regarded as learning exercises in which assessment skills are practised. G_{VA} students appeared to have more opportunities to practise their assessment skills when using the video-annotation tool to provide peer feedback, and thus significantly improved their assessment skills of nursing communication. Therefore, the peer scores in G_{VA} were more consistent with the expert scores than those in G_{non-VA}. This revealed that the video-annotation tool was also useful in improving students' assessment skills. Chant, Jenkinson, Randle, Russell, and Webb (2002) indicated that this ability to distinguish good communication from bad is crucial for nursing communication training.

The findings of this study suggest that integrating the video-annotation tool into online peer assessments could help students enhance the quality of peer feedback, thus improving the validity of peer scores and student performances. Previous studies seldom used auxiliary strategies or tools to help students offer higher-quality feedback, and they did not verify the effect of video annotation on the quality of peer feedback. This study fills the research gap by using the video annotation mechanism as guidance to provide higher-quality feedback, and it proved to be effective. Liou and Peng (2009) provided students with teacher-designed handouts, illustrations, and commenting samples to train them in the techniques of offering meaningful feedback on their peers' essays. We believe that the video-annotation tool presented in the current study plays the same role in helping students make their comments. Although our study was not anonymous in the peer-assessment process, it did not appear to affect the strictness of peer feedback, because students had to give convincing evidence in their comments by using video annotation.

Mu (2010) found that the video-annotation system could automatically link notes with video content and thus reduce students' cognitive load by enabling them to take fewer notes and focus more on video content rather than video control. Chiu et al. (2018) had similar findings when using a video-annotation

system in health education. Both studies used video annotation as a note-taking tool to improve learning from instructional video. In this study, we extended an additional benefit of video annotation in education: when video annotation was integrated into online peer-assessment activities, it became an effective peer-assessment tool. Whether or not video annotation decreased students' extraneous cognitive load is uncertain since it was not evaluated in current study. However, based on our analysis and observations, it was obvious that G_{VA} group of students took more time in their peer-assessment work than G_{non-VA} group, similar to the findings of Bargeron et al. (1999). According to cognitive load theory (Sweller et al., 1998), we consider that our students' cognitive load was "germane" cognitive load rather than "extraneous" cognitive load. The differences between them lie in that: extraneous cognitive load is the effort associated with the way information or tasks are presented to a learner, and germane cognitive load refers to the work put into creating a permanent store of knowledge or schemas. Accordingly, the non-learning related cognitive load is frequently referred to as extraneous cognitive load; and germane cognitive load means learning-related cognitive load. Instructional designers should reduce extraneous cognitive load and redirect learners' attention to cognitive processes that are directly relevant to the construction of schemas (Sweller et al., 1998).

Conclusions

This study examined the effects of video annotation on nursing students' communication performance, the quality of peer comments, and the validity of peer scoring in an online peer-assessment environment. Students in the experimental group exhibited greater improvement in their communication than those in the control group. The findings also suggested that students in the experimental group provided more useful suggestions to their peers. Additionally, peer scores in the experimental group aligned more closely with expert scores. Therefore, adding the video-annotation tool to an online video peer-assessment system could increase the rigor of the peer-assessment process and promote students' learning performance. The video-annotation tool provided students with an investigative method, encouraging them to delve deeper into the pros and cons of their peers' work. After slight adjustments to the peer-assessment scales, this online video peer-assessment model could be transferred to other educational settings, such as surgical skills training, physical education, or other nursing skills training. These fields often use video-based pedagogy and peer assessment to enhance reflective practice. Its application in such areas may well offer starkly different and valuable learning methods for students.

Nevertheless, this study has a number of limitations that must be addressed. First, the conclusions are based solely on an analysis of junior nursing college students who used the online video peer-assessment system. Experimental results may differ with students in other age groups or studying different subject areas. Second, although this study originally had an experimental design, randomised controlled trials were not used, which may limit the argument that specific educational interventions have an impact on communication training (Grant & Jenkins, 2014). Third, this study adopted technology to improve the quality of peer feedback, which was assumed to be beneficial for learning, but it did not further examine the relationship between the quality of peer feedback and learning performance. Future studies could focus on other interventions, such as the design of various types of feedback buttons, such as "suggestion" or "question," to improve the quality of peer feedback, and on contributions to students' learning performance. To deepen understanding of the effects of different types of peer feedback on learning, researchers could use other coding schemes, such as the affective, cognitive, and metacognitive aspects of comments, as adopted by Cheng et al. (2015), to analyse the quality or attributes of peer feedback.

