The effect of in-class and one-on-one video feedback on EFL learners' English public speaking competency and anxiety

Tingting Liu
Sichuan International Studies University
National Institute of Education, Nanyang Technological University

Vahid Aryadoust
National Institute of Education, Nanyang Technological University

Video feedback (VF) is a useful learning technique in acquiring public speaking skills due to its capacity to retain both verbal and non-verbal elements of multimodal communication. Previous research has focused on video self-critique, peer VF, online feedback, or one-on-one VF in the first language (L1) context and has yielded varied results regarding the impact of VF on public speaking competence and anxiety. Therefore, this quasi-experimental classroom-based study compared the impact of the one-on-one VF and in-class VF on public speaking competency and anxiety of learners of English as a Foreign Language (EFL). Three intact university-level English classes in China (n = 74) were assigned to the three conditions (in-class VF, one-on-one VF, and verbal feedback). Data on students' speaking competence were validated using many-facet Rasch measurement (MFRM). Subsequent gain score analysis and ANCOVA showed that in-class VF significantly improved students' delivery skill and global competence and reduced their speaking anxiety compared with the verbal feedback group. It is proposed that in-class VF be employed as an instructional procedure to help EFL learners improve their public speaking skills and reduce their public speaking anxiety, particularly in the teaching context of a large class size together with comparatively limited logistic and teaching resources.

Email address for correspondence: vahid.aryadoust@nie.edu.sg

Key words: public speaking competency; public speaking anxiety (PSA); video feedback; formative assessment; oral presentation; many-facet Rasch measurement (MFRM)

Introduction

The last two decades have witnessed a rise in the offering of English public speaking courses in tertiary education in context of English as a foreign language (EFL). In teaching public speaking/oral presentation skills to students, feedback and assessment constitute an essential part of the learning cycle. The available research on feedback in public speaking has focused on the effect of different feedback sources on acquisition of speaking skills (De Grez et al., 2009; van Ginkel et al., 2017), development and validation of public speaking assessment instruments (Morreale et al., 2011; Schreiber et al., 2012; Thomson & Rucker, 2002), validation of self-assessment or peer-assessment of speech (Aryadoust, 2015; De Grez et al., 2012), application of technology in providing feedback such as offering feedback on blogs (Hung, 2011; Shih, 2010; Xu et al., 2017), and video feedback (VF) (Christianson et al., 2009; Hinton & Kramer, 1998).

VF refers to a "structured process whereby students review their recorded communication with the benefit of some guidance and/or evaluation from an instructor or peers" (Quigley & Nyquist, 1992, p. 325). Research on the effect of VF on students' acquisition of public speaking skills is inconclusive. In a recent systematic review, Miskam and Saidalvi (2020) found that most research applied mixed-method or descriptive survey designs and they suggested more complex research designs with comparison groups should be used to investigate whether video technology can result in statistically significant improvement in students' public speaking competence. The extant studies that have used comparison group designs mainly investigated self-regulated video watching, peer VF, tutor-guided one-on-one VF, or online feedback (Barry, 2012; Christianson et al., 2009; De Grez et al., 2009; Deihl et al., 1970; Hinton & Kramer, 1998; Hung, 2011). By contrast, there is a dearth of research on teacher-directed VF in a plenary class as an instructional procedure. In-class VF refers to a situation wherein the teacher records students' speeches delivered in class and then comments on their speeches when playing the speech videos to the plenary class to emphasize strengths and weaknesses as well as offering suggestions to foster improvement. Observing social models constitutes the first step toward acquiring complex social behaviors (Wouters et al., 2007) and feedback was found to be the most fruitful moderator in improving students' performance (Hattie & Timperley, 2007). Therefore, we argue teacher-directed in-class VF may offer valuable learning

opportunities to both the speaker and students in the audience. However, it is still unclear how the in-class VF impacts learners' public speaking competency.

Another construct of in-class VF instruction is public speaking anxiety (PSA), as any learning activity design should consider its emotional impact. PSA is defined as "a situation-specific social anxiety that arises from the real or anticipated enactment of an oral presentation" (Bodie, 2010, p. 27). A high level of anxiety is considered to have a debilitative effect, including decreased autonomy of learning and participation in the class activity (C.-M. Chen & Lee, 2011). EFL learners may experience a higher level of anxiety because of general EFL classroom anxiety (Horwitz et al., 1986). VF has long been used as an anxiety therapy among socially or speech-anxious subjects in clinical settings (e.g., Rodebaugh, 2004; Rodebaugh et al., 2010) or in lab settings in the L1 context (e.g., Hinton & Kramer, 1998). However, few studies have examined the impact of VF as an instructional procedure on PSA in an authentic EFL classroom.

To address the research gaps, the present study adopts a quasi-experimental design to investigate the impact of a teacher's in-class VF on students' public speaking competency and anxiety.

The impact of VF on public speaking competence

Although oral presentation has traditionally been an intrinsic part of communication courses, the lack of self-viewing opportunities and teacher feedback reduces the effectiveness of these courses (Simpson et al., 2019). The video makes it possible for student speakers to view their speeches from the perspective of a listener, compare performances, and identify areas for improvement (Christianson et al., 2009; Quigley & Nyquist, 1992).

In utilizing video recording in the feedback of public speaking competence, four types of application emerge from the research: self-viewing, self-critique, peer VF, and teacher VF. Despite its effectiveness in boosting students' confidence and projecting a positive self-image (Hinton & Kramer, 1998), the mere replaying of their speaking performance did not seem to result in significant improvements in the students' speaking performance (Bankston & Terlip, 1994) or their self-perceived communication competence (Dupagne et al., 2007; Hinton & Kramer, 1998). A few studies have investigated the effect of critiquing one's performance while watching one's own videos through filling in reflection forms and questionnaires. Compared to the conventional verbal feedback, students commented that conducting self-critique while watching one's videos helped them better identify their strengths and weaknesses (Barry, 2012; Christianson et al., 2009; Hung, 2011). It also resulted in

improvements in their language and delivery (Miles, 1981) and structural development (Glenn, 1996).

Peer VF is viewed by students as a source for diverse perspectives and motivation for learning. Hung (2011) investigated how students perceived using speech Vlogs, where they provided feedback on each other videos and then submitted a self-critique. Students commented that "peer feedback points out some problems in my Vlog that can hardly be found by myself" (p. 743). However, Nikolic et al. (2018) found peer feedback did not result in significant improvement compared with the condition where students simply made a self-critique or received no feedback. Additionally, some students felt uncomfortable and exposed when commenting on their peers (Simpson et al., 2019).

Christianson et al. (2009), in their investigation of students' attitudes toward peer VF and video self-critique, reported that some students perceived self-critique to be difficult to conduct and peer feedback shallow in nature. While students find peer feedback to be facilitative to their learning, they value teachers' comments the most (Ng, 2014). Comparative studies revealed that VF offered by an experienced observer or instructor was more effective in enhancing students' performance than students watching their videos, even with self-critique, and thus a combination of practice, self-critique, and teachers' VF was recommended (Deihl et al., 1970; Mills & Pace, 1989). Research also showed that compared with self-feedback and peer feedback, teacher feedback was more useful in enhancing students' speaking competence (De Grez et al., 2009) and more effective in encouraging students' speaking behavior (van Ginkel et al., 2017).

