EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

Developing video-conferenced English speaking test (VEST) for classroom-based assessment in tertiary education

David Imamyartha, Mutiara Bilqis*, Rizki Febri Andika Hudori, Areta Puspa, Eka Wahjuningsih
Department of English Education, Faculty of Teacher Training and Education, Universitas Jember, Indonesia

***Corresponding Author**
Email: *bilqis.fkip@unej.ac.id*

Received:	Revised:	Accepted:	Published:
17 October 2022	10 January 2023	26 February 2023	28 February 2023

Abstract

A proper speaking test may pose challenges to teachers due to suboptimal assessment literacy and the laborious administration of a speaking test. Technology can help alleviate the labor while enabling teachers to ascertain the accuracy and validity of the speaking test. In this direction, this study aims to develop a videoconferenced English-speaking test (henceforth VEST) with the aid of videoconferencing technology. Aligned with the empirics on the use of technology for speaking assessment, this study addresses the gapping void in how videoconferencing technology can be harnessed to escalate the practicality and positive impacts of a classroom-based speaking test in an Indonesian EFL setting. The development adhered to the ADDIE framework involving Analysis, Design, Development, Implementation, and Evaluation. The test prototype was evaluated using a test usefulness analysis framework. The evaluation results were coupled with reflection results of test takers' experiences to guide test revision. Given the small sample size, this study has unveiled the potential of VEST to improve the praxis of speaking assessment and the resultant test takers' experience. Implications and recommendations for future studies are also discussed.

Keywords: *speaking; video-conferenced English-speaking test; VEST; technology*

INTRODUCTION

This study is inspired by the assessment praxis in a course on Speaking for Academic Purposes at an Indonesian university. A preliminary interview with two course teachers unraveled the extensive use of audio and video-recorded performance as the test delivery modes. Although the teachers aimed at assessing students' authentic performance in real-time, this had never been possible due to a large amount of time required to administer the test during online instruction. This substantially influences the accuracy and validity of the test and, as a result, lowers students' efforts to take the speaking test (Hirai and Koizumi, 2009).

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

This assessment praxis is in stark contrast to the demand for speaking performance to be evaluated by carefully considering validity issues, such as the scope of the test construct (Hirai and Koizumi, 2009; Schreiber et al., 2012). Discussion on classroom-based assessment of speaking performance has started to gain attention (e.g., Hirai and Koizumi, 2009; Kehoe et al., 2021; Schreiber et al., 2012); yet teacher-tailored assessment generally did not focus on communicative competencies due to its extensive linguistic constructs. The issues mirror the general challenges of speaking tests documented in the literature. Another study by Galaczi et al. (2011) reported the overemphasis of general assessment scales for speaking tests, especially due to the labor in administering speaking tests.

Technology can be a powerful resource to alleviate the burden of speaking tests, while allowing test raters or teachers to ensure the test validity (Nakatsuhara, et al. 2020; Wainfan and Davis, 2004). Rapid development in digital technology has aided in collecting and delivering test-takers' performances much more efficiently and simply in audio or video format. The development has transformed the praxis of examination, in which the test administrator and rater seriously pay attention to the mode and scoring of speaking tests (Nakatsuhara, et al. 2020). However, given the necessity to ensure language assessment quality as the springboard for evaluating the quality of language instruction, the development of a speaking test in online language teaching in Indonesia remains underexplored. The imbalance is manifested in the current discourse on the use of technology for language instruction where techniques, strategies, and approaches to language teaching are a massive proportion.

There is a small burgeoning body of literature concerning the employment of videoconferencing for the assessment of communicative speaking performance. Some previous works document the design properties such as maximum parallax and screen size (Grayson and Monk, 2003) and whether test takers pay attention to the video shown to them (Wagner, 2007). Another line of inquiry documents small-scale studies in language instruction through a videoconferenced speaking test (Xiao, 2007) or the portrayal of large-scale praxis using videoconferencing tools in an education university (Byrne and Staehr, 2002). This literature, however, implies a gaping void in relation to videoconferencing for testing speaking performance, which is a void in this specific research topic, particularly in the Indonesian EFL context.

In this study, the test development aims to develop a classroom-based assessment with the aid of videoconferencing technology, focusing on spoken proficiency in the contexts of paper presentation, picture-cued description, and dialogue in focus group discussion. The integration of videoconferencing technology stems from the increasing trend in online language instruction in Indonesia, which unfortunately has made the praxis of online language assessment lag behind due to practicality concerns. At this preliminary stage,

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture, 8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

we attempt to develop a videoconferenced English speaking test (VEST) to address the practicality and wash back issues highlighted in the literature. This research aims to address the following questions:

a) What is the extent of VEST usefulness from raters' and test takers' evaluation?
b) How do test takers experience their engagement in VEST?

In the following section, the authors described the rationales and theoretical underpinnings of the development and validation argument of VEST. As the present study served as a pilot study on the test development, with only six participants of the same cohort in an English Language Education program, the authors warranted careful transferability and generalizability of the research findings and implications, including the extent to which VEST was deemed *valid*. In consequence, the methodology section provided a rich description of the test blueprint and test validity assurance strategies in order to allow interested readers to judge the extent to which the current VEST design satisfied multiple validity criteria. Furthermore, rather than viewing propositional theories on test validity as absolute, interested parties can reflect on and determine the extent to which findings of this study make sense and apply to their praxis (Whitehead, 1989).

LITERATURE REVIEW

Ensuring a Validation Argument in Test Development

Test development calls forth rigorous validation (Bachman and Palmer, 1996). Validation is well-known as the extent to which theoretical underpinnings and empirical evidence are aligned with the expected appropriateness and adequacy of interpretations as well as actions on the basis of test scores (Messick, 1989). In this direction, Chapelle et al. (2003) proposed a validation process in which testers aim to establish validity arguments coupled with empirical evidence and validation theories backgrounding the tests under development.

