



Castledown

 OPEN ACCESS

Language Education & Assessment

ISSN 2209-3591

<https://www.castledown.com/journals/lea/>

Language Education & Assessment, 4(3), 105–121 (2021)
<https://doi.org/10.29140/lea.v4n3.554>

The underlying structure of Foreign Language Anxiety in integrated speaking assessment: A mixed methods study



KWANGMIN LEE ^a

YAFEI YE ^b

^a University of Iowa, USA
klee27@uiowa.edu

^b University of Iowa, USA
yafei-ye@uiowa.edu

Abstract

The aim of this mixed methods study is to identify the underlying structure of the construct of Foreign Language Anxiety in integrated listening-to-speak tasks. First, the analysis of the qualitative interviews with six postsecondary ESL learners reveals that anxiety for integrated speaking stems from four different factors: *listening*, *source integration*, *topic familiarity*, and *time pressure*. These findings informed the next study phase, the development of an initial inventory of 26 items. The questionnaire was checked for content validity and then administered to 190 respondents. Next, based on the quantitative data, Confirmatory Factor Analysis was carried out in order to identify the model that best fits the data. Model comparisons corroborate the qualitative finding that suggests FLA of integrated speaking comprises four correlated latent variables, with all model fit indices falling within an acceptable range. Overall, this study provides evidence for construct validity of FLA in integrated speaking contexts as a multifaceted structure. Findings are discussed in relation to the factors that we identified.

Keywords: integrated speaking assessment, Foreign Language Anxiety, student perceptions, case study

Introduction

In L2 assessment, integrated tasks refer to test tasks that require examinees to combine different language modalities, such as reading or listening with speaking or writing. Some of the benefits that the use of integrated tasks accords are that they align with the demands of language use in real-life contexts, providing equal amounts of background information to all test-takers, and promoting

Copyright: © 2021 Kwangmin Lee & Yafei Ye. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within this paper.

positive washback for teaching. Capitalizing on these benefits, an increasing number of large-scale tests have implemented integrated tasks as part of their testing programs (Yang & Plakans, 2012). At the same time, however, examinees engaged in an integrated task are vulnerable to the effects of construct-irrelevant variables, such as Foreign Language Anxiety (FLA). To date, a number of studies have investigated the internal structure of Foreign Language anxiety, they were primarily concerned with independent speaking in the classroom context (Cheng *et al.*, 1999; Liu & Jackson, 2008). While they have yielded some insightful findings as to the underlying structure of FLA, they cannot be generalized to the relationship between FLA and integrated language performance because the cognitive (Plakans & Gebril, 2012) and affective (Huang & Hung, 2013) complexity that integrated speaking tasks pose for test-takers is different from that posed by independent tasks. With few studies dedicated to the relationship between FLA and integrated assessment, little is known about the underlying structure of FLA in the integrated speaking context. In light of this gap, this study aims to examine the underlying structure of FLA in integrated speaking assessment by utilizing a scale developed and validated for such a context. This study contributes a piece of validity evidence to the interpretations of performances and score uses of integrated listening-to-speak assessments by providing an understanding about the relationship between integrated performance and FLA. In addition, findings of this study will be of interest to test developers and teachers alike, as they can develop a better grasp of what factors trigger FLA among test-takers during an integrated listening-to-speak task, and how it impacts their test performance.

Literature Review

Foreign Language Anxiety

As one of the most common feelings aroused in the foreign language classroom, Foreign Language Anxiety (FLA) had not been deliberately conceptualized until the mid-1980s. As a result, inconsistent findings have been generated in a number of relevant studies (Chastain, 1975; Backman, 1976) as highlighted in Scovel's (1978) review of literature. He warned that the role of anxiety in foreign language learning cannot be fully understood without specifying the type of anxiety being examined. It was Horwitz *et al.* (1986) who first conducted systematic research on FLA, arguing for the conceptualization and measurement of anxiety within the language learning context. In this vein, they proffered that FLA is "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (p. 128). Furthermore, they developed the Foreign Language Classroom Anxiety Scale (FLCAS) to measure language learners' FLA, integrating three different types of anxieties under its rubric – *communication apprehension*, *test anxiety* and *fear of negative evaluation*. Their conceptualization of FLA as consisting of different subcomponents, although not corroborated through empirical means (Thompson & Lee, 2014), has influenced the data analysis procedures and interpretations of numerous studies that followed in the ensuing years.

Much of previous research into FLA has focused on it either as predicted by a host of factors that are external or internal to the learner, or as predictive of second language performance. In a correlational study carried out by Onwuegbuzie *et al.* (2000), seven variables were found to affect the level of anxiety that college-level learners display during the input, processing, and output stages: age, expectations of achievement, perceived self-worth, perceived scholastic competence, perceived intellectual ability, perceived job competence, and number of foreign language courses taken during high school. Similarly, Kitano (2001) identified two contributing factors of FLA experienced by university students learning Japanese as a second language, including *fear of negative evaluation* and *self-perceived speaking ability*. Interestingly, the advanced-level students were found to undergo a much higher level of classroom anxiety ($r=.540$) due to the fear of negative evaluation than their

lower-level peers ($r = .237$). Also, learners who had spent some time in Japan were more vulnerable to the fear of negative evaluation ($r = .496$) compared to those students with no time spent in Japan ($r = .231$). Research into FLA as a criterion variable has indicated how it wields debilitating effects on second language performance (e.g. Elkhafaifi, 2005; Hewitt & Stephenson, 2012). For instance, Phillips (1992) found that students with high FLA generally have a negative attitude toward an oral exam and tend to produce shorter sentences with simple structures during their performance. In a study that examines FLA both as an independent variable and as a criterion variable, Matsuda and Gobel (2004) reveal that overseas experience has a significant bearing on the learners' self-confidence in speaking English in the classroom, which in turn adversely affects their speaking proficiency.

FLA and independent speaking

Some studies that looked into the interface of FLA with speaking performance of L2 learners are characterized by an alignment with independent speaking tasks. For instance, in Hewitt and Stephenson (2012), oral performance data of the participants were collected by asking them to speak about the topics that are addressed in the content courses they were taking at the time of the study.