Despite the importance of teacher feedback, studies on teacher-directed VF are rather limited. The existing studies on the impact of teacher VF are limited to lab settings where tutors provide one-on-one feedback while watching the videos with the speaker (e.g., Deihl et al., 1970) or they post online written feedback on the videos (e.g., Shih, 2010). However, we argue that timely VF in a plenary class can compensate for reduced class hours and teacher-student interactions as well as opportunities to learn from comments on peers' speeches. Given these benefits, it is important that teacher feedback, as a valuable but rare source, should be leveraged and optimized in language learning classes.

Additionally, many previous studies on the effect of VF on students' speaking competence used students' self-report rather than professional measures (e.g., Dupagne et al., 2007; Hinton & Kramer, 1998). While self-report provides insights on how communicatively competent speakers perceive themselves to be, the validity of

one's perception of one's public speaking competence as a measure of the effectiveness of VF is subject to question (McCroskey & McCroskey, 1988). An individual's self-evaluation can be inaccurate, particularly as a novice public speaker (LeFebvre et al., 2015). Additionally, speech-anxious speakers tend to underrate their competence (Rodebaugh & Rapee, 2005).

Another feature of recent studies on VF innovations is the use of online platforms for video blogs where students upload their self-recorded speech videos (Hung, 2011; Shih, 2010). Despite the opportunity to reexamine their work and receive feedback from their peers and teachers, students in Hung's (2011) study found repetitive recording and uploading of the videos time-consuming and anxiety-provoking. Additionally, these videos are often shot in the absence of a real audience and recorded or edited multiple times, whereas speeches on most occasions are delivered in the presence of an authentic audience in one shot. Thus, one can be concerned that prerecorded videos may not be a true representation of the speaker's public speaking competence.

Impact of VF on public speaking anxiety (PSA)

The most frequent negative consequences of high levels of PSA include avoidance of public speaking occasions, which in turn limits one's development in academic and professional pursuits (McCroskey, 1977). Bodie (2010) summarized three types of PSA treatment, including (i) systematic desensitization through exposure, (ii) cognitive modification of biased self-imagery, and (iii) skill training through performance feedback and specially designed courses. Exposure therapy can be conducted through in vivo exposure (an authentic physical situation), in vitro exposure (a situation imagined by patients), and virtual reality (VR) exposure therapy (a situation created by VR) (Ling et al., 2012).

Regarding cognitive modification, VF is among the most applied cognitive therapies. The most basic level of VF therapy is self-confrontation, the act of watching one's speaking performance in a video. In a study on fourth-year medical students, Schmidt et al. (2014) found that significantly less students with a higher level of anxiety were observed in the VF therapy group compared to the control group. Hinton and Kramer (1998) reported that self-confrontation resulted in a significant reduction in PSA among students with high-level and low-level anxiety but not among students with moderate to high levels of anxiety. However, no significant reduction in students' PSA was identified in Dupagne et al. (2007), Newburger et al. (1994), and Hallmark et al. (1993).

Another stream of VF therapy studies has investigated the moderating effect of cognitive preparation on PSA of socially anxious subjects. Harvey et al. (2000) suggested that cognitive preparation enhances the therapeutic effects of VF in reducing the discrepancy between self and video images. Cognitive preparation, conducted before watching one's video, includes (i) a detailed prediction of one's performance in the video, (ii) a projection of self-image in the mind, and (iii) a review of one's actual performance in the video as an observer, which aims to reduce self-focused attention during VF and demonstrate to speakers the extent of their biased self-imagery (Harvey et al., 2000). Likewise, Rodebaugh et al. (2010) also identified a significant reduction of anticipatory PSA among clinically anxious subjects with a high self-observer discrepancy in addition to improved self-perception of speaking performance. In a recent study, Chen et al. (2018) compared four combinations of VF with audience feedback, cognitive preparation, and cognitive review and found socially anxious students who received all four interventions reported a significant decrease in their PSA and a significant increase in their positive perception of speech performance compared to other groups.

Despite the reported effectiveness of VF in reducing PSA, previous studies were mostly conducted in clinical or lab settings in L1 contexts. However, to EFL learners, the act of speech-making in the EFL classroom may induce the three components of foreign language classroom anxiety (communication apprehension, test anxiety, and fear of negative comments) (Horwitz et al., 1986). This anxiety can interfere with language learners' cognitive processes, including input, processing, and output abilities (MacIntyre & Gardner, 1994), but also their feedback intake, which aims at feeding forward their future speech making. However, it is still unclear how students would respond to being exposed to public speaking situations and being openly evaluated by the instructor in EFL classrooms.

The present study

Several research gaps emerged in the literature review. Firstly, most previous studies were on self-regulated video watching, peer VF, tutor-guided one-on-one VF, or online feedback (Barry, 2012; Christianson et al., 2009; De Grez et al., 2009; Deihl et al., 1970; Hinton & Kramer, 1998; Hung, 2011). However, given the value and sustainability of teacher feedback, it is important to investigate how in-class VF could impact students' public speaking competence. Secondly, research has investigated the provision of feedback through online platforms such as the non-face-to-face written and verbal feedback to students' pre-recorded speeches on an online platform (Hung, 2011; Shih, 2010; Xu et al., 2017). While these innovations have extended our understanding of VF, in-class VF can enable teachers and students to conduct more

interactive communication where they could confirm and clarify points immediately. Thirdly, previous research has used self-assessments or self-reports as a measure of one's public speaking competence (e.g., Dupagne et al., 2007; Hinton & Kramer, 1998). However, students' self-assessment may be an inadequate and biased measure of their public speaking competence compared to the validated measures offered by professional raters (Aryadoust, 2015).

Therefore, the first objective of the present study is to investigate the impact of in-class VF on students' public speaking competence measured with validated ratings by professional raters. Teacher's comment is highly valued and is viewed as a more informative and authoritative source of evaluation, especially when novice speakers need scaffolding in understanding the rubrics and resolving disputes. Additionally, in-class VF allows students to have in vivo exposure to real-time speech in the presence of a real audience, which resembles most public speaking occasions. Many-facet Rasch measurement was used to validate ratings of the students' speech performance, as applied in previous studies of rater behavior (Knoch et al., 2018; Koizumi et al., 2019; Youn, 2018).

The second aim of the study is to examine the impact of in-class VF on students' PSA. Evaluation of students' speeches openly while playing their speech videos may expose their vulnerabilities and induce anxiety (McCroskey & Lashbrook, 1970), and thus hinder their uptake of VF. Additionally, previous research yielded varied results in terms of the impact of VF on speakers' PSA in clinical or lab settings in L1 contexts (e.g., Dupagne et al., 2007; Hinton & Kramer, 1998; Rodebaugh, 2004), whereas it is still unclear how in-class VF would impact the students' PSA in an EFL classroom context.

Therefore, the present study seeks to answer the following two research questions:

Q1: How do in-class VF and one-on-one VF affect students' public speaking competence?

Q2: How do in-class VF and one-on-one VF affect students' PSA?

Methodology

Participants

The participants in the quasi-experiment were 74 students (male = 15; female = 59) from three intact classes of the second-year English majors who represent the

intermediate to high-intermediate level of English proficiency among learners of EFL in a university in China. As the participants were enrolled in the same compulsory course as English majors, their performance would be minimally influenced by work in other courses. The participants were 19 to 20 years old, with 80% female and 20% male, which is the typical composition of English majors in China.