There is no one specific validation theory that explains the process relevant to videoconferenced speaking tests (Kim and Craig, 2012). Chapelle et al. (2003) argued that there is no distinctive guideline on what needs to be included in a validity argument. They believe that high-stakes assessment calls forth more rigor in validation considerations than low-stakes one. Throughout a validation, test developers need to collect all evidence suitable to test validity in consideration of resultant scores and the data supporting a validity argument, be it qualitative or quantitative.

Attending to a progressive validity definition proposed by Messick (1989), Chapelle (1994) developed a validity table to generate a validity argument. She delved into the justification of using C-tests for measuring vocabulary mastery among L2 learners. Test developers aim at collecting evidence against or in favor on the basis of construct validity, utility/relevance, value implications, and social impacts as the components for building validity arguments. Following the validity argument proposed by Chapelle (1994), Read and

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

Chapelle (2001) extended the coverage by including the impact of validation measures which serve as the bearing factors to test development and validation. Figure 1 portrays the roles of agents as well as the objective of gathering validity arguments throughout a test development. Generally, test objectives stemmed from the impact of stakeholders on the measures needed to formulate validity considerations.

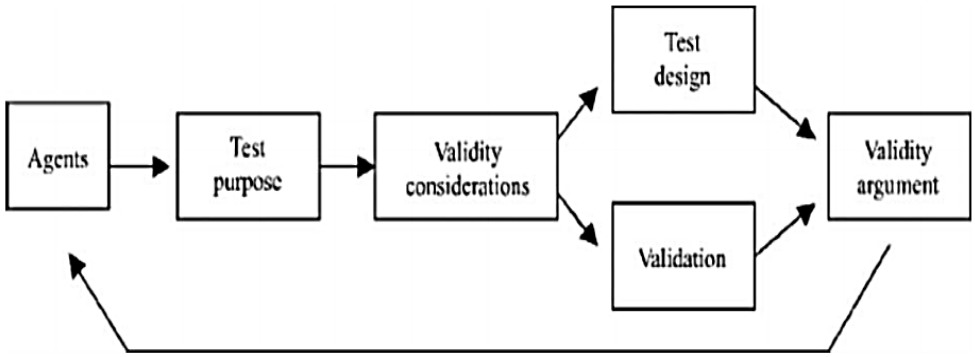

Diagram 1. Cyclical process in building validity argument (Chapelle, 1994)

The decision making bound to these concerns influences test development and validation, and eventually, the validity argument is amassed to evince test validity. The arguments comprise both positive and negative attributes of the test, providing continuous feedback for the cyclical process.

One strategy for building a validity argument refers to the usefulness analysis developed by Bachman and Palmer (1996). They pointed out six properties of test usefulness in evaluating theoretical and empirical underpinnings of test development: reliability, authenticity, interactiveness, practicality, construct validity, and impact. Reliability deals with test score consistency between similar tests and test takers' performance. Construct validity describes how accurately a test measures a specified set of competencies. Interactiveness deals with the extent and also type of engagement (Bachmand and Palmer, 1996) in accomplishing a task. Authenticity deals with how much test performance corresponds to the actual language use in real settings. Impact is concerned with the consequence of a test on educational and social systems, including individuals within. Practicality is about the match between test design, development, its administration, and required resources.

Writing test specifications (henceforth test specs) can aid in building a validity argument. Test specs are described as a test blueprint and generally consist of guiding language and sample items. The former consists of a general description, prompt, response, and specification supplement. Kim (2006) pointed out measures to build validity arguments through auditing a spec-

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture, 8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

driven test. This audit requires a test developer to record the evidence at each phase of developing test specs. This evidence manifests in an audit, the trail of which is carried out in the chronological order of validation (Schwandt and Halpern, 1988).

The Roles of Test Modes in Speaking Assessment

Weir et al. (2013) argue that face-to-face speaking test is widely operationalized in language testing and throughout the course, it has been found advantageous in terms of validity considerations, such as the positive effect on language learning and the foundational interactional construct. Howbeit, the "right-now-right-here" property of a face-to-face test and the challenges of involving trained examiners in a large-scale test have resulted in low practicality.

The nature of resource-intensive demands associated with a face-to-face speaking test has promoted practical alternatives, such as 1) a semi-direct speaking test which involves machine-assisted prompts and grading by human raters, and 2) the automatic speaking test. In spite of studies that have reported on test scores and difficulty level comparisons between computer-mediated tests, face-to-face tests, and construct comparability (Kiddle and Kormos, 2011), empirical studies and theoretical footings which exceed a mere score comparison have underlined the construct-related difference between these test modes. It lies in the fact that speaking represents a social interactional and cognitive process (Galaczi, 2010; van Moere, 2012).

Weir and Saville (2002) have underlined language skills activated in different modes. These differences are found with respect to cognitive validity since the mode of test delivery affects resultant cognitive processing. For example, the face-to-face format includes perceiving input from interlocutors, adapting to different viewpoints and topics, and making judgments regarding the accommodation of the language used by the interlocutor (Field, 2011). In addition, the perceptions of test takers have been found different with regard to the test mode delivery, with some studies (Kenyon and Malabonga, 2001) reporting that test takers voice nervousness and limited control in a semi-direct test since their role is dictated by machine, which nullifies the possibility of seeking help in the case of test taker difficulty.

Videoconferencing for English Speaking test

It is widely acknowledged that face-to-face interviews denote a well-known strategy to test students' linguistic competencies. However, they can be somewhat challenging to administer due to practicality issues (Bachman and Palmer, 1996). In addition, access to competent raters is difficult and expensive in most Indonesian universities. In order to address these challenges, the present study highlights the use of a videoconferencing tool, such as Zoom as an alternative to alleviate the burden of administering a large-scale speaking test. In addition to its potential of mirroring a face-to-face speaking test, a videoconferenced speaking test has been well known to address the lack of recordings and telephones which have been extensively used in speaking tests,

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

particularly due to the issue of authenticity, validity, and interaction. This study aims to propose videoconferencing to tackle the issues.