Evidence for the underlying structure of FLA for independent speaking has been documented in a number of studies, aided by the use of the FLCAS. A good illustration includes Liu and Jackson (2008), who, in the context of Chinese EFL learners at college level, identified precisely the same structure - *communication apprehension*, *test anxiety* and *fear of negative evaluation* - as was conceptualized by Horwitz et al. (1986). Furthermore, a similar three-factor structure has been identified with data provided by Taiwanese college-level EFL learners, as analyzed by Chou (2018). While Thompson and Lee (2014) reported a similar component structure of the FLCAS for college-level multilingual learners in Korea, they added and interpreted a factor that has not been detected in the previous literature: *fear of ambiguity in English*. They identified a significant inverse relationship taking hold between fear of ambiguity and both experience abroad and English proficiency. That is, as the length of staying abroad and the level of English proficiency increased, fear of ambiguous situations decreased.

FLA and integrated speaking

A number of studies have been conducted to identify the processes and products of integrated writing performance, contributing to construct validity of the test task. However, less attention has been paid to integrated speaking assessment. Huang and Hung (2018) provide a good illustration of which latent components exist for integrated speaking of Taiwanese college-level EFL learners in terms of strategy use. Even though strategy use might not seem directly relevant to FLA, Huang and Hung (2018) can provide indirect insights into what factors or elements L2 test-takers focus on in taking integrated tasks and by extension, what potentially makes them feel anxious in an evaluative setting. This factor-analytic study corroborates a model in which three different, yet interrelated, factors underlie the construct of strategy use in integrated speaking. The first factor is *discourse synthesis strategy use*, which prompts the use of such strategies as *selecting* contents from the source text and *connecting* into a coherent response. Also identified in the model were *cognitive strategy use* and *communication strategy use*, which refer to the processing and manipulation of the source materials to create a response and the process of solving linguistic difficulties, respectively. While this study provides insight into the way in which L2 learners allocate cognitive resources in performing an integrated speaking task, it does not overtly address affective factors, such as motivation or anxiety.

In the literature, there are as of yet only a handful of studies identifying the comparability of the effects of FLA on independent and integrated speaking performance (Huang & Hung, 2013), the way in which FLA combines with other individual differences (e.g. language proficiency and topical knowledge) to affect integrated speaking performance (Huang *et al.*, 2016), and examinees' strategic behaviors to alleviate the adverse effects of FLA on integrated speaking performance (Barkaoui *et al.*, 2013). Taken together, these studies suggest the importance of incorporating individual difference factors in accounting for the nature of the construct of integrated speaking.

Taken together, in spite of the growing number of research studies addressing FLA in terms of its underlying structure by way of factor-analytic methods, a vast majority of them have drawn attention to independent speaking in the classroom context. Fewer studies have been carried out to investigate the internal structure of the anxiety that L2 examinees undergo in an evaluative setting. No research has thus far identified the underlying structure of FLA in terms of its latent components in the context of integrated speaking assessment. Addressing this question could contribute the theoretical understanding to the field with regard to the psychometric structure of FLA elicited by integrated speaking assessment. Also, by considering how FLA is triggered either by a single factor (i.e. unidimensional structure) or a host of factors (i.e. multidimensional structure), test developers and teachers in the classroom can better cope with test-takers'/language learners' anxiety in testing situations. Employing mixed methods research, proceeding from qualitative exploration to quantitative corroboration, the current study aims to address the following research questions:

- 1) Which aspects of listening-to-speak assessment make test-takers feel anxious?
- 2) Does a set of contributing factors identified as part of RQ1 find statistical support through Confirmatory Factor Analysis (CFA)?

Methodology

Mixed-methods research

In order to address the research questions, the current study is framed as mixed methods research (referred to hereafter as MMR). Among a number of research designs available in MMR, this study draws on an exploratory sequential design for the purpose of instrument development and validation (Creswell & Plano Clark, 2011). The exploratory sequential design, according to Creswell and Plano Clark (2011), is the collection and analysis of qualitative data followed by the development of an instrument whose functionality is determined on the basis of the results of quantitative testing. The motivation to place qualitative research at the beginning is grounded in the fact that only few theoretical accounts of the nature of underlying structure of FLA in integrated speaking performance exist. Therefore, initial exploration as means to learn about and generate hypotheses regarding the internal structure of FLA for integrated speaking was deemed a necessary step to take. The subsequent quantitative analysis was expected to either support or refute the qualitative findings.

The following phases were included in the current study: 1) collecting and analyzing qualitative interview data with an eye on which factors contribute to ESL students' anxious feelings when performing an integrated listening-to-speak task, 2) developing a questionnaire that measures anxiety based on the qualitative findings, 3) administering the designed questionnaire with a much larger sample of participants to validate the instrument, and 4) interpreting to what extent the quantitative results align with the initial qualitative findings. These procedures are illustrated in Figure 1.