Instruments

Assessment tool for public speaking competency

The assessment scale, Public Speaking Competency Instrument (PSCI) (Thomson & Rucker, 2002), is one of the well-established rubrics for assessing public speaking competence. The PSCI is a 20-item measurement tool with a 5-point scale per item (the total ranging from the lowest possible 25 points to the highest possible 100 points). The rating scale has five subscales: introduction (items 1-3, clarification of the speech purpose), body (items 4-8, sufficiency and organization of the supporting materials), conclusion (items 9-12, review of the speech purpose and key points), delivery (items 13-19, use of the voice and body), and global competence (item 20, global assessment of the speaker's competence). This assessment scale has proved to be highly reliable with Cronbach's Alpha > 0.8 for 4 out 5 subscales and the factor analysis showed the general factor explained 32.2% of the variance (Thomson & Rucker, 2002). Additionally, both the PSCI and the public speaking skills covered in the course instructions are based on Lucas (2001), which is among the most widely used English public speaking teaching materials in L1 and EFL contexts (Bodie, 2010; Junko & Heffernan, 2008; Schreiber et al., 2012). Therefore, the rating scale is suitable for this study as the principles of public speaking should be applicable in this context.

Assessment tool for PSA

Personal Report of Confidence as a Speaker (PRCS) (Paul, 1966) is one of the best-validated public speaking anxiety assessment scales. PRCS, which consists of 30 items in a True-False format, assesses the behavioral and affective response to public speaking situations. The scores range from 0 (i.e., no fear of public speaking) to 30 (i.e., the highest level of fear). PRCS has been widely employed in research studies and treatments which have shown it to have reasonable internal consistency (Daly, 1978; Klorman et al., 1974). It has been used in both L1 contexts (Hancock et al., 2010) and L2 or EFL contexts (Alek et al., 2020; L. H. Chen & Wang, 2015), and was found to be reliable in the present study as a pretest (30 items; $\alpha = 0.703$) and posttest (30 items; $\alpha = 0.758$).

Procedures

After obtaining ethical clearance, three intact classes were invited to the study. The participants signed consent forms to indicate their willingness to voluntarily participate in the experiment. Two experienced teachers in the English public speaking course who are also judges in English speech contests were recruited as raters. In the rater training session, both raters and one of the researchers discussed the descriptors and standards of the rating rubrics and reviewed several speech samples to align the performances with the rubrics. Additionally, the raters marked 10 speeches and their ratings were compared and discussed to examine their understanding of the rubrics. When there were discrepancies in ratings, one of the researchers would offer a third-party opinion in the discussion until agreement was reached.

With regard to training for the participants, they were introduced to fundamental public speaking skills, including the generation of thesis statements, organizing the speech, supporting the speech with evidence, and delivering the speech.

Next, one of the researchers reviewed seven speech topics from a national English-speaking contest, which were used in speech practices in the English public speaking course for former batches of students. The raters and researchers then discussed these topics and chose two of them which were of similar difficulty level and format as topics for the pretest and posttest. Before the treatment, all participants took the pretest in which they each delivered a 3-minute prepared speech on Topic A (Dream vs. Reality) and answered the PRCS.

The treatment period lasted for 10 weeks in the 90-minute public speaking class. All three groups shared the same instructor, who followed the feedback procedure based on Thomson and Rucker's (2002) PSCI. The intact classes were assigned to one of the three conditions: control group (n = 25), in-class VF group (n = 24), and one-on-one VF group (n = 25). The instructor assigned a controversial topic to all students two weeks in advance to develop into 3-minute persuasive speech, invited five speakers to deliver their speeches in class, and provided feedback to them orally. Different students were invited to give the speech each time to ensure that all students would receive the same amount of exposure and feedback.

In the feedback sessions, the instructor first invited five students to deliver their speeches and then offered comments on the speeches. The control group received verbal feedback where all students (speaker students and audience students) listened

to the instructor's verbal feedback without viewing the speakers' videos. It took between 5 and 8 minutes to comment on each speech.

For the in-class VF group, all students watched the speakers' videos, while listening to the instructor's feedback. The instructor played each speaker's video and stopped it whenever she wanted to provide feedback and emphasize relevant speech skills. For instance, at the end of the introduction, the instructor would pause the video and ask the students if they could identify the thesis statement and realize that the speaker had moved on to the body of the presentation. When students found the thesis explicit and the structure clear, the instructor would comment on what probably made the thesis and structure of the speech clear. On the other hand, when students found the thesis and structure ambiguous or imprecise, the instructor would offer suggestions on how to improve the thesis. Strategies that were discussed included the rephrasing of the thesis, adding signposts, using longer pauses between the introduction and body, or slowing down when articulating the thesis. It would take approximately 15 minutes to give feedback on each speech.

For the one-on-one VF group, all students received the teacher's verbal comments without watching the videos, which would take about 5 to 8 minutes for each one; the speakers then spent 7 to 10 minutes watching their speech videos in a one-on-one meeting with the instructor. Again, the instructor would periodically pause the video to emphasize specific speaking skills and offer suggestions to align the verbal comments to their performance in the videos. For example, the instructor would ask the speakers to identify problematic body language, such as excessive gestures and frowning, and offer suggestions, like practicing in front of a mirror.

After the treatment, the students took the posttest, delivering a 3-minute prepared speech on Topic B (Life Initiative vs. Convenience) and then completed the PRCS. Data were collected and pre-processed for statistical analysis (see Table 1 for a summary of the procedures).

Table 1. Experiment procedures

Procedure	Notes
Preparation	● Raters receiving training *Discussion of the rating scale and sample speech* *Pilot ratings and discussion to align application of rating scale* ● Students learning fundamental public speaking skills

Pretest	Instruments ● Topic A (Dream vs. Reality) ● PSCI for speaking competence ● PRCS for speaking anxiety
Treatment	● Control group (*n* = 25) *Verbal feedback without video for all students (5-8m/speech)* ● In-class VF group (*n* = 24) *Verbal feedback with video for all students (10-15m/speech)* ● One-on-one VF group (*n* = 25) *Verbal feedback without video for all students (5-8m/speech) & Verbal feedback with videos for speaker students (7-10m/speech)*
Posttest	Instruments ● Topic B (Life Initiative vs. Convenience) ● PSCI for speaking competence ● PRCS for speaking anxiety

Notes: m = minutes; PSCI = Public Speaking Competence Instrument; PRCS = Personal Report of Confidence as a Speaker; VF = video feedback

Data analysis

Data validation using Rasch measurement

The pretest and posttest data each include two raters' ratings on 20 items from five criteria of the PSCI for the 74 participants. As the public speaking competence measurement is a rater-mediated assessment, we submitted the ratings for each of the five criteria for the pretest and posttest separately to many-facet Rasch measurement (MFRM) (Linacre, 1994). Specifically, we investigated the impact of the following facets on students' scores: students' speaking ability, rater severity, and item difficulty. Facets computer program, Version 3.83.4 (Linacre, 2021) was used for these analyses. The mathematical expression of this three-facet MFRM applied is expressed as follows:

$$\log\frac{Pnijk}{Pnijk-1} = B_n - D_i - C_j - F_k$$

, where P_{nijk}, is the probability that student speaker n of ability B_n is observed by rater j of severity C_j in category k of item i of difficulty D_i as opposed to the probability P_{nijk} – 1 of the speaker being observed in k – 1 category (Linacre, 1994, 2012a). Accordingly, the probability of a speaker being awarded a score by the rater on each item is a function of his or her ability and item difficulty, scoring rubrics, and rater severity/lenience (Linacre, 1994). MFRM adjusts the ratings for the effect of each of these facets and generates "fair scores", which partial out the effect of other facets on

the scores so that the effect of "bias" on the test scores is eliminated (Linacre, 1994). The fair scores were used in the inferential statistical analysis to answer the research questions of the study.