Videoconferencing offers a significantly closer experience resembling face-to-face tests than telephone interviews and recordings. This mode of testing allows learners to rely on the facial and vocal cues that will otherwise be absent in other modes (Wainfan and Davis, 2004). In addition, the possibility of directly engaging with an interlocutor amplifies the sense of social presence (Nakatsuhara, et al., 2020). All of these aspects are viewed as vital for ensuring successful communication (Gruba, 1997; Wainfan and Davis, 2004). By implication, a videoconferenced speaking test holds the potential to provide a more authentic and robust medium for speaking assessment.

Unfortunately, the use of videoconferencing technology for administering a speaking test in language class has been underexplored (Wang, 2004, 2006; Xiao, 2007). Also, its implementation in a classroom-based or large-scale speaking test has only received scant emphasis. Heins et al. (2007) found that most studies on videoconferencing in English language teaching have only delved into reporting classroom exchanges with distant classrooms and speakers, and remote tutoring. Project reviews on videoconferencing mechanisms have also received extensive emphasis (Lee, 2007; Wang 2004).

This study addresses one issue concerning how test delivery affects test takers' performance, which is more crucial than how videoconferencing is harnessed for language assessment or whether students or test takers are satisfied with the technology. Recruiting forty English learners in Korea, Craig and Kim (2010) investigated the differences with regard to test takers' anxiety between videoconferenced and face-to-face speaking tests. Their findings documented no significant difference in anxiety levels between the two modes of test, and thus the selection of test mode was proven insignificant threat to test validity. Further investigation on the validity of the videoconferenced speaking test was initiated by Kim and Craig (2012). Their analysis focused on scores from both modes, which dealt with functional competence, coherence, accuracy, and fluency, as well as interactiveness. In addition, they investigated forty test takers' feedback on anxiety in both modes by integrating "nervousness" and "comfort" with the speaking test, test rater, and test environment. Their analysis identified no significant difference in terms of analytic and global scores. They also found that these two modes were convenient to most test takers (Kim and Craig, 2012, p. 268). By implication, the videoconferenced speaking test held several useful characteristics, consisting of practicality, construct validity, interactiveness, reliability, authenticity, and impact (Bachman and Palmer, 1996). Another difference was concerned with the higher anxiety prior to the face-to-face test.

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture, 8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

METHODOLOGY

Test Development and Validation Implications of Test Objective

This study employed Kim's (2006) audited spec-driven test, which drew upon the development of the spec-driven test (Davidson and Lynch, 2002). The language test was developed through Kim's (2006) ADDIE model, which involved Analysis, Design, Development, Implementation, and Evaluation. In addition, Schwandt and Haplern's (1988) concept of the audit was implemented by analyzing any written documentation throughout the test development.

As the basis for developing the test blueprint and validation assurance strategies, we relied on the course learning objectives and then worked on other test attributes by referring to these objectives and the target test takers. What follows are the blueprint and validity assurance strategies.

Table 1. Test Blueprint and Validity Assurance Strategies

No	Design	Descriptions
1	**Goal of the test**	This test aims to assess communicative speaking competence in an academic context. It emulates the target language behaviors found in a general and academic setting, e.g. campus.
2	**Target test takers**	• The test is designed for undergraduate students (S1) majoring in the English Language Education program as the requirement for passing a course entitled Speaking for Academic Purposes. As stated in the course profile below, "*This course is designed to develop students' speaking skills in advanced level, such as employing important language functions in presenting current issues, panel discussions and debates. It is also designed to enhance students' ability in giving presentation such as presenting research articles".* • This target performance (advanced level) corresponds to the C1 level in the CEFR benchmark (please check link for description of CEFR). What follows is the descriptor of C1-level speaking performance: *"Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices"* (Council of Europe, 2001)
3	**Inferences Test screen familiarity**	• Test takers need to be familiar with the use videoconferencing
4	**Use: low-stakes testing**	• The test aims at providing feedback and suggestion for improving students' language proficiency in the classroom level
5	**Performance**	Given a context based on a picture and/or written

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

	Indicators (Scoring criteria)	scenario, the test takers are assessed with regard to the following performance indicators, with their respective weighing: a) Fluency and Coherence (25%) Taking part effectively in picture-cued monologue and extended dialogue through face-to-face communication by providing relevant and intelligible responses b) Pronunciation (25%) Using accurate articulation, intonation, and stress patterns of a wide range to deliver descriptions and explanation related to an assigned topic in an academic setting c) Lexical Resource (25%) Using appropriate lexical items of a wide range to deliver basic to complex ideas relevant to an assigned topic in an academic setting. d) Grammatical Range and Accuracy (25%) Using diverse structural items accurately to deliver complex ideas relevant to an assigned topic in an academic setting.
6	**Validity consideration**	a) Contexts, topics, and situations are made relevant to the course syllabus b) Test constructs need to be aligned with the course learning objectives c) Test takers need to be given a description or briefing on the videoconferencing tool used in the test d) Raters need to use an informative scoring rubric to help the delivery of feedback with the aid of Google form
7	**Validity Assurance Strategies**	***Standardization of test setting*** a) Discussion with the instructors teaching the course to check the alignment between the test and the course objectives and the clarity of test guidelines as well as scoring rubric b) Trial to prospective test takers to check the clarity of the test guideline ***Standardization of rater*** a) This test requires a rater with a qualification of at least a master's degree in English Language Teaching and language proficiency of at least C1 (e.g. IELTS band at least 7 or TOEFL score of at least 500) b) Training with sample performance for inter-rater scoring is compulsory. Sample performance **at band 7** is available on this YouTube channel (please check

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture, 8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

link for sample).