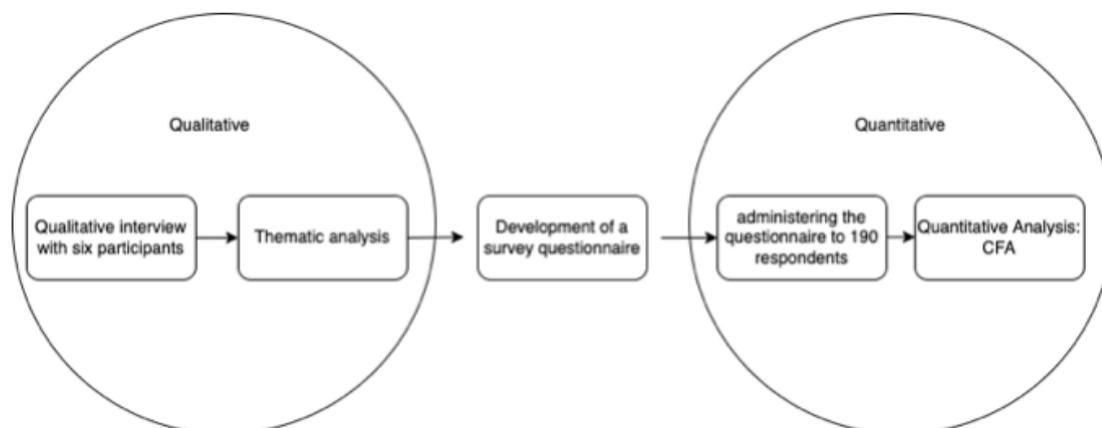


Figure 1 *Mixed-methods research procedures*

Methods

Phase I: Qualitative data collection and analysis

Participants

We limited eligibility criteria for participation to those who are currently enrolled in post-secondary education in a U.S. institute. In the qualitative section of the study, participants were recruited through convenience sampling. They were six nonnative-English-speaking students currently enrolled in a large mid-western university in the U.S. at the time of data collection. They came from three different L1 backgrounds, with Chinese, Persian, and Indonesian equally distributed. Only one participant was male, and all the rest were female. Their fields of study included Business, Education, Finance, and Neuroscience. All of them identified their current proficiency in the English language as high-intermediate. As for educational level, three were pursuing a bachelor's degree, one master's, and two doctoral.

Table 1 *Qualitative interview participants*

Participant number	First language	Major	Educational level
1 (Yuri)	Chinese	Finance	Bachelor's
2 (Ivo)	Chinese	Language education	Master's
3 (Nimo)	Persian	Business	Ph.D.
4 (Arez)	Persian	Neuroscience	Master's
5 (Victor)	Indonesian	Finance	Bachelor's
6 (Susan)	Indonesian	Language education	Ph.D.

Instruments and procedures

Speaking test. The first phase of the qualitative research of the current study was dedicated to having the participants become familiar with independent and integrated speaking tasks. Even though all six participants stated that they had taken the TOEFL test as means to gain admissions to higher education in the U.S., a pre-interview test was administered to refresh their memories and to

help them related to FLA more easily in the subsequent interview. Participants were invited to take a simulated listening-to-speak test and be interviewed subsequently about their feelings and emotions throughout the process. The test was delivered in the format of a lecture, which was adapted from a TED-talk entitled “The History of Tea.” The researchers transcribed the original clip and modified the content for enhanced clarity and intelligibility. The revised script was checked for content validity with two ESL experts, and necessary modifications were made accordingly. To better approximate the real academic context in which such a discourse would likely occur, the researchers invited a native-speaking language instructor to read the script for the participants, rather than playing the video-recorded version of the script. As she read the script, she wrote several keywords on the blackboard, such as people’s names or the years in which certain historical events happened, pausing at a moment when the participants’ cognitive load is likely burdened. After listening to the lecture, participants were asked to (1) verbally summarize the lecture they just listened to, and (2) tell a personal story related to tea. For each question, they were given two minutes to prepare and one minute to speak.

Qualitative interviews. Upon finishing the test, the participants were asked to participate in a 15-minute interview to explain and describe the challenges involved in performing such a task with a specific focus on anxiety. The interviews were semi-structured and intentionally flexible, meaning that, while following the protocols prepared in advance, the researchers used a conversational approach to delve in-depth into the aspects they considered meaningful. The purpose of the interview was to have a better grasp of which factors made participants feel anxious during the speaking test. The whole process was video-recorded.

Qualitative Data Analysis: Thematic Analysis

The interview data were analyzed qualitatively following the protocols of thematic analysis, an approach to analyzing qualitative data to elicit the themes in the data (Clarke & Braun, 2017). Utility of thematic analysis resides in generating useful hypotheses or theories by exploring rich information conferred by data, rather than explaining data in light of existing theory. Justification for using thematic approach in the current study is a need to develop a model that specifies the relationship between FLA in integrated speaking assessment and dimensions that underlie this relationship, which would form the basis of subsequent quantitative testing. We followed inductive coding procedures to “allow for discovery of themes that do not fit a pre-existing coding scheme” (Polio & Friedman, 2017). We initially examined two interview transcriptions by coding line-by-line to obtain a feel for how the transcriptions are organized and a preliminary idea of which prospective themes could be developed out of them. One important consideration that we made was the themes’ relevance to the construct of integrated speaking FLA. Similarly, how the participants relate to FLA as they perform an integrated speaking task was a focus. We analyzed the qualitative data separately applying thematic analysis procedures, and had several meetings to discuss each other’s findings as means to enhance the trustworthiness of the analyses.

Scale development based on qualitative findings

Based on the qualitative findings that FLA for integrated speaking may consist of four different factors – *listening*, *source integration*, *time pressure*, and *topic familiarity* - the researchers crafted an initial list of 26 Likert-type items, with each item measured on a 6-point scale, ranging from *strong disagree* to *strongly agree* (See Appendix A for complete survey questionnaire). Similar to the structure of FLCAS (Horwitz et al., 1986), the questionnaire asks respondents to describe whether they agree or disagree with statements involving factors that might cause FLA when performing integrated listening-to-speak tasks. For example, we developed a questionnaire item – I feel nervous

that I might miss particular points while listening to the script (Q8) – to represent one aspect of the derived listening factor (See Appendix A for complete questionnaire). A few statements were negatively worded to minimize extreme response biases. Before the questionnaire was administered, the researchers asked two experts who are familiar with questionnaire design and understand the selected topic to read through all the items. They evaluated if the questionnaire successfully captured the construct being investigated, and checked for common errors and possible confusions. Necessary revisions were made accordingly. The survey was delivered to a larger population of students who speak English as a second language enrolled in the same institution as for the qualitative interview participants.

Phase II: Quantitative data collection and analysis

Participants

In the quantitative section of the study, 190 participants provided their response to the survey. Hence, a much wider range of L1 backgrounds were represented. Participants' ages ranged from 18 to 45 years. The current proficiency level of English language as self-reported by the participants ranged from elementary to advanced. While a total of eight disciplines were represented in our participant population, the majority came from medicine, public health, education, and business and administration. While the fact that the population has been sampled from a single educational institute in mid-west U.S. might adversely affect the generalizability of the findings, it is expected that such a limitation could be effectively offset by the participants' broad diversity in L1 background, educational level, and field of study.

Quantitative Data Analysis: CFA

We ran a first round of CFA analysis based on the initial inventory of the survey questionnaire. The main purpose of conducting CFA analysis was to ensure that the initial hypothesized model, which was elicited through a qualitative means, functioned adequately in line with model fit indices. Thus, the decision to bypass Exploratory Factor Analysis (EFA) is justified.