MFRM analysis further estimated infit and outfit mean square (MnSq) indices—the compatibility of the observed score with the score expected by the model (Bond & Fox, 2015). Infit MnSq detects unexpected ratings near the person's ability, such as when an average student receives an unexpectedly high or low score by one or more raters (Aryadoust, 2015; Linacre, 2012b). Outfit MnSq detects anomalies in scores that are far from the person's ability, such as when a low-ability student is given an unexpectedly high score (Aryadoust, 2015; Linacre, 2012b). MnSq values between 0.5 and 1.5 indicate tolerable variation from the measures expected by the model, with values below 0.5 indicating over-predictability of measures (overfit) and values greater than 1.5 indicating possible serious distortions of the measure (underfit) (Bond & Fox, 2015; Linacre, 2012b). While a MnSq greater than 1.5 indicates inaccuracy in the measures, a MnSq smaller than 0.5 suggests no severe problems (Linacre, 2012b). Therefore, in this study, we excluded the underfit data (infit MnSq or outfit MnSq > 1.5) and retained only the fair measures of students' ability of the fitting data for the inferential statistics reported below.

Next, we estimated the reliability and separation indices of the speaker. Reliability in MFRM varies between 0 and 1 and the speaker reliability index indicates that the speaker ordering can be reproduced if this sample of speakers were given another parallel set of items measuring the same construct (Bond & Fox, 2015). Separation is another expression of reliability and indicates the number of statistically distinguishable strata in the data (Aryadoust, 2015; Bond & Fox, 2015).

Inferential statistics

To address the two research questions, SPSS, version 25 (IBM Corp., 2017) was used for data analysis. To answer research question 1, we applied ANCOVA on the posttest scores, using the pretest scores as the covariate. However, the analysis failed to meet the assumptions of ANCOVA, which are no significant difference between comparison groups and no interaction between covariate and independent variable (Field, 2013). Therefore, we decided to perform an ANOVA on gain scores (posttest score minus pretest score) to determine whether there was a significant difference in the improvement in participants' public speaking competence (5 subscales + total) between groups. The fair measures for students' ability obtained in the MFRM analysis were used in the gain score ANOVA with all misfit data removed from the study. It should be noted that ANCOVA and gain score analysis are two major

solutions for partialling out pre-existing difference in pretest-posttest design (Knapp & Schafer, 2009; Roever & Phakiti, 2017). Gain score analysis does not rely on conditioning and is a useful alternative to ANCOVA (Kim & Steiner, 2021; Knapp & Schafer, 2009), especially when the purpose of research is to observe the amount of the gain (Wright, 2006).

To answer research question 2, we further conducted ANCOVA, using the pretest anxiety score as the covariate to examine whether there was a significant difference between the three groups in their speaking anxiety. The assumptions of ANCOVA were met and the Bonferroni post hoc test was used to adjust for multiple comparisons. ANCOVA was run to partial out pre-existing differences in participants' anxiety level in the pre-test (Knapp & Schafer, 2009).

Results

Data validation using MFRM

12 separate MFRM analyses [(5 subscores + 1 total score) * 2 = 12] were performed on the subscores and total scores of the three groups' speaking competence.

Table 2 presents statistics for the speaker, including observed scores, fair measures, fit statistics, standard error of measurement, and reliability statistics of the five subscales and the total of both pretest and posttest generated in the analyses. The range of observed scores are the average scores awarded by the two raters on each item of the five subscales. The fair measure was obtained by adjusting the disparity between severe and lenient raters by computing the expected rating that would be awarded by a rater with a level of severity of zero logits (Eckes, 2005). There was no significant difference between the observed score and fair measure, indicating the leniency and severity of raters cancelled each other out. The fair measures were further expressed in logits, with a medium standard error of measurement except for global competence, suggesting fair precision in the ability estimates for the students. The medium to high reliability coefficients generated for the total and five subscales, except for the subscale of global competence, indicated that the marks would be reproducible if the test was readministered to the students. The validity of the data was further evidenced by the high separation in the students' ability measure, which indicated proper variability in the students' ability. Subscale 5 (global competence) generated a high standardized error of measurement, a low reliability coefficient, and low separation, probably because it included only a single item to rate a global impression of the students' performance. Finally, there were around 10 out of 74 misfit data for each subscale and

the total, indicating some erratic patterns in the marks. These non-fitting measures of the students' ability can be attributed to construct-irrelevant factors and thus were eliminated from the inferential statistical analysis.

Table 2. Student's ability, fit, and reliability estimated using MFRM

Test	Subscale	Observed Score Range	Fair Measure Range	Measure (*Logits*) Range	SE Range	Infit MnSq Range	Outfit MnSq Range	Separation	Reliability
Pretest	Introduction	2.00-3.67	2.00-3.64	-3.68-2.81	0.66-1.03	0.15-3.45	0.15-3.5	1.14	0.57
	Body	2.00-3.60	2.02-3.58	-4.03-2.70	0.59-0.76	0.09-3.09	0.09-3.71	1.65	0.73
	Conclusion	1.00-3.88	1.02-3.87	-2.58-3.28	0.61-1.90	0.27-2.41	0.25-2.25	2.51	0.86
	Delivery	1.64-3.57	1.63-3.57	-4.01-1.71	0.45-0.48	0.38-2.40	0.37-2.39	2.44	0.86
	Global Competence	1.50-3.50	1.50-3.50	-12.04-11.99	1.49-8.69	0.00-1.95	0.00-1.95	0.00	0.00
	Total	38.00-71.50	38.30-70.59	-1.91-2.22	0.23-0.63	0.00-7.52	0.00-7.53	2.69	0.88
Posttest	Introduction	2.33-4.17	2.32-4.15	-6.82-3.74	0.84-1.58	0.03-4.98	0.03-4.51	1.90	0.78
	Body	2.40-3.60	2.37-3.63	-2.91-2.91	0.69-0.71	0.30-2.31	0.26-2.61	1.73	0.75
	Conclusion	2.38-3.75	2.33-3.75	-5.36-1.38	0.70-0.84	0.21-2.49	0.18-4.22	1.24	0.61
	Delivery	2.43-4.36	2.43-4.36	-3.17-7.20	0.56-0.68	0.28-2.10	0.24-2.38	3.30	0.92
	Global Competence	2.00-4.00	2.05-3.95	-9.66-9.66	2.13-7.65	0.01-7.67	0.01-7.82	0.00	0.00
	Total	48.50-80.50	48.19-81.85	-3.88-3.97	0.29-0.62	0.00-9.00	0.00-9.00	3.48	0.92

Research question 1

Descriptive statistics

The gain scores (posttest scores minus pretest scores) of the five subscales include introduction (M = 0.221, SD = 0.288), body (M = 0.189, SD = 0.229), conclusion (M = 0.592, SD = 0.489), delivery (M = 0.470, SD = 0.298), and global competence (M = 0.504, SD = 0.506) as well as the total score of the fair measure of students' speaking competency (M = 7.697, SD = 4.631). The gain scores were normally distributed as indicated by their skewness coefficients (0.006–1.259), and kurtosis coefficients (0.911–1.379).