Acknowledgements

References

Adwan, J. (2016). Dynamic online peer evaluations to improve group assignments in nursing e-learning environment. *Nurse Education Today, 41*, 67–72. https://doi.org/10.1016/j.nedt.2016.03.028

Bargeron, D., Gupta, A., Grudin, J., & Sanocki, E. (1999). Annotations for streaming video on the Web: System design and usage studies. *Computer Networks, 31* (11–16), 1139–1153. https://doi.org/10.1016/S1389-1286(99)00058-4

Bouzidi, L., & Jaillet, A. (2009). Can online peer assessment be trusted? *Journal of Educational Technology & Society, 12*(4), 257–268. Retrieved from https://www.j-ets.net/collection/published-issues/12_4

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press. https://doi.org/10.17226/9853

Cevik, Y. D., Haslaman, T., & Celik, S. (2015). The effect of peer assessment on problem solving skills of prospective teachers supported by online learning activities. *Studies in Educational Evaluation, 44*, 23–35. https://doi.org/10.1016/j.stueduc.2014.12.002

Chant, S., Jenkinson, T., Randle, J., Russell, G., & Webb, C. (2002). Communication skills training in healthcare: A review of the literature. *Nurse Education Today, 22*, 189–202. https://doi.org/10.1054/nedt.2001.0690

Cheng, K.-H., Liang, J.-C., & Tsai, C.-C. (2015). Examining the role of feedback messages in undergraduate students' writing performance during an online peer assessment activity. *The Internet and Higher Education, 25*, 78–84. https://doi.org/10.1016/j.iheduc.2015.02.001

Chiu, P.-S., Chen, H.-C., Huang, Y.-M., Liu, C.-J., Liu, M.-C., & Shen, M.-H. (2018). A video annotation learning approach to improve the effects of video learning. *Innovations in Education and Teaching International, 55*(4), 459–469. https://doi.org/10.1080/14703297.2016.1213653

Cho, K., Schunn, C., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology, 98* (4), 891–901. https://doi.org/10.1037/0022-0663.98.4.891

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*, 287–322. https://doi.org/10.3102/00346543070003287

Gielen, S., Peeters, E., Dochy, F., Onghena, P., & Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction, 20*, 304–315. https://doi.org/10.1016/j.learninstruc.2009.08.007

Grant, M. S., & Jenkins, L. S. (2014). Communication education for pre-licensure nursing students: Literature review 2002-2013. *Nurse Education Today, 34*(11), 1375–1381. https://doi.org/10.1016/j.nedt.2014.07.009

Hammer, M., Fox, X., & Hampton, M. D. (2014). Use of a therapeutic communication simulation model in pre-licensure psychiatric mental health nursing: enhancing strengths and transforming challenges. *Nursing and Health, 2*(1), 1–8. https://doi.org/10.13189/nh.2014.020101

Hewitt, B. L. (2000). Characteristics of interactive oral and computer-mediated peer group talk and its influence on revision. *Computers and Composition, 17*(3), 265–288. https://doi.org/10.1016/S8755-4615(00)00035-9

Hsia, L.-H., Huang, I., & Hwang, G.-J. (2016). A web-based peer-assessment approach to improving junior high school students' performance, self-efficacy and motivation in performing arts courses. *British Journal of Educational Technology, 47*(4), 618–632. https://doi.org/10.1111/bjet.12248

Hulsman, R. L., & van der Vloodt, J. (2015). Self-evaluation and peer-feedback of medical students' communication skills using a web-based video annotation system. Exploring content and specificity. *Patient Education and Counseling, 98*(3), 356–363. https://doi.org/10.1016/j.pec.2014.11.007