The researchers derived the CFA model, and assessed whether the observed data structure matches the hypothesized baseline model via various model fit measures obtained by the model results. In line with Kline's (2016) suggestions, absolute fit indices, such as the chi-square (χ^2), Root Mean Square Error of Approximation (RMSEA), as well as the Standardized Root Mean Square Residual (SRMR) are reported as part of evaluating if the proposed model reflects a good fit to the data. Values that fall below .08 are regarded as acceptable or reasonable as far as RMSEA and SRMR are concerned. The chi-square divided by degrees of freedom (χ^2/df) is examined, with a value indicating a good fit to the data falling between 2 and 3 (Kline, 2016). Also, comparative fit measures including the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI) were estimated to examine the degree to which the proposed model is superior to a rival model. Approximation to 1.0 is associated with good model fit for the comparative fitting measures (Brown, 2006). Then, as the final step, model comparisons between the proposed baseline model and other equally plausible rival models were conducted. This was done to determine a model that best explains the survey response data.

Findings

Qualitative findings

Four major themes related to the cause of FLA when performing an integrated listening-to-speak task were identified through thematic analysis.

Definition of the factors

1. *Listening* refers to the sensations of anxiety or nervousness that participants attribute to the aspects of the listening passage or the cognitive processes that go into understanding it in the test task.
2. *Source integration* refers to anxiety or nervousness that comes from the process of incorporating information provided as part of the source text to organize and convey arguments for the summary task.
3. *Topic familiarity* refers to the degree to which participants feel familiar with a given topic in the test task due to prior experience associated with it.
4. *Time pressure* refers to the psychological stress imposed by the time limit that participants experience during the speaking task. Either the planning time or the speaking time, or both, could be the source of pressure.

Listening. When asked what difficulties they encountered, or what made them feel uncomfortable during the listening-to-speak task, all participants recalled they struggled while listening to the lecture for various reasons. Arez thought it was difficult to catch particular points because remembering everything she heard was impossible. Similarly, Ivo mentioned it was crucial to determine which points were essential amid too much information. However, she also felt this was rather difficult to accomplish.

So sometimes I lost myself because I don't know whether this point is important or not. Sometimes I cannot tell, so I just write everything maybe I think is important, but after the whole essay. I take a lot of detail stuff.

Susan and Victor admitted they did not like listening in the first place, for they were either not good at it or had unpleasant test-taking experiences.

Victor: Probably a little bit uncomfortable because it reminds me to take my English test, like the listening part. So it recalled my memory.

Susan: Because I had to listen. I don't like listening. I don't have any interest.

Nimo expressed difficulty associated with jotting down all information given by the passage. Yuri went so far as to express suspicion of the usefulness of taking notes for listening, to the extent that the logic and flow that the notes should manifest do not readily emerge because she tends to write “too quickly.” Asked to elaborate which coping strategies they have used to counter the negative effects of their manner of note-taking, Nimo suggested that he used “signs for creating hierarchical or graphic notes” that “help you follow from the beginning to the end.” Yuri, on the other hand, shared what she usually does in anticipation of an upcoming lecture in the non-testing context. She “prepare[s] before the lecture,” so she pays relatively more attention to the “whole outline, logic stuff, not the very detailed stuff.”

Source integration. The participants' responses were sharply divided as to which task type, between summary (i.e. integrated task) and sharing personal experience (i.e. independent task), creates more anxiety for them. Arez and Nimo considered the independent task the more challenging of the two, as it requires “mak[ing] up a story” in the absence of background knowledge or personal experience to base the spoken response on. According to Yuri's account, having no background knowledge made it difficult to handle the requirements of the independent task, while the lecture helped her recall the

memory of the history of tea. As Nimo participant mentioned, not having “anything in mind, you should make up something that you can tell in a specific time.”

By contrast, Ivo, Yuri, Victor, and Susan who related to the anxiety-provoking aspects of the integrated task, which explicitly requires proper interpretation and integration of sources for subsequent speaking. Ivo linked the challenges of the summary task to her own state of language proficiency, where her “comprehension is lower than output,” alluding to the realization that integrated speaking performance is not the receptive skill nor the productive skill, but rather an integration of all the language skills involved in a given task. Also, Victor and Susan provided their understanding that performance in the integrated task depends heavily on fulfilling the expectations of the rater. The following are two excerpts that shed light on the participants’ concerns to align their spoken response to the demands of the task in an appropriate manner.

Victor: *I think, personally, it is more difficult to summarize because I think that, if you are supposed to summarize something, you are expected to have some sort of result, maybe, like more fine results. You are supposed to talk about this and this and that or in this order.*

Susan: *Summary was more difficult because you have to get the point first.*

Topic familiarity. Topic familiarity was another theme that demonstrated a starkly different perception among participants. As far as the topic of the lecture—the history of tea—is concerned, the participants coming from Chinese-L1 background found it easy to address. The Chinese participants, Yuri and Ivo, mentioned that they are familiar with the history of tea, as well as the characters mentioned as part of the lecture. At the same time, reasonably enough, they expressed doubt as to whether they would have had the same level of comfort in the test had the topic been one of “European.” On the other hand, non-Chinese background participants were not as familiar with the history of tea as their Chinese counterparts. However, they did not find the topic extremely challenging to both understand and speak about. This is because, as pointed out by Nimo, “tea is very popular, and at least people have heard its name and have drunk it or seeing other people consuming it.”

Time pressure. The participants disclosed varying degrees of anxiety depending on the phase of the speaking test. For instance, they largely took the preparation time to be sufficient, which, in turn, affected their general perception of the test. To be specific, Arez participant mentioned that the test was only moderately difficult due to a host of factors, which includes relatively easy topic, clearly intelligible accent of the lecturer, and ample preparation time. However, the timing factor, as it relates to the speaking part of the test, was found to have a debilitating effect on anxiety of the participants. For the summary task, Yuri participant stated that “time was not enough to give all the data,” thus “I can tell you more if I have more time.” Victor brought up the difficulty of managing time, saying “I did not realize that one minute was too quick.” Another interesting point has been made by Nimo in whose perception the time factor in the independent task context goes hand in hand with topic familiarity when he mentioned:

In a test, because the time is one minute, I cannot just answer for 20 seconds and stop. If it was like that, because sometimes you don't have something to say for one minute. Maybe I don't have any experience with tea, but I need to just make some sentences to fill that one minute. So that is something that makes it stressful.