The General linear model (GLM) ANOVA

We used students' fair measures of speaking competence (represented by five subscales and total) generated in the MFRM (misfit data excluded) to compare the gain scores in each subscale and the total using GLM ANOVA. The Levene's test showed that the data were homogenous in scores for all subscales and total ($p > 0.05$). The results indicated that there was a significant difference in delivery [$F(2,53) = 6.193$, $p = 0.004$, partial $\eta^2 = 0.189$] and global competence [$F(2,58) = 8.891$, $p < 0.001$, partial $\eta^2 = 0.235$] and a small to medium effect size, with the feedback methods accounting for 18.9% of the variance in the students' speaking delivery and 23.5% of the variance in the global speaking ability (see Table 3). The Bonferroni post hoc test showed a significant difference between the control group and the in-class VF group in both speech delivery ($p = 0.003$) and global competence ($p < 0.001$), which indicated that the in-class VF significantly improved students' delivery and general speaking competence (see Table 4).

Table 3. ANOVA GLM of the gain scores between three groups

Subscale	Source	SS	df	MS	*F*	*Sig.*	ηp^2	Levene's Test
Introduction	Group	0.081	2	0.040	0.479	0.622	0.018	2.866 ($p = 0.066$)
	Error	4.475	53	0.084				
Body	Group	0.013	2	0.006	0.117	0.890	0.004	0.578 ($p = 0.564$)
	Error	2.821	52	0.054				
Conclusion	Group	1.032	2	0.516	2.272	0.115	0.090	0.633 ($p = 0.536$)
	Error	10.447	46	0.227				
Delivery	Group	0.925	2	0.462	6.193	**0.004**	0.189	1.680 ($p = 0.196$)
	Error	3.956	53	0.075				
Global	Group	3.609	2	1.805	8.891	**0.000**	0.235	0.026 ($p = 0.974$)
Competence	Error	11.772	58	0.203				
Total	Group	107.391	2	53.695	2.661	0.080	0.094	1.509 ($p = 0.231$)

	Error	1029.150	51	20.179

Notes: SS = sum of squares; *df* = degree of freedom; MS = mean square; Sig. = significance; ηp^2 = partial eta squared; Levene's statistics is based on mean

Table 4. The Bonferroni post hoc tests for multiple comparisons of speaking competence

Sub-Scale	Group (I)	Group (J)	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval	
						Lower Bound	Upper Bound
Delivery	Control	In-class	-0.308*	0.088	**0.003**	-0.524	-0.091
	In-class	One-on-one	0.163	0.090	0.229	-0.060	0.386
	One-on-one	Control	0.145	0.091	0.355	-0.081	0.370
Global Competence	Control	In-class	-0.601*	0.143	**0.000**	-0.952	-0.249
	In-class	One-on-one	0.257	0.139	0.210	-0.086	0.599
	One-on-one	Control	0.344	0.143	0.057	-0.008	0.696

Notes: Based on estimated marginal means

*The mean difference is significant at the 0.05 level.

[b] Adjustment for multiple comparisons: Bonferroni

Research question 2

Checking the ANCOVA conditions

Both the pretest score (M = 16.65, SD = 4.483) and posttest score (M = 14.38, SD = 5.012) were normally distributed, as indicated by their skewness coefficients (0.276 for pretest and 0.097 for posttest), and kurtosis coefficients (-0.630 for pretest and -0.472 for posttest). An ANOVA test of the students' pretest scores for anxiety showed that there was no significant difference between the three groups ($F(2,71) = 0.599$, $p > 0.05$). Additionally, the presumption of homogeneity of regression slopes was met as there was no interaction between the covariate (pretest anxiety) and the independent variable (grouping) ($F(2,68) = 0.392$, $p > 0.05$).

ANCOVA of the public speaking anxiety

ANCOVA showed that there was a significant difference in students' public speaking anxiety among the three groups, with a small effect size ($F(2,70) = 7.329$, $p = 0.001$, partial $\eta^2 = 0.173$). The Bonferroni post hoc test showed the significant difference was between the control group and the in-class VF group, indicating that in-class VF significantly reduced students' speaking anxiety (see Table 5).

Table 5. Bonferroni post hoc test for multiple comparison of speaking anxiety

Group (I)	Group (J)	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval	
					Lower Bound	Upper Bound
Control	In-class	4.586*	1.211	**0.001**	1.616	7.555
In-class	One-on-one	-1.635	1.193	0.524	-4.560	1.290
One-on-one	Control	-2.951	1.214	0.053	-5.927	0.026

Notes: Based on estimated marginal means

*The mean difference is significant at the 0.05 level.

[b] Adjustment for multiple comparisons: Bonferroni

Discussion

This study examined the effect of in-class VF and one-on-one VF on EFL learners' English public speaking competency and anxiety. In what follows, the findings of the study are discussed.

Research question 1

Significant differences were found between the control group and the in-class VF group in subscale 4 (delivery) and subscale 5 (global competence), whereas no significant improvement was found between groups in the subscales of introduction, body and conclusion.

The first point worth noticing is that significant difference was only captured in the subscales of delivery and global competence. This result is in line with Bankston and Terlip (1994), which found that among the 16 criteria of the assessment scale, students' attention to three criteria of delivery changed significantly across time (i.e., eye contact, pitch, and loudness), whereas the content or language-related criteria such as clear focus and constructing precise sentences remained unchanged. One possible reason is that in-class VF offered the students the opportunity to see in the camera aspects of their delivery, such as their use of gestures and eye contact, and adjust the discrepancies between their projected and actual behaviors. Just as Wilhelm (2014) stated, VF allowed students to view and review their videos and rediscover the projection of self-image, both verbally and physically. Similarly, Woolley's earlier study (1960) found that students tended to be more attentive to their delivery skills rather than the content of their speech when watching their videos.

In addition, this result indicated the importance of delivery to general public speaking competence, as can be concluded from the inclusion of delivery skills in most public

speaking rubrics (e.g., Schreiber et al., 2012; Thomson & Rucker, 2002). Mehrabian (1971) found that the impact of facial expression, vocal expression, and verbal expression accounted for 55%, 38%, and 7% of information links respectively. Therefore, it can be argued that delivery skill is crucial to general speaking competence and VF can be an expedient and effective way of enhancing the overall speaking competence.

By contrast, students did not make any significant progress in developing the introduction, body, and conclusion of their public speech. This partially resonates with a previous study by Aryadoust (2015), wherein the items measuring the opening and closing remarks were found to have high difficulty. Opening remarks can be more taxing since this is the time when the speakers have the highest amount of anxiety, which can affect the quality of the presentation (Behnke & Sawyer, 2004). On the other hand, closing remarks need well-developed integration skills to synthesize the main points of the presentation, and these are among the most difficult competencies to master. They take more time to develop effectively and, therefore, teachers should be patient with the slow progress of their students.

Another point that interested us is that the significant difference was found only between the in-class VF group and the control group. Social cognitive theory is well suited as an explanation for this result (Bandura, 2005). According to the theory, the social cognitive processes (in learning) start with the observation of a social model and continue with repeated performance, following which learners would progress to a final self-regulated stage (Wouters et al., 2007). The theory also postulates that developing complex behaviors requires both self-regulative operations and corrective feedback systems to translate the procedural knowledge into proficient performances (Bandura, 2005). It may be posited that compared with the control group, the in-class VF group had better model observation and feedback uptake. While the class reviewed their own or peers' performance in the video, the instructor would pause in due time to discuss and underscore the speaking techniques applied by the speaker, followed by suggestions for future improvement. Students thus associated, in a more concrete fashion, the speaker's strengths and weaknesses with the specific behaviors in the video.