Hulsman, R. L., Harmsen, A. B., & Fabriek, M. (2009). Reflective teaching of medical communication skills with DiViDU: assessing the level of student reflection on recorded consultations with simulated patients. *Patient Education and Counseling, 74*, 142–149. https://doi.org/10.1016/j.pec.2008.10.009

Lai, C.-Y. (2016). Training nursing students' communication skills with online video peer assessment, *Computers & Education, 97*, 21–30. https://doi.org/10.1016/j.compedu.2016.02.017

Liang, J. C., & Tsai, C. C. (2010). Learning through science writing via online peer assessment in a college biology course. *The Internet and Higher Education, 13*(4), 242–247. https://doi.org/10.1016/j.iheduc.2010.04.004

Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of Computer Assisted Learning, 17*, 420–432. https://doi.org/10.1046/j.0266-4909.2001.00198.x

Liou, H. C. & Peng, Z. Y. (2009). Training effects on computer-mediated peer review. *System, 37*(3), 514–525. https://doi.org/10.1016/j.system.2009.01.005

Lu, R. L., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning, 6*(2), 100–115. Retrieved from https://www.ncolr.org/jiol/issues/pdf/6.2.2.pdf

Mu, X. (2010). Towards effective video annotation: An approach to automatically link notes with video content. *Computers & Education, 55* (4), 1752–1763. https://doi.org/10.1016/j.compedu.2010.07.021

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401. https://doi.org/10.1007/s11251-008-9053-x

Ryan, A., Walshe, N., Gaffney, R., Shanks, A., Burgoyne, L., & Wiskin, C. M. (2010). Using standardized patients to assess communication skills in medical and nursing students. *BMC Medical Education, 10*, art. 24. https://doi.org/10.1186/1472-6920-10-24

Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merrienboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education, 27*, 443–454. https://doi.org/10.1080/0260293022000009311

Smith, H., Cooper, A., & Lancaster, L. (2002). Improving the quality of undergraduate peer assessment: A case study from psychology. *Innovations in Education and Teaching International, 39*, 71–81. https://doi.org/10.1080/13558000110102904

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Thorndike, R. L., & Hagen, E. P. (1977). *Measurement and education in psychology and education* (4th ed.). New York, NY: John Wiley & Sons. https://doi.org/10.1177/001316447903900141

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*, 249–276. https://doi.org/10.3102/00346543068003249

Tsai, Y. C., & Chuang, M. T. (2013). Fostering revision of argumentative writing through structured peer assessment. *Perceptual & Motor Skills, 116* (1), 210–221. https://doi.org/10.2466/10.23.PMS.116.1.210-221

Tseng, S. C., & Tsai, C. C. (2007). On-line peer assessment and the role of the peer feedback: A study of high school computer course. *Computers & Education, 49 (4)*, 1161–1174. https://doi.org/10.1016/j.compedu.2006.01.007

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition, 21*(2), 217–235. https://doi.org/10.1016/S8755-4615(04)00008-8

Van der Pol, J., van der Berg, B. A. M., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*, 1804–1817. https://doi.org/10.1016/j.compedu.2008.06.001

Wiskin, C. M., Allan, T. F., & Skelton, J. R. (2003). Hitting the mark: negotiated marking and performance factors in the communication skills element of the VOICE examination. *Medical Education, 37*(1), 22–31. https://doi.org/10.1046/j.1365-2923.2003.01408.x

Wu, C.-C., & Kao H.-C. (2008). Streaming videos in peer assessment to support training pre-service teachers. *Journal of Educational Technology & Society, 11*(1), 45–55. Retrieved from https://www.j-ets.net/collection/published-issues/11_1

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education, 11*(3-4), 186–193. https://doi.org/10.1016/j.iheduc.2008.06.005

Zhao, Y. (1998). The effects of anonymity on computer-mediated peer review. *International Journal of Educational Telecommunications, 4*(4), 311–345. Retrieved from https://www.learntechlib.org/j/IJET/