Similar to the role of timing in posing as a source of anxiety in completing an independent task, Nimo highlighted the challenge inherent in covering a nice summary within one minute. Also, the limited amount of time allowed for the summary task, coupled with a desire to cover as much detail

as possible, posed a double bind for participants. This is exemplified when Yuri mentioned: “For the summary part, maybe I try to cover a lot detail stuff, so I didn’t finish my end part very well. So I finish it maybe very urgent. You can feel that.”

Quantitative findings

Model comparisons

In line with the qualitative interview findings, the researchers established the four-factor correlated model as the baseline. This baseline model, referred to hereafter as Model 1, consists of *listening*, *integrated speaking*, *topic familiarity*, and *time pressure*. Before comparing potential models, the researchers calculated the reliability and validity of the developed scale. When estimated at a factor level, the coefficient omegas of all the factors within the developed scale except for *topic familiarity* exceeded the benchmark value of .75 (*listening* $\omega = .863$; *topic familiarity* $\omega = .662$; *source integration* $\omega = .822$; *time pressure* $\omega = .836$). As for validity, convergent validity was established by virtue of Average Variance Extracted (AVE) going above .5, once again except for *topic familiarity*. We attribute the relatively low reliability and validity of this factor to a high percentage of items that were removed from the original inventory.

Model fit indices showed Model 1 to fit the survey questionnaire data reasonably well ($\chi^2/df = 2.14$, TLI = .933, CFI = .928, RMSEA = .079, SRMR = .048). First, the chi-square divided by degrees of freedom is 2.14, which falls within an acceptable range (Kline, 2016). Absolute values of model fit (e.g., TLI and CFI), exceeded an acceptable limit of .92, indicating that the explanatory power of the model in relation to the data is reasonably strong. The amount of either remaining or unexplained variance was reasonable, as shown by both the SRMR and RMSEA going under .08.

This study went further than calculating model fit indices for the baseline model to comparing them with those of rival models with a view to establishing plausibility of the baseline model over their competing models. The first competing model of the baseline model, Model 2 represents the possibility that FLA for integrated speaking assessment is a unidimensional construct. The implication of this model is that learners who are engaged in an integrated listening-to-speak test task would be simultaneously vulnerable to the effects of both cognitive skills—*listening* and *integrated speaking*—and externally imposed test conditions—*time pressure* and *topic familiarity*. The plausibility of this hypothesis was supported by the observation that, while both the CFI and TLI of Model 2 were above the acceptable level of .92, the RMSEA and the SRMR went below the threshold of .08 (see Table 2).

Another plausible model that the current study hypothesized is the one in which *listening* and *source integration* are integrated as one factor. This model, in other words, does not differentiate the receptive skill (i.e., listening) and the productive skill (i.e., source integration) that go into integrated speaking performance. Comparing the baseline model against the hypothesized model where no differentiation is made between the receptive and productive skills was expected to provide insights into the nature of the integrated assessment. Termed Model 3 in the current study, a three-factor model was seen to take hold in terms of model fit indices (TLI = .923, CFI = .937, RMSEA = .079, SRMR = .043).

Seeing that the baseline model as well as all the hypothesized competing models functioned well in terms of model fit indices, the researchers carried out a series of chi-square difference tests to discern the model with the best fit to the data. The result was that all the models were statistically different from one another in terms of explaining the data, with the baseline four-factor model functioning

significantly better than the other two rival models (Model 1 vs. Model 2, $p < .005$; Model 1 vs. Model 3, $p < .005$; Model 2 vs. Model 3, $p < .005$). Taken together, the baseline four-factor model lends quantitative support to the qualitative findings that suggest the existence of four constructs under FLA for integrated listening-to-speak tasks.

Table 2 Model comparisons

	TLI	CFI	RMSEA	SRMR	Chi-square difference
Model 1 (baseline 4-factor model)	.928	.933	.079	.048	
Model 2 (unidimensional model)	.940	.951	.076	.040	Model 1 vs. Models 2 and 3 $p < .005$
Model 3 (3-factor model)	.927	.941	.073	.049	Model 2 vs. Model 3 $p < .005$

