

## Integrated Listening/Speaking Skill Assessment: The Role of Ambiguity Tolerance, Cognitive/Metacognitive Strategy Use, and Foreign Language Anxiety

Karim Sadeghi<sup>1\*</sup>, Neda Bakhshi<sup>2</sup>

### ARTICLE INFO

#### Article History:

Received: April 2024

Accepted: June 2024

#### KEYWORDS

Ambiguity Tolerance  
Cognitive/Metacognitive  
Strategy Use  
Foreign Language  
Anxiety  
Integrated  
Listening/Speaking  
Assessment

### ABSTRACT

Assessing language skills in an integrative form has drawn the attention of assessment experts in recent years. While some research data exists on integrative listening/reading-to-write assessment, there is comparatively little research literature on listening-to-speak integrated assessment. Also, little attention has been devoted to the role of individual attributes within the context of integrated assessment. The objective of the current research was to investigate the relationship between integrated listening/speaking assessment and individual characteristics of ambiguity tolerance (AT), use of cognitive/metacognitive strategies, and foreign language anxiety (FLA). Oxford Quick Placement Test was used to homogenize 60 EFL learners in terms of language proficiency (B2-C1). Additionally, integrated listening/speaking performances were collected using sample TOEFL-iBT tests. The transcribed spoken samples were evaluated by two raters using TOEFL-iBT rubrics in terms of overall description, delivery, language use, and topic development. Additionally, information on individual characteristics was gathered by means of 3 different questionnaires. Data analysis revealed that FLA had a negative relationship while AT and the use of cognitive and metacognitive strategies had a positive correlation with integrated listening/speaking test performance. Individual differences have generally been neglected in the assessment literature, but this study revealed that performance on integrated listening/speaking tests can be affected by language-irrelevant constructs such as individual attributes in addition to test-takers' language competence.

### 1. Introduction

Speaking is a substantial language skill, crucial for daily communication, international mobility, employment, and higher education in the present era (Fulcher, 2015; Isaacs, 2016). According to Isaacs (2016), speaking is the most challenging of the four language skills to master when learning second and foreign languages. Despite its remarkable place in language pedagogy, speaking has been noticeably challenging for assessors since the ability to speak authentically is to a large extent dependent on the success of listening. As such, rather than considering speaking as an isolated skill or an independent construct, language educators and assessors are advised to regard it as an integrated attribute (Crossley & Kim, 2019).

Integrated language assessment tasks have garnered support owing to numerous benefits. For instance, integrated tests have been claimed to offer authenticity given that they are generally designed to simulate authentic language use (Butler et al., 2000); integrative tests also provide high predictive

<sup>1</sup> Dhofar University, Oman, Email: [ksadeghi@du.edu.om](mailto:ksadeghi@du.edu.om)

<sup>2</sup> Urmia University, Iran, Email: [nedaibakhshii97@gmail.com](mailto:nedaibakhshii97@gmail.com)

validity, induce more positive washback, boost learners' motivation, and offer context-specific diagnostic feedback (Wesche, 1987). Integrated language test tasks also supply test-takers with prior textual or aural input for test-takers to generate spoken or written responses (Brown et al., 2005). The TOEFL iBT, for example, evaluates test takers' spoken ability through exercises that combine speaking, reading, listening, and writing. However, despite the integrated nature of speaking and recommendations to teach and test it as such, due attention has not been devoted to assessing speaking in the form of an integrated construct. Accordingly, one aim of the current project was to advocate the speaking assessment as an integrated construct composed of listening and speaking.

When it comes to assessing any language skill, apart from language competence, learners' individual characteristics are also believed to have an impact on test results. As Bachman and Palmer (2010) recognize, learners' individual characteristics are not basically considered as a component of language skill although they can have a crucial effect on learners' test performance. In a similar vein, Kunnan (1998) highlights that individual features might have a fundamental influence on learners' performance, sometimes in ways thoroughly irrelevant to the specific skill being assessed. Consequently, it is crucial for test developers and users to pay significant attention to the essence and extent of impacts that individual characteristics can exert on second language (L2) test performance. Such an approach to the role of individual attributes helps minimize the effect of non-language related features on test results, ensuring that the test scores more accurately reflect language capability (Kunnan, 1995).

Given the intervening role of individual characteristics in language assessment, another aim of the present study was to evaluate the standing of some major individual attributes in language learners' performance on integrated listening/speaking test tasks. The individual characteristics which were focused on in this work were: ambiguity tolerance (AT), use of cognitive/metacognitive strategies, and foreign language anxiety (FLA). This does not mean that other individual characteristics are any less important or may not have a link with performance on integrated listening/speaking tasks. The reason why these three were selected is that there is abundant evidence (see next section) on the role of these features in language learning as well as their place in skill-based language assessment. However, there appears to be no research studies investigating the link between these attributes and integrated language tests in general, and integrated listening/speaking assessment in particular.