Standardization of rating procedure

c) Standardization applies in the scoring rubric and test setting (test duration and procedure)
d) The moderation requires no more than a 2-point difference between raters
e) Statistical analysis involves reliability test with SPSS

Empirical Experiment

Participants

This experiment involved two Indonesian teachers of English, one male, and one female, with individual teaching experience of 6 years. The first teacher, a male, had been extensively assigned to teach all three speaking courses (English for Survival, Public Speaking, and Academic Speaking). He taught these courses in even and odd semesters, therefore implying a stronger background and experience in speaking instruction and assessment than the other. The other teacher, a female, had been assigned to one of the three speaking courses. This course was only offered in the even semester. These teachers were purposively involved to find out different perspectives and evaluations on the test design, which aimed at ensuring that the test design was applicable to different test raters or teachers. Prospective raters were also chosen by considering the ideal linguistic proficiency required for the assessment. Prospective raters needed to meet C1 level (advanced language user) within the CEFR scale, as indicated by a language proficiency certificate in TOEFL with at least 500 points, or IELTS with at least a band score of 6. The test takers were 6 students from the English Education Department, involving 5 females and one male. Their identities were only reported as initials in this study. Their ages ranged from 20 to 21. These students had taken all three speaking courses, so their experiences in taking the speaking test were supposedly uniform.

Procedure

Figure 2 portrays the overall procedure for developing the videoconferenced speaking assessment, which adhered to the ADDIE phases throughout the process. *Analysis* was conducted in the first phase, followed by *Design* in the second phase. *Development* was operative from the third up to the sixth phase. *Implementation* was conducted in the seventh phase, and *Evaluation* took place in the eighth phase.

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

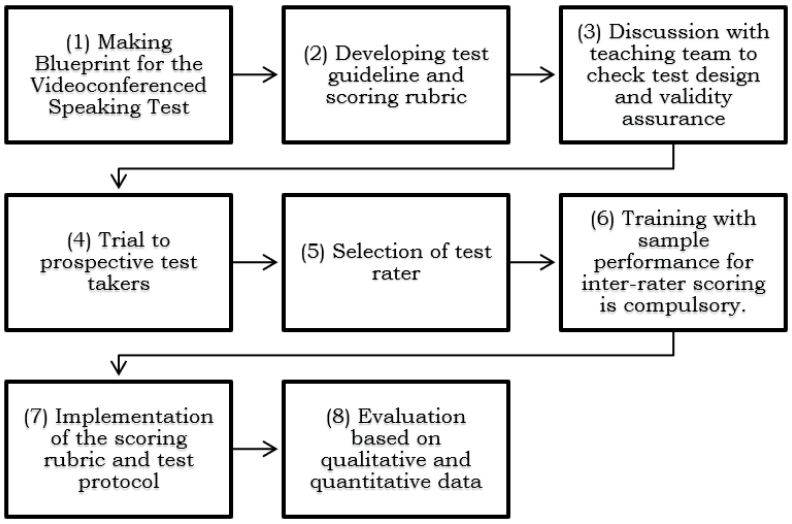

Diagram 2. Procedure of Test Development

The analysis focused on the course description and course learning objectives as the basis for designing the test blueprint, test prompt, and scoring rubric. Afterward, the interim prompt, test guideline and scoring rubric were evaluated through discussion involving a pair of teachers in the speaking course. This aimed to check the alignment between the test and the course objective and gain teachers' feedback on the initial test design. In light of assuring the practicality and intelligibility of the test from the students' perspectives, a trial was conducted with 6 prospective test takers. The following figure exemplified one sample performance in the second test section, describing the topic on a card.

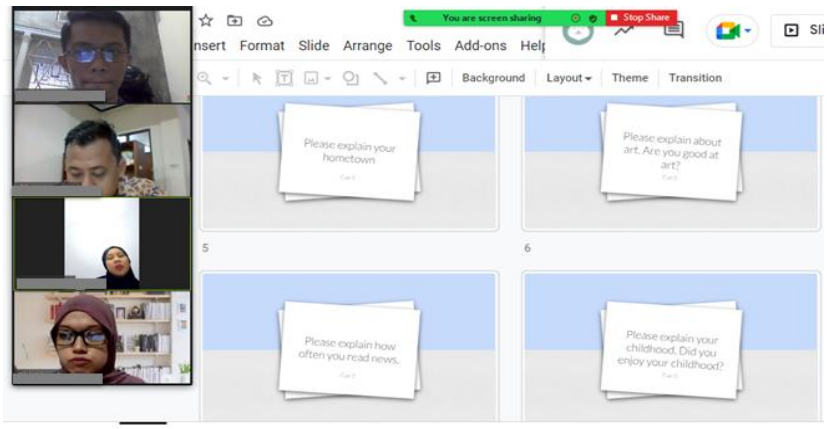

Figure 1. Sample of Test Administration in the Tryout

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture*, *8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

After the trial, test evaluation sheets were distributed to the raters and test takers, to evaluate the test's usefulness and gain feedback for revision. After the tryout, a moderation of raters' scoring was applied by setting a maximum of the 2-point difference between raters based on a 10-scale scoring rubric. The test results and responses from the tryout participants served as the basis for evaluating and revising the test.

Data Collection and Analysis

Quantitative data were taken from the test score reliability between raters, which was collected from the test scores garnered during the trials. The scores were evaluated by a reliability analysis using SPSS. The qualitative data were garnered through two interviews using *Bahasa Indonesia* with raters and test takers. The interview was run to reveal teachers' feedback on how the test design and guidelines facilitated the administration of the speaking test. In addition, interviews were carried out with the test takers after the trial. Both of the post-test interviews focused on the attributes of test usefulness, with a test evaluation sheet employed to elicit responses from the raters and test takers. Eight additional questions adopted from Kim and Craig (2012) were raised during the post-test interview with test raters and test takers, to derive suggestions for test revision. Each of these questions was concerned with computer familiarity, comfort/anxiety, raters' effect, environment, gestures and facial expressions, interest, speaking opportunity, and topic/situation effect. Prior to analysis, all data from the transcribed interviews and evaluation sheets were crisscrossed among authors for examining information consistency and accuracy. All of the qualitative data were analyzed deductively by referring to a test usefulness framework (Bachman and Palmer, 1996) and test experience (Kim and Craig, 2012), in consideration of comprehensive test evaluation.