Additionally, observing other students' speech fosters learning via "vicarious reinforcement" and "vicarious punishment", so that people can benefit from the successes and weaknesses of others as well as from their own experiences (Bandura, 1977). By observing their peers' performance, students would enact desired behaviors and refrain from undesired ones to achieve successful performance. Notably, good speeches were responded to with applause, better audience engagement, and

favorable comments from the instructor (De Grez et al., 2009), although the unsuccessful speeches were not harshly criticized, to avoid the reinforcement of negative affect.

Compared with one-on-one VF, in-class VF offered both the speakers and non-speakers opportunities to critically appraise their own or their peers' performance in the video against the criteria and obtain professional feedback to verify their assessment. Contrastingly, non-speaker students in the one-on-one VF group were deprived of such vicarious learning opportunities, although the speaker student received the one-on-one VF. This suggests that the amount of exposure to VF, even though it is provided on peers' performance, has significant effects on students' public speaking competence. Thus, we suggest that the teacher should make good use of the video and assist students to notice not only their own strengths and weaknesses but also get inspired from their peers' performance. Nevertheless, we suggest that before foregrounding vicarious learning via VF, it would be important that the teacher highlights the point that reviewing video recorded speeches in class serves only educational ends and that students need not feel embarrassed about the critical comment on their presentation.

Research question 2

We found that the in-class VF group experienced significantly lower PSA compared with the control group. This can be attributed to the therapeutic effect of VF, which can provide a more objective source of information concerning one's social performance (Rapee & Hayman, 1996). A major source of speaker anxiety is distorted self-image, which leads to underestimations of their performance (Aryadoust, 2015), overestimations of the visibility of anxiety symptoms, and large discrepancies between self and objective-observer perceptions of performance (Aryadoust, 2015; Rodebaugh et al., 2010; Rodebaugh & Rapee, 2005). Compared with the control group, the in-class VF group had more exposure to their own and their peers' images in the videos, which helped them to obtain a more objective estimation of their ability, taking the observer's perspective. In contrast, the socially anxious individuals in the control group did not make any progress in the alleviation of their anxiety, likely because they would base the appraisal of their performance on a mental representation of how they appeared to an observer (Rapee & Hayman, 1996).

Again, interestingly, no significant difference was found in PSA between the one-on-one VF group and control group or between the in-class VF group and one-on-one VF group. This indicated that the one-on-one VF was effective in reducing PSA but not so effective as in-class VF. This finding provides an answer to the controversies

apropos of one-on-one VF on performance and anxiety. On the one hand, the one-on-one VF group were exposed to the videos of their speech performance, which helped to correct their biased negative self-image. On the other hand, compared to the in-class VF group, the one-on-one VF group had less exposure to their fear of being negatively evaluated in class, and thus were less likely to achieve improvements.

Additionally, compared to the one-on-one VF group, the in-class VF group could have conducted their self-initiated cognitive preparation, although the instructor did not intentionally undertake it. Harvey et al. (2000) suggested that cognitive preparation conducted before self-observation in VF can result in a more objective viewing of one's performance and minimize the discrepancies between dysfunctional self-image and video evidence. Cognitive preparation directs participants to make a detailed prediction and visualization of their performance in the video, and then shift their attention from their subjective feelings to their actual performance (Harvey et al., 2000). Cognitive preparation has been found to enhance the effects of VF as it leads to greater improvement in participants' self-evaluations of performance in comparison to VF alone or an exposure control (Rodebaugh, 2004; Rodebaugh et al., 2010).

In our study, while listening to the instructor's comment on their peers' performance, the in-class VF group could have conducted their self-initiated cognitive preparation by predicting their performance in the video and projecting self-image against the strengths and weaknesses of their peer's performance. This self-initiated cognitive preparation could have helped to reduce the anxiety level of the in-class VF group by transferring their attention from subjective feelings to criteria-related aspects of their performance when receiving VF. However, the cognitive preparation in an implicit self-initiated fashion rather than in an explicit instructor-directed means may account for the absence of significant difference in the PSA between the in-class and one-on-one VF group. This warrants future study on the effect of instructor-directed cognitive preparation before the VF on students' PSA.

Conclusion

In sum, the study showed that in-class VF was effective in improving speakers' delivery skills and global competence and in reducing their PSA, although there was no difference between the in-class and one-on-one VF conditions. Thus, we propose that VF be conducted as an instructional procedure in the classroom to expose students to in vivo public speaking experiences and facilitate the acquisition of public speaking skills. While we believe in the value of one-on-one feedback, in-class VF is perhaps more feasible for universities with larger class size, limited instruction hours,

or limited human and lab resources. Due to the tendency to reduce the instruction hours in university curricula (De Grez et al., 2009), it is impractical to have all students deliver their speeches and offer them VF on a one-on-one basis in one class session. Additionally, in contexts where no teaching assistant or tutor is hired, in-class VF would allow students to reap the benefits of feedback on a few selected students' speeches, which represent students' strengths and weaknesses. Although we did not find significant differences between the effect of one-on-one VF and in-class VF, we believe that in-class VF saves previous time and would be more focused and specific as it has the necessary attributes of effective feedback summarized by Miskam and Saidalvi (2020).

In-class VF is also facilitative in representing the non-verbal information which is hardly representable through verbal comments, thus reinforcing the application of a wide array of verbal and non-verbal skills in public speeches. It should be noted that when applying in-class VF, the instructors must be cautious about the content and language of their comments, as speech-making is a daunting experience for students (Bodie, 2010). Positive feedback should be provided generously as it has been shown to increase students' task engagement and confidence and decrease problem behaviors (Pinter et al., 2015). In pointing out students' weaknesses, it would also help for the teacher to phrase his/her comments in a pattern of expressing anticipation for desired behaviors in the students' future speech. Having said that, instructors are still highly recommended to conduct in-class VF as students generally demonstrate positive attitudes towards it (Bourhis & Allen, 1998; Miskam & Saidalvi, 2020).

Despite our efforts to enhance the validity and reliability of the research, the limitations of this study need to be acknowledged. Due to practicality considerations, the study used three intact classes as a convenience sample. However, the fact that this research simulated real-class instruction can enhance the ecological validity and transferability of the results. Additionally, future research should engage more trained raters, where feasible, to add to the validity of the performance ratings.

In our discussion of the results, we proposed that the enhancement of the speaking skills and reduction of PSA of the in-class VF group could be attributed to the "vicarious experience" and "cognitive preparation" they gained from receiving VF on their peers' performance. In this process, students could have initiated self-assessment or peer assessment in their minds and then confirmed or adjusted their assessment when receiving the instructor's feedback. Therefore, future studies should explore how students engage with teacher and peer feedback and self-assessment to add to the impact of VF (Price et al., 2011) and examine how students' engagement relates to their speaking competence.

Acknowledgements

This study was supported by Sichuan International Studies University, Curriculum and Teaching Renovation, No. JY1869218 and Nanyang Technological University (NTU) Research Scholarship.