Corresponding author: Chin-Yuan Lai, yuanlai@nutc.edu.tw

Appendix A. CCS scale (adopted from Ryan et al., 2010, pp. 2-3)

Grade	Communication skills	Professional Attitudes
A	Excellent use of appropriate questioning styles - including effective information gathering. Appropriate levels of eye contact and body posture. Excellent active listening. Clear understanding of demonstrating empathy, rapport-building and acknowledgement of emotional responses.	Students at this level are likely to appear highly professional, confident and sincere. Advice, when offered, will be given appropriately and responsibly. Students interacting with the role-player will demonstrate good levels of respect, and show no signs of prejudice. Students will engage responsibly with the screening process.
B	Generally appropriate use of question styles. Students will use eye contact and body language appropriately and in an engaged manner. Students will use a variety of questions to gather information. They will listen actively, demonstrate empathy consistently, and handle emotion in the role-play in a generally appropriate manner.	Students at Grade B demonstrate professionalism and sincerity, although may occasionally appear uncertain. Advice, if offered, will mostly be appropriate and responsible. Students will be respectful and demonstrate attitudes free of prejudice. Students will engage with the screening process and generally answer questions appropriately.
C	Although students may occasionally be inconsistent or erratic, they will gather information adequately and attempt different questioning styles. Although they may occasionally interrupt inappropriately, they generally listen well, Eye contact and body language will mostly be appropriate, with occasional inconsistencies. Students will demonstrate some empathy and respond to emotion, but perhaps in a clumsy fashion or limited way.	Self-presentation is adequate. Advice, if offered, may at times be inappropriate or appear slightly irresponsible, but this is more likely due to lack of insight than an intention to mislead the role-player. Role-players will be treated respectfully, although clumsiness from the student may result in the role-player occasionally feeling uncomfortable. Students will take the occasion seriously.
D	Limited range of question styles, or erratic and confused questions students may not listen well, interrupting and impeding role-player concerns. Eye contact and body language may not be appropriate to the content of the encounter. There may be little or no demonstration of empathy, little or no response to patient emotion, and inadequate reflection. Skills used within the encounter will neither be appropriately contextualized, nor meaningfully demonstrated (for example a warm smile during the breaking of bad news).	Students at this grade will appear unprofessional and may appear to lack sincerity. They may give the impression of a lack of care or interest in the role-play. Advice, if offered, will likely be inappropriate and/or poorly presented. Student responses may sometimes appear stigmatising.
E	Students at this level will have serious deficiencies in their skill set. They will demonstrate few if any of the following skills: eye contact, body language, active listening, demonstrating empathy and responding to patient emotion. They will demonstrate little understanding of the context.	Students will have serious difficulties presenting a professional manner. Little interest in the scenario, role-player, or assessor. Students are likely to demonstrate an uncaring interpersonal style that could be interpreted as arrogant, prejudicial or stigmatising.

Appendix B. Coding scheme for analysis of peers' comments (adapted from Hulsman & van der Vloodt, 2015, p. 358)

Dimension	Category	Definition	Examples[a]
Retrospective categories:	Behavior	A reference is made to a specific behavior.	• "Repeat client's reply" • "Speak before client finish talking"
	Motive	A reason is provided for the behavior. Reasons can be refer-ring to personal intentions, to the situation, or to patient behaviors	• "*Due to nervousness*, appear tight" • "*Put oneself in client's shoes*, express concern"
	Effect	The consequence of the behavior is described. Consequences can be referring to personal goals, to the situation, or to patient behaviors.	• "Shouldn't ask "why" *that makes client's feeling accused*" • "Didn't clearly say one's name; *client might confuse them with other nurses*"
Prospective categories:	Suggestion	A specific suggestion for an alternative behavior is provided.	• "Could observe client before asking question" • "Before the end of last meeting, should make a conclusion and express one's feelings"
	Goal	The proposed consequence of the alternative behavior is described. Goals can be referring to personal goals, to the situation, or to patient behavior.	• "Could talk with client about other topics *to prevent her from obsessed with illusions*" • "Should have facial expressions, *that looks more sincere*"

[a]Examples were taken from participants' comments.