FINDINGS

The Usefulness of VEST

The first aspect of test usefulness under analysis was reliability. Descriptive analysis was performed on trait-specific scores from six test takers. Compensating the small sample size, this analysis sought to identify raters' agreement in each component of the scoring rubric, i.e., fluency and coherence, lexical resources, grammatical range and accuracy, and pronunciation. This resulted in a total of twenty-four individual trait-specific scores from each rater. The analysis result is hereby presented (Please see Appendix 6 for the complete score report).

Table 2. Summary of Inter-rater Agreement Analysis

Aspects	Mean	Rater 1	Rater 2	Range	Variance
Score Means	7.729	7.708	7.750	.042	.001

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

Score Variances	.640	.543	.737	.194	.019

The reliability analysis between rater 1 and rater 2 demonstrated score means of 7.708 and 7.75 respectively. The analysis identified Cronbach's Alpha of 0.866 with an inter-rater correlation of 0.773. The statistics clearly demonstrated fairly high agreement between raters. Likewise, a decent inter-rater correlation also confirmed the absence of a rater effect on scoring. Raters' feedback on the test evaluation sheet also informed the evolution of test specs. Their feedback is summarized in Table 3.

Table 3. The Evaluation of Test Usefulness by Raters & Test Takers

Quality	Evaluation	Mean
Reliability	1. The scoring rubric includes clear gradation.	4.88
	2. The scoring rubric uses clear descriptors to differentiate test takers' performance	4.88
	3. The test platform (Zoom™) and guideline are familiar to teachers.	5.00
	4. The test environment (as in the guideline) can be made consistent for test takers.	4.63
Construct Validity	1. The test objective is relevant to the course objectives in the English Education Department.	5.00
	2. The test construct includes linguistic skills relevant to target test takers.	4.75
	3. The test construct includes topics relevant to target test takers.	4.25
	4. The test requires strategies to perform interactive spoken communication.	4.75
Authenticity	1. The test situation/setting is similar to the communication in class situation/setting.	4.13
	2. The test tasks (describing picture, monologue, and dialogue) are generally found in academic setting.	4.88
Interactiveness	1. The test tasks allow the interaction between test takers and raters.	4.50
	2. The test platform, Zoom, allows the interaction between test takers and raters.	4.88
Impact	1. The test helps test users identify students' strengths and weaknesses.	4.50
	2. The test helps test users determine the success of learning activities.	4.63

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture*, *8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

	3. The test helps test users to revise and improve the course design and learning strategies.	4.88
Practicality	1. The administrative details are clearly explained before the test.	4.88
	2. Students can finish the test within the timeframe.	4.50
	3. The test can be administered smoothly without procedural "glitches".	4.75

Test Takers' Experiences

The data from the interview were deductively analyzed by relying on the factors driving test experiences. Table 4 displays the themes concerning the test takers' experiences.

Table 4. The Results of the Post-Test Interview

Quality	Themes
Comfort/anxiety	Test takers were nervous only upon test commencement, yet gradually felt comfortable with the test.
Computer familiarity	Videoconferencing technology was part of the test takers' study routine.
Interest	Test takers were willing to take another future test due to the practicality and interactivity concerns.
Environment	No significant problem was found, except for the absence of a test timer for the test takers.
Rater effect	Rater accommodation and feedback were vital to increase test takers' comfort and confidence.
Topic/situation	The topic/situation was familiar to test takers.
Speaking opportunity	Test takers acknowledged sufficient speaking opportunities in the three-section speaking test to satisfy their test experiences.
Gestures/facial expressions	Gestures/facial expressions were fully accommodated and supportive to test takers' performance.

First, all of the test takers found the test setting and tasks comfortable as videoconferencing was a regular activity in their study. This allowed them to make sure that their surrounding was comfortable and supportive for their optimal performance, compared to a face-to-face setting where generally many other test takers are present. Since they were unfamiliar with the overall test guidelines, they were slightly nervous upon beginning the test; yet they eventually had everything under control, especially in the conversation section. The opportunity of seeing raters'

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

gestures and facial expressions made the test more humanistic due to its close resemblance to face-to-face communication. SA mentioned that "*it was really assuring to see raters' expressions and that improved the comfort and confidence in taking the test*". This was in line with DK's statement claiming that "*the video helped to reflect on the ongoing performance and monitor the gestures and facial expressions".* Although anxiety was in fact present, this had nothing to do with the test delivery mode. JW explained that *"the anxiety occurred because of the lack of preparation for the test, so earlier preparation would lead to better performance"*.

Second, test takers' familiarity with videoconferencing played a vital role in maximizing their performance. They had been using videoconferencing for the last two years due to emergency remote teaching and found no issues engaging in a videoconferenced speaking test. The preference for videoconferencing, over face-to-face tests, was confirmed in the interview. SB stated the benefit of *"arranging the test setting and situation, such as room setting and noise cancellation, and this was helpful to make sure maximum performance"*. Another participant, ND also valued "*the opportunity of using and receiving non-linguistic cues, which mirrored face-to-face interaction*".