References

Alek, A., Marzuki, A. G., Farkhan, M., & Deni, R. (2020). Self-assessment in exploring EFL students' speaking skill. *Al-Ta' Lim Journal, 27*(2), 208–214. https://doi.org/10.15548/jt.v27i2.613

Aryadoust, V. (2015). Self-and peer assessments of oral presentations by first-year university students. *Educational Assessment, 20*(3), 199-225. https://doi.org/ /10.1080/10627197.2015.1061989

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (2005). The evolution of social cognitive theory. In K. Smith & M. Hitt (Eds.), *Great minds in management: The process of theory development* (pp. 9–35). Oxford University Press. http://www.uky.edu/~eushe2/BanduraPubs/Bandura2005.pdf

Bankston, R. G., & Terlip, L. A. (1994). *The effects of videotaping on student performances in the basic communication course* (ED378604). ERIC. https://files.eric.ed.gov/fulltext/ED378604.pdf

Barry, S. (2012). A video recording and viewing protocol for student group presentations: Assisting self-assessment through a Wiki environment. *Computers & Education, 59*(3), 855–860. https://doi.org/10.1016/j.compedu.2012.04.008

Behnke, R. R., & Sawyer, C. R. (2004). Public speaking anxiety as a function of sensitization and habituation processes. *Communication Education, 53*(2), 164–173. https://doi.org/10.1080/0363452041000168242 9

Bodie, G. D. (2010). A racing heart, rattling knees, and ruminative thoughts: Defining, explaining, and treating public speaking anxiety. *Communication Education, 59*(1), 70–105. https://doi.org/10.1080/03634520903443849

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Bourhis, J., & Allen, M. (1998). The role of videotaped feedback in the instruction of public speaking: A quantitative synthesis of published empirical research.

Communication Research Reports, 15(3), 256–261. https://doi.org/10.1080/08824099809362121

Chen, C.-M., & Lee, T.-H. (2011). Emotion recognition and communication for reducing second-language speaking anxiety in a web-based one-to-one synchronous learning environment: Emotion recognition for reducing second-language anxiety. *British Journal of Educational Technology, 42*(3), 417–440. https://doi.org/10.1111/j.1467-8535.2009.01035.x

Chen, J., McLean, J. E., & Kemps, E. (2018). Combined audience and video feedback with cognitive review improves state anxiety and self-perceptions during speech tasks in socially anxious individuals. *Behavior Therapy, 49*(2), 165–179. https://doi.org/10.1016/j.beth.2017.11.006

Chen, L. H., & Wang, D. M. (2015). Effect of systematic desensitization and cognitive restructuring on English speaking anxiety. *Proceedings of the 2015 International Conference on Social Science, Education Management and Sports Education,* 83-86. https://doi.org/10.2991/ssemse-15.2015.21

Christianson, M., Hoskins, C., & Watanabe, A. (2009). Evaluating the effectiveness of a video-recording based self-assessment system for academic speaking. *Language Research Bulletin, 24*, 1–15.

Daly, J. A. (1978). The assessment of social-communicative anxiety via self-reports: A comparison of measures. *Communication Monographs, 45*(3), 204 – 218. https://doi.org/10.1080/03637757809375966

De Grez, L., Valcke, M., & Roozen, I. (2009). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers & Education, 53*(1), 112–120. https://doi.org/10.1016/j.compedu.2009.01.005

De Grez, L., Valcke, M., & Roozen, I. (2012). How effective are self- and peer assessment of oral presentation skills compared with teachers' assessments? *Active Learning in Higher Education, 13*(2), 129–142. https://doi.org/10.1177/1469787412441284

Deihl, E. R., Breen, M. P., & Larson, C. U. (1970). The effects of teacher comment and television video tape playback on the frequency of nonfluency in beginning speech students. *The Speech Teacher, 19*(3), 185–189. https://doi.org/10.1080/03634527009377819

Dupagne, M., Stacks, D. W., & Giroux, V. M. (2007). Effects of video streaming technology on public speaking students' communication apprehension and competence. *Journal of Educational Technology Systems, 35*(4), 479–490. https://doi.org/10.2190/5947-4W72-303L-Q578

Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. *Language Assessment Quarterly, 2*(3), 197–221. https://doi.org/10.1207/s15434311laq0203_2

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Glenn, R. J. (1996). *Using video to enhance content and delivery skills in the basic oral communication course: Summarizing the uses and benefits* (ED399600). ERIC. https://eric.ed.gov/?id=ED399600

Hallmark, J. R., Hanson, T. L., Padwick, G., Abel, D., & Stewart, P. (1993). *Communication apprehension remediation: The interaction effect of video self-observation and gender* (ED363902). ERIC. https://files.eric.ed.gov/fulltext/ED363902.pdf

Hancock, A. B., Stone, M. D., Brundage, S. B., & Zeigler, M. T. (2010). Public speaking attitudes: Does curriculum make a difference? *Journal of Voice, 24*(3), 302–307. https://doi.org/10.1016/j.jvoice.2008.09.007

Harvey, A. G., Clark, D. M., Ehlers, A., & Rapee, R. M. (2000). Social anxiety and self-impression: Cognitive preparation enhances the beneficial effects of video feedback following a stressful social task. *Behaviour Research and Therapy, 38*(12), 1183–1192. https://doi.org/10.1016/S0005-7967(99)00148-5

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Hinton, J. S., & Kramer, M. W. (1998). The impact of self-directed videotape feedback on students ' self - reported levels of communication competence and apprehension. *Communication Education, 47*(2), 151 – 161. https://doi.org/10.1080/03634529809379119

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Hung, S.-T. (2011). Pedagogical applications of vlogs: An investigation into ESP learners' perceptions: Pedagogical applications of vlogs. *British Journal of Educational Technology, 42*(5), 736–746. https://doi.org/10.1111/j.1467-8535.2010.01086.x

IBM Corp. (2017). *IBM SPSS statistics for Windows* (25.0) [Computer software]. IBM Corp.

Junko, O., & Heffernan, N. (2008). Factors predicting oral presentations in EFL classrooms. *The Asian EFL Journal, 10*(1), 65–78.

Kim, Y., & Steiner, P. M. (2021). Gain scores revisited: A graphical models perspective. *Sociological Methods & Research, 50*(3), 1353–1375. https://doi.org/10.1177/0049124119826155

Klorman, R., Weerts, T. C., Hastings, J. E., Melamed, B. G., & Lang, P. J. (1974). Psychometric description of some specific-fear questionnaires. *Behavior Therapy, 5*(3), 401–409. https://doi.org/10.1016/S0005-7894(74)80008-0

Knapp, T. R., & Schafer, W. D. (2009). From gain score t to ANCOVA f (and vice versa). *Practical Assessment, Research, and Evaluation, 14,* Article 6. https://doi.org/10.7275/YKE1-K937

Knoch, U., Fairbairn, J., Myford, C., & Huisman, A. (2018). Evaluating the relative effectiveness of online and face-to-face training for new writing raters. *Papers in Language Testing and Assessment, 7*(1), 61-86.

Koizumi, R., Kaneko, E., Setoguchi, E., Innami, Y., & Naganuma, N. (2019). Examination of CEFR-J spoken interaction tasks using many-facet Rasch measurement and generalizability theory. *Papers in Language Testing and Assessment, 8*(2), 1-33.

LeFebvre, L., LeFebvre, L., Blackburn, K., & Boyd, R. (2015). Student estimates of public speaking competency: The meaning extraction helper and video self-evaluation. *Communication Education, 64*(3), 261–279. https://doi.org/10.1080/03634523.2015.1014384

Linacre, J. M. (1994). *Many-facet Rasch measurement* (2nd ed). MESA Press.

Linacre, J. M. (2012a). *Many-facet Rasch measurement: Facets tutorial 1.* https://www.winsteps.com/a/ftutorial1.pdf

Linacre, J. M. (2012b). *Many-facet Rasch measurement: Facets tutorial 2.* https://www.winsteps.com/a/ftutorial2.pdf

Linacre, J. M. (2021). *Facets computer program for many-facet Rasch measurement* (3.83.4) [Computer software]. Winsteps.com.