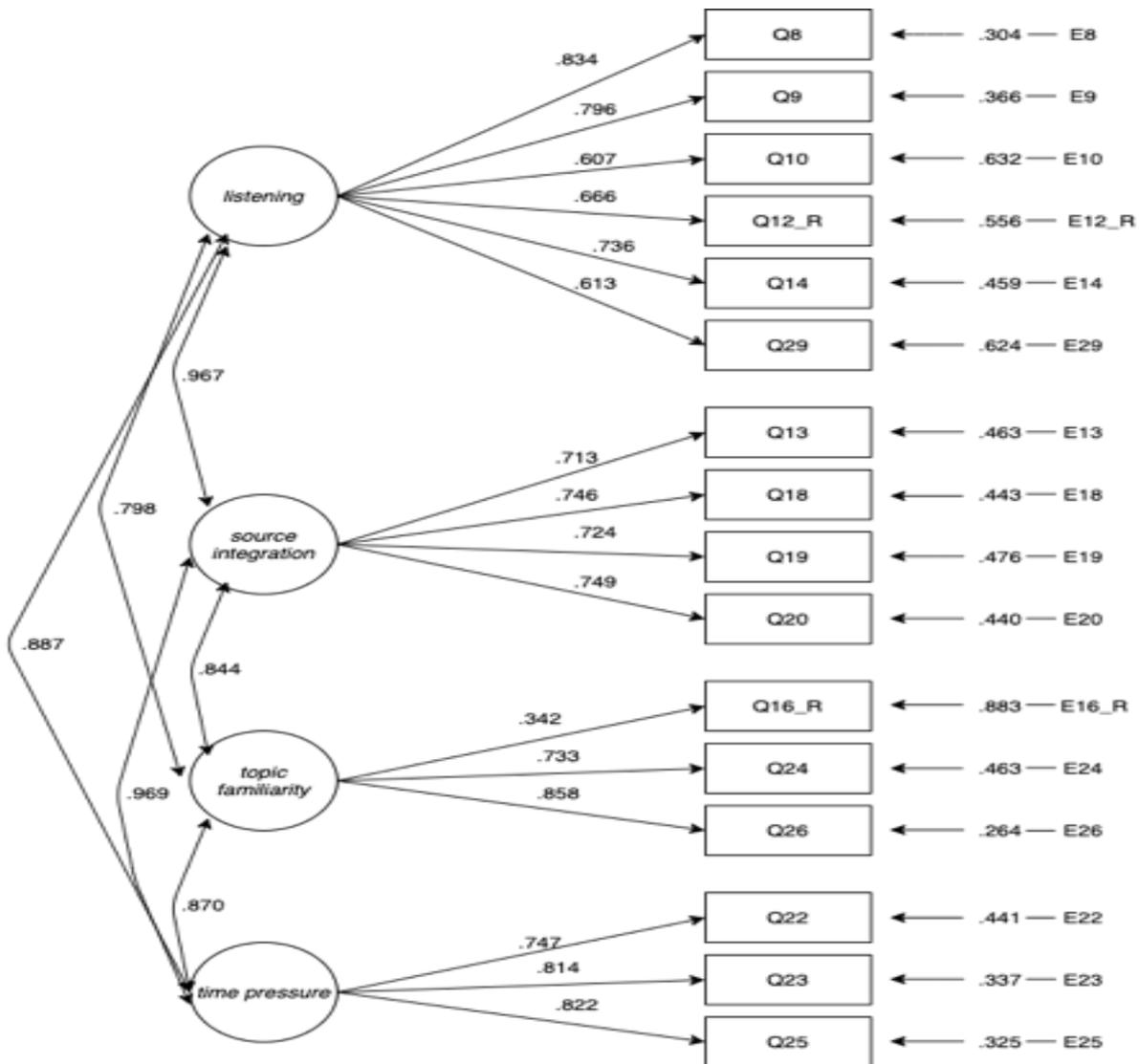


Figure 2 Factor structure of FLA of integrated listening-to-speak task

Note: All the paths are statistically significant at $\chi^2 = .05$ level.
 “_R” as in E12_R and E16_R means “reverse-coded.”

Discussion

The current study has confirmed that FLA for integrated speaking tasks is a multidimensional construct. To be specific, underpinning the four-factor model of this specific type of FLA were *listening*, *integrated speaking*, *topic familiarity*, and *time pressure*. The interfactor correlations were fairly high among factors, with some going above the threshold value of .85, as suggested by Kline (2016). Utilization of two-phase mixed methods research was crucial in the conduct of this research study. One thing to note is that the three-factor model, which does not differentiate *listening* and *integrated speaking*, is also statistically supported. Nevertheless, the researchers opted to maintain the four-factor model as the baseline in recognition of the qualitative findings that support its plausibility. The ensuing section will address each of the identified factors in detail.

Listening and source integration

Previous research has been divided as to the role of the receptive skill in the performance of the integrated task (Delaney-Asencion, 2008; Plakans, 2009; Plakans & Gebril, 2013; Watanabe, 2001). In the current study, the *listening* factor speaks to anxiety aroused by the receptive aspect of integrated speaking performance, while the *source integration* factor aligns more closely to anxiety induced by the productive skill. One notable finding of our research study is that both the baseline four-factor model, which separates *listening* and *source integration*, and a three-factor model, which collapses the two factors, are both statistically plausible. That is, whether FLA relates to the *listening* and *integrated speaking* factors separately, as the baseline model suggests, or it straddles the two phases as one, like the three-factor model indicates, does not make a significant difference from a purely quantitative standpoint. On one hand, the finding that *listening* and *source integration* can be integrated as one factor partly confirms prior research studies that failed to establish the psychometric separability of the receptive and productive skills in the performance of the integrated tasks. (Delaney-Asencion, 2008, Watanabe, 2001). The current study reported a correlation coefficient of .967 between *listening* and *source integration*, which goes above the threshold value for determining collinearity. On the other hand, the finding that *listening* and *source integration* are independent factors finds support from the model fit indices that pertained to the baseline four-factor model, which confirms prior research indicating the importance and independence of the receptive skill in integrated performance (Plakans, 2009; Plakans & Gebril, 2013). Also, the participants in the qualitative interviews provided accounts that speak to the separability of *listening* and *source integration*, as exemplified by the disagreement among them as to which task between integrated and independent tasks was more challenging for them. Although this study was conducted to identify the nature and structure of FLA in relation to integrated speaking assessment, the findings do not seem to finalize the debate surrounding the role and effect of the receptive skill in integrated assessment.

Time pressure

One important finding that relates to the *time pressure* factor is how it emerged as an important factor in the configuration of anxiety for integrated tasks. The qualitative data suggested that while participants largely took the preparation time to be sufficient, the time allotted for speaking was not enough to cover as many details as they would have liked. With this qualitative finding informing the initial model for CFA, the quantitative analysis found that the estimated correlations of *time pressure* with the *listening* and *source integration* factors are .887 and .969, respectively, with $p < .005$ for both. Thus, through the use of CFA, the current study shows the correlational relationship that *time pressure* has with comprehension and performance factors, *listening* and *source integration*. Many previous studies on anxiety reported its inverse relationships with L2 integrated speaking

performance score (e.g., Huang, 2018; Huang & Hung, 2013), and the current study builds on them by making explicit the role and effect of *time pressure*, an externally imposed variable, on shaping the comprehension and production of integrated speaking tasks, *listening* and *source integration*. However, no claims can be made about the directionality between *time pressure* and *listening* and *source integration* (Brown, 2006); it is unclear, from a quantitative perspective, whether anxiety coming from *time pressure* is caused by anxiety grounded in *listening* or vice versa.