With the practicality and close resemblance to a face-to-face speaking test, the test takers reported interest in taking a future VEST to assess their proficiency. The online setting implied fewer temporal and spatial hurdles on the part of the test takers. This interest also stemmed from the humanistic communication of videoconferencing to mediate both linguistic and non-linguistic cues. Interlocutors' facial and gestural expressions played important roles in facilitating continued communication since this was able to support their communication strategies. One participant, SB, explained that *"it was nervous at first, but seeing raters' responses and receiving their feedback made the test experience relaxing and less stressful"*. What is more, RS believed that "*using gestures in the videoconferenced speaking test would help listeners understand the messages".*

Test takers reported no issues with the use of videoconferencing for assessment purposes. Given extensive videoconferencing experiences in their study, they acknowledged the value of such a technology-mediated environment to build their communication strategies, while lowering anxiety due to the indirect encounter with raters. In this direction, SB further explained that "*it was comforting to take such videoconferenced test because of the opportunity of hearing and seeing raters' responses and the indirect communication through Zoom".* However, the absence of a test timer on the part of test takers posed problems during the test. SA mentioned that "*it would be a lot easier for test taker to have test timer shown on screen to monitor the overall test performance".*

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture*, *8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

Fifth, raters' ability to give feedback and confirmation through linguistic and non-linguistic cues was deemed pivotal for sustained communication. Test takers highly valued this affordance of videoconferencing as it helped them to ensure fine performance as they received feedback, as the case in the conversation section. Eventually, this helped to make the test experience as natural as possible. SA also acknowledged that she "*appreciated raters' responses as this helped make sure that the speaking performance was fine".* She further mentioned that "*noticing raters' excitement in the talk really increased the motivation to speak more".* Voicing the same idea, ND clarified that *"raters' responses could reduce anxiety by clarifying and confirming test taker's ideas".*

Finally, test takers confirmed a close resemblance between face-to-face communication and VEST, thanks to the inclusion of three different speaking modes across sections. These modes or tasks were seen as authentic to their daily communication experiences in an academic setting. In addition, the topics were generally found to be familiar to test takers. SA explained that *"the three tasks were simply like the ones regularly encountered in class communication, so these helped to reach maximum performance and keep the communication going as long as possible".* The possibility of choosing the pictures or cards to talk about was also deemed useful in helping test takers perform at their best. JW pointed out that "*the selection of pictures would improve the likelihood of extended talk and confidence during the test due to comprehensive background knowledge".*

The aggregates of raters' feedback and test takers' reflections were taken into account for developing the validity arguments and the test spec evolution. Table 5 describes the evolution. The final versions of the test guideline, test prompt, and scoring rubrics are included in the appendices.

Table 5. Test Spec Evolution

Test Versions	Evolution/Revision
Spec version 1.0	1. A sample video from YouTube is necessary for pilot scoring. 2. The nine-scale IELTS band descriptor is simplified into 7 scales for practicality reasons.
Spec version 2.0	1. The whole test sections are all recorded for future re-grading and moderation between raters. 2. Each rater should take a turn in guiding the conversation in Section Three.
Spec version 3.0	1. A countdown timer is shown on the screen throughout each section. 2. Topics and cards should also relate to language and language education.

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

DISCUSSION: VERDICT ON TEST USEFULNESS

The validity argument was garnered by attending to the test usefulness framework. This overall verdict was grounded within the test specs evolution, qualitative data, and speaking test results. The argument portrayed the positive and negative sides, which were drawn from the test objectives, design decisions, validity considerations, and validity assurance strategies.

First, the employment of the videoconferencing tool could be viewed as a reliable mode of assessment with regard to the targeted constructs. This is in line with previous works reporting on the comparability between face-to-face and technology-mediated speaking assessments. Nevertheless, the evidence arising out of empirics was rather limited due to a small number of test takers and, by implication, the narrow diversity of linguistic proficiency. In addition, the brevity of the briefing for the test raters, only conducted in a one-hour online meeting, might also influence their accuracy and stringency or leniency upon scoring. These required further clarification in future research.

With regard to construct validity, the validity argument drew on the test development process by involving situations and topics specified in the course objectives. These decisions attended to test development theories proposed by Kim (2006). Notwithstanding, complete speaking competence constructs might be hardly identified on the basis of empirical footing in the interpretations of test results, particularly due to the small sample size, narrow variety in language proficiency, and the tasks in the speaking test which were not distinctively developed for a wide range of proficiency levels.

The speaking test was considered relevant to the test takers' language use in a real-life setting, especially mediated by computer technology, such as the one via Zoom. This confirmed the authenticity of the VEST. Following Bachman and Palmer (1996), authenticity, situations, and topics of test tasks are authentic when they are deemed meaningful for test takers in their real use of language. The qualitative findings evinced that the test takers were familiar with the situations and topics involved in the speaking test. Otherwise, authenticity is likely to deteriorate due to the limitation of the correspondence between test takers' linguistic proficiency level and the test tasks.

Fourth, test interactiveness was finely achieved in VEST since it resembled face-to-face interaction. Howbeit, the video quality, small screen size, and an internet connection issue limited the flexibility in using and perceiving facial and gestural expressions between interlocutors. This might significantly influence test takers' speaking performance. As such, technological affordances, engagement based on Second Language Acquisition (SLA) concepts, and strategies used by test takers may become potential research areas in the future.

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture*, *8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

Also, the influence of videoconferencing on test takers' performance, testing experience, and positive washback on their test preparation call forth further research. In Indonesia, where the technological divide still prevails, videoconferencing might offer prospective benefits for increasingly popular online modes of learning and testing; yet this also holds some drawbacks in terms of practicality and, more importantly, educational equality among language learners. Those in resource-constrained regions may find VEST a daunting task and thus hesitate to even prepare their performance.

Eventually, the practicality associated with VEST has been escalated with the improvement in information, communication, and technology (ICT) in Indonesia. Specifically, the price of personal computers and videoconferencing devices, such as webcam, has declined. This generates increasingly wider access to VEST. Table 6 summarizes the overall test validity arguments based on the findings.