Ling, Y., Brinkman, W.-P., Nefs, H. T., Qu, C., & Heynderickx, I. (2012). Effects of stereoscopic viewing on presence, anxiety, and cybersickness in a virtual reality environment for public speaking. *Presence: Teleoperators and Virtual Environments, 21*(3), 254–267. https://doi.org/10.1162/PRES_a_00111

Lucas, S. (2001). The art of public speaking (7th ed.). McGraw Hill.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning, 44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

McCroskey, J. C. (1977). Oral communication apprehension: A summary of recent theory and research. *Human Communication Research, 4*(1), 78–96. https://doi.org/10.1111/j.1468-2958.1977.tb00599.x

McCroskey, J. C., & Lashbrook, W. B. (1970). The effect of various methods of employing video-taped television playback in a course in public speaking. *The Speech Teacher, 19*(3), 199–205. https://doi.org/10.1080/03634527009377821

McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports, 5*(2), 108–113. https://doi.org/10.1080/08824098809359810

Mehrabian, A. (1971). *Silent messages.* Wadsworth Publishing Company.

Miles, P. L. (1981). Student video self-critiques. *Communication Education, 30*(3), 280–283. https://doi.org/10.1080/03634528109378481

Mills, G. E., & Pace, R. W. (1989). What effects do practice and video feedback have on the development of interpersonal communication skills? *The Journal of Business Communication (1973), 26*(2), 159–176. https://doi.org/10.1177/002194368902600205

Miskam, N. N., & Saidalvi, A. (2020). Using video technology to improve oral presentation skills among undergraduate students: A systematic literature review. *International Journal of Psychosocial Rehabilitation, 24*(5), 5280–5291. https://doi.org/10.37200/ijpr/v24i5/pr2020235

Morreale, S., Backlund, P., Hay, E., & Moore, M. (2011). Assessment of oral communication: A major review of the historical development and trends in the movement from 1975 to 2009. *Communication Education, 60*(2), 255–278. https://doi.org/10.1080/03634523.2010.516395

Newburger, C., Brannon, L., & Daniel, A. (1994). Self-confrontation and public speaking apprehension: To videotape or not to videotape student speakers? *Basic Communication Course Annual, 6,* Article 18. https://ecommons.udayton.edu/cgi/viewcontent.cgi?article=1246&context=bcca

Ng, E. M. W. (2014). Using a mixed research method to evaluate the effectiveness of formative assessment in supporting student teachers' wiki authoring. *Computers & Education, 73,* 141–148. https://doi.org/10.1016/j.compedu.2013.12.016

Nikolic, S., Stirling, D., & Ros, M. (2018). Formative assessment to develop oral communication competency using YouTube: Self- and peer assessment in engineering. *European Journal of Engineering Education, 43*(4), 538–551. https://doi.org/10.1080/03043797.2017.1298569

Paul, G. L. (1966). *Insight vs desensitization in psychotherapy.* Stanford University Press.

Pinter, E. B., East, A., & Thrush, N. (2015). Effects of a video-feedback intervention on teachers' use of praise. *Education and Treatment of Children, 38*(4), 451–472. https://doi.org/10.1353/etc.2015.0028

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education, 36*(8), 879–896. https://doi.org/10.1080/03075079.2010.483513

Quigley, B. L., & Nyquist, J. D. (1992). Using video technology to provide feedback to students in performance courses. *Communication Education, 41*(3), 324–334. https://doi.org/10.1080/03634529209378892

Rapee, R. M., & Hayman, K. (1996). The effects of video feedback on the self-evaluation of performance in socially anxious subjects. *Behaviour Research and Therapy, 34*(4), 315–322. https://doi.org/10.1016/0005-7967(96)00003-4

Rodebaugh, T. L. (2004). I might look OK, but I'm still doubtful, anxious, and avoidant: The mixed effects of enhanced video feedback on social anxiety symptoms. *Behaviour Research and Therapy, 42*(12), 1435–1451. https://doi.org/10.1016/j.brat.2003.10.004

Rodebaugh, T. L., Heimberg, R. G., Schultz, L. T., & Blackmore, M. (2010). The moderated effects of video feedback for social anxiety disorder. *Journal of Anxiety Disorders, 24*(7), 663–671. https://doi.org/10.1016/j.janxdis.2010.04.007

Rodebaugh, T. L., & Rapee, R. M. (2005). Those who think they look worst respond best: Self-observer discrepancy predicts response to video feedback following a speech task. *Cognitive Therapy and Research, 29*(6), 705–715. https://doi.org/10.1007/s10608-005-9634-9

Roever, C., & Phakiti, A. (2017). *Quantitative methods for second language research: A problem-solving approach.* Routledge. https://doi.org/10.4324/9780203067659

Schmidt, M., Freund, Y., Alves, M., Monsel, A., Labbe, V., Darnal, E., Messika, J., Bokobza, J., Similowski, T., & Duguet, A. (2014). Video-based feedback of oral clinical presentations reduces the anxiety of ICU medical students: A multicentre, prospective, randomized study. *BMC Medical Education, 14*(1), 103. https://doi.org/10.1186/1472-6920-14-103

Schreiber, L. M., Paul, G. D., & Shibley, L. R. (2012). The development and test of the public speaking competence rubric. *Communication Education, 61*(3), 205–233. https://doi.org/10.1080/03634523.2012.670709

Shih, R.-C. (2010). Blended learning using video-based blogs: Public speaking for English as a second language students. *Australasian Journal of Educational Technology, 26*(6), 883–897. https://doi.org/10.14742/ajet.1048

Simpson, T., Holden, K., Merrick, D., Dawson, S., & Bedford, L. (2019). Does video feedback & peer observation offer a valid method of reinforcing oral presentation training for undergraduate biochemists? *Higher Education Pedagogies, 4*(1), 262–283. https://doi.org/10.1080/23752696.2019.1587717

Thomson, S., & Rucker, M. L. (2002). The development of a specialized public speaking competency scale: Test of reliability. *Communication Research Reports, 19*(1), 18–28. https://doi.org/10.1080/08824090209384828

van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2017). The impact of the feedback source on developing oral presentation competence. *Studies in Higher Education, 42*(9), 1671–1685. https://doi.org/10.1080/03075079.2015.1117064

Wilhelm, T. (2014). Humanising language teaching magazine for teachers and teacher trainers. *Humanising Language Teaching, 16*(1). http://old.hltmag.co.uk/feb14/sart05.htm

Woolley, T. R. (1960). Potential utility of VTR as an auxiliary feedback device. *The Speech Teacher, 9*(2), 135–139. https://doi.org/10.1080/03634526009377075

Wouters, P., Tabbers, H. K., & Paas, F. (2007). Interactivity in video-based models. *Educational Psychology Review, 19*(3), 327–342. https://doi.org/10.1007/s10648-007-9045-4

Wright, D. B. (2006). Comparing groups in a before–after design: When t test and ANCOVA produce different results. *British Journal of Educational Psychology, 76*(3), 663–675. https://doi.org/10.1348/000709905X52210

Xu, Q., Dong, X., & Jiang, L. (2017). EFL learners' perceptions of mobile-assisted feedback on oral production. *TESOL Quarterly, 51*(2), 408–417. https://doi.org/10.1002/tesq.335

Youn, S. J. (2018). Rater variability across examinees and rating criteria in paired speaking assessment. *Papers in Language Testing and Assessment, 7*(1), 32–60.