The motivation to pursue a link between these variables and integrated listening/speaking assessment was the following assumptions: learners with higher degrees of AT, greater use of cognitive/metacognitive strategies, and lower levels of FLA may perform better in integrated listening/speaking tests. Investigating the relationship between non-language related constructs (like the individual characteristics mentioned above) and performance on language tests is significant because the observed test scores can reflect test-takers' language competence (in this case their integrated listening/speaking ability) as well as other irrelevant variables, the share and contribution of which should be properly established so as any claims about score validity can be substantiated.

## **2. Review of Literature**

### **2.1. Integrative Assessment**

Recently, integrated skills assessment has attracted some research attention among language testing and assessment experts (Crossley & Kim, 2019; Frost et al., 2021; Plakans & Gebril, 2012; Weigle & Parker, 2012; Zhang & Wilson, 2023). The field of language assessment has move toward more authentic testing methods, and integrated language assessment has emerged as a key area of interest (Yu, 2013). Mapping the Speaking Module of IELTS against authenticity criteria in applied linguistics (Bachman and Palmer's (2010) Test Usefulness Model) and general education (Herrington & Herrington, 2006), Souzandefar (2024), for instance, found that the Speaking test of IELTS suffers from authenticity and recommended integrated assessment among others to bring more authenticity to speaking assessment. Integrated test tasks are those that require test takers to listen to and/or read source texts, and then incorporate information from these texts into spoken or written responses, and such tests are increasingly used to simulate real-world language use (Lewkowicz, 1997). Language abilities were mainly assessed in isolation in the past; however, throughout the decades, language tests have evolved to serve multiple purposes, acknowledging the interconnected nature of language skills (Huang & Hung, 2010; Huang et al., 2016).

Although since decades ago the integration of language skills has been recognized in assessment terms, leading to dichotomies such as discrete-point versus integrative testing, only recently has the distinction between integrative tests (such as cloze and dictation) and integrated assessment (such as listening/ reading-then-speaking or reading/listening-then-writing) received some research attention. Frost et al. (2020), for instance, observe that due to the real-life nature of integrated assessment tasks, these are increasingly used in high-stakes tests such as TOFBL iBT, since such assessment practices yield more appropriate and accurate evidence of test-takers' language proficiency than independent tests do. The same is echoed by Rukthong and Brunfaut (2020) who posit that despite very limited research evidence, and although it is still not clear what such integrated assessments measure, their use is growing in second language assessment circles. Similarly, Crossley et al. (2014) note that interest in using integrated listening/speaking assessment is on the rise despite the lack of enough empirical evidence on the role and nature of stimulus material and text integration in candidates' performance and in human ratings of their performance.

Studies on integrated assessment have primarily been conducted on the assessment of writing rather than speaking skill (Cubilo & Winke, 2013; Cumming et al., 2005; Doubet & Southall, 2017; Plakans & Gebril, 2012; Weigle & Parker, 2012). Few studies have investigated speaking skill in an integrated format. One early example is Lee (2006), who examined the reliability of scores in the TOEFL-iBT speaking test, including independent and integrated tasks, and found that integrated listening/speaking tasks have higher reliability. In a similar vein, Frost et al. (2012) used discourse analytic techniques to investigate the validity of integrated listening-speaking tasks in a pilot Oxford English language test. The researchers aimed to find out how stimulus from a listening activity is integrated into speaking performance in a summary task and whether the scores reflected what was involved in the performance. Based on their analyses, the researchers recommended integrated listening-then-speaking test tasks as appropriate measures of speaking proficiency. Rukthong (2021) likewise investigated differential performance of Thai students on multiple-choice (MC) listening test items and compared that with their performance on an integrated listening/speaking test (what she calls listening-to-summarize (LTS) tasks). The findings indicated that LTS tasks engaged advanced listeners and led them to employ some of the listening strategies used in real-life contexts, leading the researcher to recommend the use of such integrated tasks for assessing language abilities.

More closely linked with the theme of our work is a study by Frost et al. (2021). Concerned that comprehension of listening/reading input can raise questions about the integrated speaking score as well as aiming to reveal the impact of the source material on strategy use by test takers, Frost et al. (2021) analyzed speaking performances of 120 TOEFL iBT test-takers as well as 38 verbal reports from the same candidates. The results indicated differential use of test-taking strategies by more advanced test-takers, implying that there is a link between strategy use and performance on integrated speaking assessment. Another relevant study is Barkaoui et al. (2013), who studied the role of strategic behavior in integrated and independent speaking performance. They investigated the relationship between strategy use and integrated listening, reading, and speaking performance in TOEFL-iBT by 30 Chinese EFL learners and found no significant relationship between strategy use and integrated versus independent speaking performance. These two studies are a precursor to our study where, in addition to strategy use, AT and FLA were also considered as predictors of performance on integrated listening/speaking tasks. The sparse studies that have so far investigated integrated listening/speaking assessment have fully neglected the significance of individual characteristics, like ambiguity tolerance, and foreign language anxiety as well as the use of metacognitive strategy use in their inquiry.