Table 6. Test Validity Arguments

Quality	Positive Attributes	Negative Attributes
Reliability	1. The spec-driven test development helped to elicit relevant speaking performances. 2. The test scores by inter-rater grading demonstrated fine agreement.	1. The small number of test takers made it difficult to gain richer portraits of proficiency levels. 2. Due to brevity of training and briefing for raters, an interviewer effect was likely to occur and thus reduce the scoring accuracy.
Construct Validity	1. Test topics, task, and setting were designed similar to the target language performance. 2. The constructs were developed based on the course outline and standardized language benchmarks, i.e., CEFR. 3. All constructs involved were put under observation on test takers' performances.	1. The small sample nullified empirical level cuts.
Authenticity	1. The test topics, tasks, and setting were selected from	1. Some test takers were not used to having a

EduLite *Journal of English Education, Literature, and Culture*
Vol. 8, No. 1, February 2023, pp. 242-264

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

	a specified domain of language use, i.e., an academic setting. 2. All topics, tasks, and settings corresponded to the test takers' real lives.	test via videoconference.
Interactiveness	1. Test takers have been widely familiar with videoconferencing. 2. VEST was highly interactive. 3. Videoconferencing allowed interlocutors to use and interpret non-linguistic communication for more interactiveness.	1. VEST might limit the use of non-linguistic cues due to small screen size. 2. Studies were called upon to unravel the strategies test takers used to engage with the test tasks.
Impact	1. The test aimed at giving a positive and exciting experience to participants by the use of videoconferencing technology. 2. VEST was aimed at positive washback-impacts on test takers, presumably encouraging them to prepare the test better and improve their proficiency independently.	1. The drawbacks in terms of practicality and, more importantly, educational equality may pose serious challenges among test takers, with those having technological resources being at an advantage.
Practicality	1. VEST was developed by considering common technology available to Indonesian students. 2. Online test administration was found flexible without significant spatial and temporal hurdles.	1. Test administration and test takers' performance might be hampered due to connection issues.

CONCLUSION

This study aimed to demonstrate the validation of a VEST by employing the procedure of spec-driven test development while attending to the concept of validity theories by means of emulating constructs engaged in a targeted language course. A validity argument was collected throughout the test development and administration, which involved both qualitative and

quantitative data as the basis for determining conceptual robustness and validity of the test as well as test spec evolution.

Attending to the test purpose, the authors formulated validation considerations and validation assurance strategies which were continuously revisited and refined to make validation decisions. In the end, the authors carried out a usefulness analysis that was built upon the audit trail of the test spec evolution and the analysis results of qualitative and quantitative data. Test spec evolution documented the selection of test tasks, formats, settings, topics, grading, and technology. These attributes supported the rationales of the test validity argument related to positive and negative sides. The positive sides were vital in generating validation evidence, while the latter helped identify the test drawbacks and potentials for future improvement. Quantitative analyses demonstrated that the test helped to gain reliable scoring between raters. This implied that the test prompt, guideline, and scoring mechanism were all easily perceivable to raters. In addition, the overall procedure of VEST was well executed by raters.

The qualitative findings unveiled that the test takers showed positive responses in terms of comfort level, environment, computer familiarity, opportunities for speaking, and situation as well as topics. Notwithstanding, some test takers were hampered by low technological resources and the absence of non-linguistic cues in their performance. These findings contribute to improving the test validity of VEST in the Indonesian English language teaching (ELT) context.

The validity argument was aggregated in the test usefulness analysis. Reliability was achieved due to the lack of significant differences between the raters' scores. Construct validity was built by observing raters and test takers throughout the tryout. Authenticity was ensured through the test topics and settings as well as situations that were made similar to those in the test takers' lives. Furthermore, videoconference elevated the interactiveness between the rater and test taker, despite limited non-linguistic cues. The overall test formats resulted in positive washback, which is likely to influence students' learning. This has been made possible by the affordability of computer technology. The findings confirm the authors' assertion that videoconferencing assessment is suitable to complement and elevate the current praxis of speaking tests made possible by technology. This helps to increase the opportunity of engaging more qualified and trained raters to assess students' performance and the effectiveness of language learning.

The small-scale sample in this study sheds light on areas of test development worth further investigation. The fact that raters also taught the test takers might have influenced their scoring attributes. Involving a greater number of participants will help gain ideal and more robust reliability

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

analysis results. By the same token, a larger scale would also garner more comprehensive data as the basis for test evaluation and revision.

AUTHOR STATEMENT

Author 1: Conceptualization, collecting data, analyzing data, writing the manuscript, and compiling references. **Author 2:** Collecting data, analyzing data, compiling references, and proofreading. **Author 3:** Collecting data, analyzing data, compiling references, and proofreading. **Author 4:** Collecting data, analyzing data, compiling references, and proofreading. **Author 5:** Supervision, providing advice, and proofreading.

ACKNOWLEDGEMENTS

We would like to thank all of the members of our research group for their tremendous input, insight, and assistance in every step in the research. We sincerely thank everyone who helped us along the way, especially our students who volunteered to be our research participants. This was an independently funded study that was part of our research group's agenda.

REFERENCES

Alderson, J.C., and Banerjee, J. (2002). Language testing and assessment (Part 2). *Language Teaching, 35,* 79–113. https://doi.org/10.1017/S0261444802001751

Bachman, L.F., and Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press. https://octovany.files.wordpress.com/2013/12/language-testing-in-practice-bachman-palmer.pdf

Byrne, G., and Staehr, L. (2002, June). *International internet-based videoconferencing in distance education – a low cost option.* Paper presented at the Informing Science IT Education Conference Proceedings, Information Science Institute, California, USA. https://doi.org/10.28945/2451 .

Chapelle, C. (1994). Are C-tests valid measures for L2 vocabulary research? Second Language Research, 10, 157–187. https://doi.org/10.1177/026765839401000203

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council of Europe. https://www.coe.int/en/web .

Craig, D.A., and Kim, J. (2010). Anxiety and performance in videoconferenced and face-to-face oral interviews. Multimedia-Assisted Language Learning, 13, 9–32. https://doi.org/10.15702/MALL.2010.13.3.9 .

Field, J. (2011). Cognitive validity. In L. Taylor (Ed.), *Examining speaking: Research and practice in assessing second language speaking* (Studies in Language

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture, 8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

Testing, Vol. 30) (pp. 65–111). Cambridge, UK: Cambridge University Press. http://hdl.handle.net/10547/338249 .