Even though the high correlation between *time pressure* and *listening* and *source integration*, as they go over and above the benchmark value of .85 (Kline, 2016), could be a reason to incorporate different factors into a single factor, this study did not combine the *time* factor with other factors with which it highly correlates. Since test-takers might feel anxious by the perceived lack of time in relation to task completion, it is possible that *time pressure* applies to any or all stages in the integrated speaking task, thus correlating highly with other constructs which rely heavily on the passage of time. The findings provide support to the view that the anxiety felt in a language test is situation-specific (Spielberger, 1983), as imposing strict time limits is one of the most typical characteristics of any or all speaking tests. Hence, it is likely that the respondents of the survey questionnaire in the current study related with the timing demands of integrated speaking tasks, which may increase their level of anxiety. Also, the extent to which *time pressure* affects integrated speaking performance via *listening* and *source integration* is not clear from the findings of the current study, since it is not under the purview of CFA, but requires the use of Structural Equation Modeling (SEM). Noting the adverse effects of timing in language tests, Immerman (1980) claimed that removing the time limit from language tests would significantly reduce test anxiety of test-takers. However, considering practicality concerns that go into administering language tests, this recommendation might not be a viable option for test developers.

Topic familiarity

Another important finding of the current study is that topic familiarity poses as an anxiety-provoking element in the completion of an integrated speaking task for L2 test-takers. This finding accords with Huang *et al.* (2016), who found that the impact of the topic on the level of anxiety that test-takers experience in an integrated speaking task varies depending on the topic. Furthermore, the qualitative interview analysis revealed that the level of anxiety felt by each interview participant about the topic—the history of tea—differed along L1 background lines. Thus, some qualitative interview participants who were familiar with the topic felt relatively comfortable speaking for the task, while those with a relatively low level of familiarity to the topic expressed anxiety and nervousness. This finding is in line with Qiu and Lo (2017), who suggested that test-takers feel enthusiastic and confident and show more interest toward familiar topics, but nervousness and anxiety are aroused for those test-takers who perform on unfamiliar topics. In a similar study, Phung (2017) concluded that topic familiarity is one of the determining factors in L2 learners' task preference. Specifically, interview data indicated familiar topics are viewed favorably by participants while unfamiliarity resulted in their negative affective reactions such as disliking the task or feeling confused.

Considering that one of the biggest claims that researchers and practitioners advocating for increasing use of integrated test tasks is with regard to equalizing test-takers' differential topical knowledge, the fact that the participants remained on different footing in terms of topic familiarity in this study requires meticulous attention. As noted by Huang *et al.* (2018), the provision of input materials in an integrated speaking test might not effectively offset the effects of the initial level of topical knowledge that test-takers might bring to the test. Going one step further, it should be noted that differential topic familiarity made some participants feel anxious about not understanding enough of the listening passage for the subsequent stage of integrated speaking. Thus, test-takers

whose topic familiarity is comparatively low for an integrated speaking test task might be placed in a double bind of having to deal with FLA, another source of construct-irrelevant variance, to minimize the detrimental effects on their test performance.

Limitations of the Study

Even though the study was designed to minimize the undue effects of extraneous factors, inferences drawn from the results are constrained as the study targeted a limited number of ESL students from one university and focused solely on one type of integrated assessment. Enhanced generalizability could come about with replications involving a larger sample size and other integrated assessment task types. Another limitation exists with respect to the way in which the quantitative data were collected: self-report. As FLA is situation-specific (Spielberger, 1983), self-reported measures of anxiety might differ from the actual level of anxiety aroused by the testing context to a great extent. Thus, it might be worthwhile to probe into the students' FLA during the test utilizing proper research methods and instruments. Also, future research should keep promoting our understanding of how FLA interacts with various forms of integrated tasks.

Practical Implications and Conclusion

The present study investigated the factors underlying Foreign Language Anxiety of ESL students engaged in an integrated listening-to-speak task through the development and validation of a questionnaire tailored for such a context. To this end, we combined research methods, initiating from qualitative exploration to quantitative confirmation. To be specific, four factors were identified—*listening*, *source integration*, *topic familiarity*, and *time pressure*—through thematic analysis of interviews. Subsequently, the qualitative findings were subjected to Confirmatory Factor Analysis to see whether a statistical process holds them up. As with related studies in the literature, FLA for integrated speaking tasks is a multi-dimensional construct supported by multiple underlying sub-constructs.

Even though this study focused on FLA in the context of integrated speaking assessment, the findings might provide important suggestions for practice. Among all the factors identified as part of the CFA-based analysis of integrated speaking FLA, teachers in the classroom have the authority to directly control *topic familiarity* and *time pressure*. Since they know the topics with which the learners are familiar, teachers could design an integrated speaking test assessing for the topic that the learners could talk about with relative ease and less anxiety. If an unfamiliar topic is presented as part of an integrated speaking test, however, the teacher could allow learners to repeat the task several times to reduce both cognitive and affective burdens. Also, by imposing flexible time limits, teachers could help learners to perform better than when they are under tight and strict time limits, as would be the case with high-stakes language tests. Providing these accommodations might compromise practicality to a certain extent. However, a tradeoff is that teachers could gather more accurate validity evidence as to a learner's level of integrated speaking performance.

One important implication that applies to both test developers and teachers in the classroom is to make their expectations as to source use as explicit as possible. As one of our qualitative participants suggested, test-takers are oftentimes left wondering how best to utilize the stimulus sources in performing for integrated language assessment. For instance, a rubric demonstrating specific requirements for source use could be communicated with test-takers/learners. This way, learners might feel less anxiety stemming from lack of understanding about what they are expected to do as far as source use is concerned.