Galaczi, E. D. (2010). *Face-to-face and computer-based assessment of speaking: Challenges and opportunities*. In L. Araújo (Ed.), Computer-based assessment of foreign language speaking skills (pp. 29–51). Luxemburg: European Union. https://www.researchgate.net/publication/281090096 .

Galaczi, E. D., Ffrench, A., Hubbard, C, and Green, A. (2011). Developing assessment scales for large-scale speaking tests: a multiple-method approach. *Assessment in Education: Principles, Policy and Practice, 18*(3), 217-237. http://dx.doi.org/10.1080/0969594X.2011.574605

Grayson, D., and Monk, A. (2003). Are you looking at me? Eye contact and desktop video conferencing. *ACM Transactions on Computer-Human Interaction (TOCHI), 10*(3), 221–243. http://dx.doi.org/10.1145/937549.937552 .

Gruba, P. (1997). The role of video media in listening assessment. *System, 25*, 335–345. https://doi.org/10.1016/S0346-251X(97)00026-2 .

Heins, B., Duensing, A., Stickler, U., and Batstone, C. (2007). Spoken interaction in online and face-to-face language tutorials. *Computer Assisted Language Learning, 20*(3), 279–295. http://dx.doi.org/10.1080/09588220701489440 .

Hirai, A., and Koizumi, R. (2009). Development of a practical speaking test with a positive impact on learning using a story retelling technique. *Language Assessment Quarterly, 6*(2), 151-167. http://dx.doi.org/10.1080/15434300902801925

Kehoe, M., Niederberger, N., and Bouchut, A-L. (2021). The development of a speech sound screening test for European French-speaking monolingual and bilingual children. *International Journal of Speech-Language Pathology, 23*(2), 135-144. https://doi.org/10.1080/17549507.2020.1750699

Kenyon, D., and Malabonga, V. (2001). Comparing examinee attitudes toward computer-assisted and other proficiency assessments. *Language Learning and Technology, 5*(2), 60–83. http://dx.doi.org/10125/25128

Kiddle, T., and Kormos, J. (2011). The effect of mode of response on a semidirect test of oral proficiency. *Language Assessment Quarterly, 8*(4), 342–360. doi:10.1080/15434303.2011.613503.

Kim, J.T. (2006). The effectiveness of test-takers' participation in development of an innovative web-based speaking test for international teaching assistants at American colleges (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign. https://www.proquest.com/docview/622019942 .

Kim, J., & Craig, D. A. (2012). Validation of a videoconferenced speaking test. Computer Assisted Language Learning, 25(3), 257–275. https://doi:10.1080/09588221.2011.649482 .

E-ISSN: 2528-4479, P-ISSN: 2477-5304
http://jurnal.unissula.ac.id/index.php/edulite
DOI: http://dx.doi.org/10.30659/e.8.1. 242-264

Lee, L. (2007). Fostering second language oral communication through constructivist interaction in desktop videoconferencing. *Foreign Language Annals, 40*, 635–649. http://dx.doi.org/10.1111/j.1944-9720.2007.tb02885.x

Messick, S. (1989). Validity. In R.L.Linn (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan. https://doi.org/10.3102/0013189X018002005 .

Nakatsuhara , F., Inoue, C., and Taylor, L. (2020): Comparing Rating Modes: Analysing Live, Audio, and Video Ratings of IELTS Speaking Test Performances, *Language Assessment Quarterly*. https://doi.10.1080/15434303.2020.1799222

O'Sullivan, B., Weir, C. J., and Saville, N. (2002). Using observation checklists to validate speaking-test tasks. *Language Testing, 19*(1), 33–56. doi:10.1191/0265532202lt219oa

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

Schwandt, T.A., and Halpern, E.S. (1988). Linking auditing and meta evaluation. Newbury Park, CA: Sage. https://doi.org/10.1177/109821408901000405 .

van Moere, A. (2012). A psycholinguistic approach to oral language assessment. Language Testing, 29(3), 325–344. doi:10.1177/0265532211424478

Wagner, E. (2007). Are they watching? Test-taker viewing behavior during an L2 video listening test. *Language Learning and Technology, 11*(1), 7–8. http://dx.doi.org/10125/44089

Wainfan, L., and Davis, P.K. (2004). *Challenges in virtual collaboration: videoconferencing, audio conferencing, and computer-mediated communications.* Santa Monica, CA: RAND Corporation. https://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG273.pdf

Wang, Y. (2004). Supporting synchronous distance language learning with desktop videoconferencing. *Language Learning and Technology, 8*(3), 90–121. http://dx.doi.org/10125/43997

Wang, Y. (2006). Negotiation of meaning in desktop videoconferencing-supported distance language learning. *ReCALL, 18*(1), 122–146. http://dx.doi.org/10.1017/S0958344006000814 .

Weir, C. J., Vidakovic, I., and Galaczi, E. (2013). *Measured constructs* (Studies in Language Testing, Vol. 37). Cambridge, UK: Cambridge University Press. http://hdl.handle.net/10547/338250.

Whitehead, J. (1989). Creating a Living Educational Theory from Questions of the Kind, 'How do I Improve my Practice?' Cambridge Journal of Education, 19(1), 41–52. https://doi.org/10.1080/0305764890190106

How to Cite (APA Style):
Imamyartha, D., Bilqis, M., Hudori. R. F. A, Puspa, A., Wahyuningsih, E. (2023). Mood developing a video-conferenced English-speaking test (VEST) for classroom-based assessment in tertiary education. *EduLite: Journal of English Education, Literature, and Culture*, *8* (1), 242-264. http://dx.doi.org/10.30659/e.8.1.242-264

Xiao, M. (2007). *An empirical study of using internet-based desktop videoconferencing in an EFL setting* (Unpublished dissertation). Ohio University. http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1194703859.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.