References

- Asencion-Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, 7, 140–150. <http://dx.doi.org/10.1016/j.jeap.2008.04.001>
- Backman, N. (1976). Two measures of affective factors as they relate to progress in adult second-language learning. *Working Papers in Bilingualism*, 10, 100–122.
- Barkaoui, K., Brooks, L. A., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, 34(3), 304–324. <https://doi.org/10.1093/applin/ams046>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Chastain, K. (1975). Affective and ability factors in second language acquisition. *Language Learning*, 25, 153–161. <https://doi.org/10.1111/j.1467-1770.1975.tb00115.x>
- Cheng, Y., Horwitz, E. K., & Schiller, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, 49(3), 417–446. <http://dx.doi.org/10.1111/0023-8333.00095>
- Chou, M. (2018). Speaking anxiety and strategy use for learning English as a Foreign Language in full and partial English-medium instruction contexts. *TESOL Quarterly*, 52(3), 611–633. <http://doi.org/10.1002/tesq.455>
- Clarke, V., & Braun, V. (2017) Thematic analysis, *The Journal of Positive Psychology*, 12(3), 297–298. <http://doi.org/10.1080/17439760.2016.1262613>
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *The Modern Language Journal*, 89(2), 206–220. <https://doi.org/10.1111/j.1540-4781.2005.00275.x>
- Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *The Modern Language Journal*, 96(2), 170–189. <https://doi.org/10.1111/j.1540-4781.2011.01174.x>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. A. (1986). Foreign language classroom anxiety. *Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Huang, H. D., & Hung, S. A. (2013). Comparing the effects of test anxiety on independent and integrated speaking test performance. *TESOL Quarterly*, 47(2), 244–269. <https://doi.org/10.1002/tesq.69>
- Huang, H. D., & Hung, S. A. (2018). Investigating the strategic behaviors in integrated speaking assessment. *System*, 78, 201–212. <https://doi.org/10.1016/j.system.2018.09.007>
- Huang, H. D., Hung, S. A., & Hong, H. V. (2016). Test-taker characteristics and integrated speaking test performance: A path-analytic study. *Language Assessment Quarterly*, 13(4), 283–301. <https://doi.org/10.1080/15434303.2016.1236111>
- Huang, H. D., Hung, S. A., & Plakans, L. (2018). Topical knowledge in L2 speaking assessment: Comparing independent and integrated speaking test tasks. *Language Testing*, 35(1), 27–49. <https://doi.org/10.1177/0265532216677106>
- Immerman, M. A. (1980). *The effects of eliminating time restraints on a standardized test with American Indian adults* (ED196584). ERIC.
- Kitano, K. (2001). Anxiety in the college Japanese language classroom. *Modern Language Journal*, 85, 549–566. <https://doi.org/10.1111/0026-7902.00125>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

- Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *Modern Language Journal*, 92(1), 71–86. <https://doi.org/10.1111/j.1540-4781.2008.00687.x>
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32(1), 21–36. <https://doi.org/10.1016/j.system.2003.08.002>
- Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (2000). Cognitive, affective, personality, and demographic predictors of foreign-language achievement. *The Journal of Educational Research*, 94(1), 3–15. <https://doi.org/10.1080/00220670009598738>
- Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, 76, 14–26. <https://doi.org/10.2307/329894>
- Phung, L. (2017). Task preference, affective, response, and engagement in L2 use in a US university context. *Language Teaching Research*, 21(6), 751–766. <https://doi.org/10.1177/1362168816683561>
- Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. *Journal of English for Academic Purposes*, 8(4), 252–266. <https://doi.org/10.1016/j.jeap.2009.05.001>
- Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18–34. <https://doi.org/10.1016/j.asw.2011.09.002>
- Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source test use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217–230. <https://doi.org/10.1016/j.jslw.2013.02.003>
- Polio, C., & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research*. New York: Routledge.
- Qiu, X., & Lo, Y. Y. (2017). Content familiarity, task repetition and Chinese EFL learners' engagement in second language use. *Language Teaching Research*, 21(6), 681–698. <https://doi.org/10.1177/1362168816684368>
- Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, 28, 129–142. <https://doi.org/10.1111/j.1467-1770.1978.tb00309.x>
- Thompson, A. S., & Lee, J. (2014). The impact of experience abroad and language proficiency on language learning anxiety. *TESOL Quarterly*, 48(2), 252–274. <https://doi.org/10.1002/tesq.125>
- Watanabe, Y. (2001). Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions. (Unpublished doctoral dissertation). University of Hawaii.
- Yang, H., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly*, 46(1), 80–103. <https://doi.org/10.1002/tesq.6>

Appendix

Inventory of survey questionnaire items

<p>Theme 1: Listening</p> <ul style="list-style-type: none"> · If I were to take a listening-to-speak test, I would worry a great deal before taking it. (Q7) · I feel nervous that I might miss particular points while listening to the script. (Q8) · I feel uncomfortable taking notes for listening in an English test. (Q9) · When the listening passage presents too much information, it is stressful to determine which is important. (Q14) · When listening to an English passage, it is hard for me to stay focused without visual materials. (Q29)
<p>Theme 2: Source Integration</p> <ul style="list-style-type: none"> · Covering detailed information for a summary task is stressful for me. (Q13) · When speaking, I get worried if I cannot cover important points delivered by a listening passage. (Q18) · I feel uncomfortable if I have to make up ideas to talk about for a speaking test. (Q19) · I feel nervous if I have to speak based on information provided by a listening passage. (Q20) · I find myself forgetting what to mention in a speaking test because of feeling anxious. (Q25)
<p>Theme 3: Topic familiarity</p> <ul style="list-style-type: none"> · Having to talk about a topic new to me makes me feel stressed. (Q15) · The topic does not affect my feelings when taking a test. (Q16_R) · I feel relaxed if given a topic I am familiar with. (Q17) · Having background knowledge about a topic in a speaking test boosts my confidence. (Q21) · For a speaking test, I need more time to prepare for a new topic. (Q26)
<p>Theme 4: Time Pressure</p> <ul style="list-style-type: none"> · I get stressed seeing how much time is left for a speaking test. (Q9) · When taking a speaking test, I worry about not being able to finish on time. (Q22) · For a speaking test, I feel that I could have done a lot better than I actually did if I had more time to prepare. (Q24) · For a speaking test, I feel that I could have done a lot better than I actually did if I had more time to speak. (Q32)
<p>Others</p> <ul style="list-style-type: none"> · The presence of other people in the test makes me nervous. (Q11) · If I knew I was going to take a listening-to-speak test, I would feel confident and relaxed. (Q12_R) · My stress level goes up when taking a speaking test. (Q23) · I worry about consequences of a low score in the speaking test. (Q27) · I worry about making grammar mistakes while taking a speaking test. (Q28) · I feel confident if I am well-prepared for speaking test. (Q30) · I would feel more confident as I take more and more speaking tests. (Q31)