## **2.2. Individual Attributes**

**2.2.1. Cognitive/Metacognitive Strategies.** Individual attributes such as the use of cognitive/metacognitive strategies have primarily been investigated in the context of language learning. Metacognitive strategies are test-takers' conscious mental processes for guiding and managing their cognitive strategy processing for having good performance (Schraw, 1998). On the other hand, cognitive strategies are problem-solving techniques that learners or test takers use to handle the learning/testing tasks and facilitate the acquisition of knowledge or skill (O'Malley & Chamot, 1989). Numerous studies have for example investigated the relationship between learning strategies (as an example of cognitive strategies) and success in learning an L2 (Goh, 1998; Green & Oxford, 1995;



Zhang & Seepho, 2013). Comparatively fewer studies have delved into the role of individual variables in language assessment; and despite relatively substantial attention paid to the place of individual characteristics in learning a second language, only scant literature exists on the relationship between cognitive/metacognitive strategy use and language assessment, and less so in the context of integrated listening/speaking assessment. Phakiti (2006), for instance, investigated the link between strategy use and performance in reading tests and found a significant positive relationship between cognitive and metacognitive strategy use and reading test performance of 384 EFL learners. This study signifies that performance on an isolated reading test depends not only on the test taker's reading competence and language proficiency but also on his/her non-language related capabilities such as how he/she can use certain cognitive/metacognitive strategies. Similarly, Purpura (1998) found that high and low-proficiency language learners use cognitive and metacognitive strategies differently in second language reading comprehension tests. Swain et al. (2009) and Barkaoui et al. (2013) compared the strategic behaviors between two types of TOEFL-iBT tasks and reported that the connection between strategy use and test performance might differ due to various factors such as learners' characteristics, task type, and context. The only study to attend to the link between strategy use and integrative speaking is Crossley and Kim (2019). They compared text integration (e.g., lexical and syntactic information, and cohesion) and strategy use (e.g., note taking) in performance on integrated listening/speaking tests and found that the influence of textual features was more dominant than strategy use in learners' performance. This result implies that there are other factors involved in test performance whose share should not be neglected in interpreting test scores, one of which is strategy use.

**2.2.2. Ambiguity Tolerance.** In addition to strategy use, AT and FLA are important individual attributes that may influence performance on integrated assessment. AT is defined by Budner (1962, p. 29) as "the tendency to perceive ambiguous situations as desirable". As far as AT is concerned, there are studies which have examined the role of AT in non-integrated language tests. Kamran and Maftoon (2015), for example, found that learners with a high AT achieved higher scores in reading comprehension tests. Similarly, Kamran (2011) investigated the difference between Iranian intermediate male and female learners with different levels of AT in discrete point reading and writing tests. The results indicated that the participants with an average AT level scored the highest in reading and the lowest in writing; also there were no significant differences between levels of AT across gender. These studies reflect the role a test-takers' AT level can play in their test performance, the nature and the exact contribution of which require recognition and further research.

**2.2.3. Foreign Language Anxiety.** The final individual attribute chosen to be studied in this project was FLA. Horwitz et al. (1986) define FLA as "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). Several studies have examined the association between FLA and oral assessment, and most studies have uncovered a moderate negative correlation between them (Hewitt & Stephenson, 2012; Oya et al., 2004). The relationship between FLA and speaking is expected to be reverse but it is not clear whether the same degree and type of relationship will be valid when another component is added to the speaking assessment, that is listening. Indeed, very few studies have explored the relationship between FLA and integrated speaking/listening assessment. One such study is Lee and Ye (2021) who conducted a mixed methods study to identify the structure of FLA in integrated listening/speaking tasks with 190 elementary-advanced EFL learners. Based on a qualitative analysis, they identified the following factors that cause FLA to interact differently with assessment: source integration, topic familiarity, and time pressure. This study was meant to identify the sources of FLA, rather than whether different levels of FLA could have a differential relationship with the integrated task performance, which this study seeks to establish. Huang and Hung (2010) similarly studied anxiety reactions of test takers in both independent and integrated tasks. The findings showed that learners preferred integrated over independent tasks, even when reading-to-speak tasks provoked additional anxiety. Likewise, although this study examined FLA in integrated speaking/listening tasks, it did not consider the relationship between these variables and the contribution of FLA in integrated task performance, a gap this study aims to bridge.

All in all, despite the importance of the ambiguity tolerance (AT), cognitive/metacognitive strategy use, and foreign language anxiety (FLA) in the process of language learning and performance, little attention has been paid to the involvement of these individual features in the assessment process, a gap our study was meant to bridge. The limited evidence coming from independent assessment (Ariyanti, 2016; Fitriani & Apriliawati, 2015; Resnik et al., 2023; Zhang & Seepho, 2013) implies that individual characteristics might have an influence on learners' performance in integrated assessment too. However, given that these constructs have rarely been examined in the context of integrated speaking assessment, the relevant body of knowledge needs to be built as a result of studies similar to ours. Accordingly, the present project set out to shed light on the role of AT, cognitive/metacognitive strategy use, and FLA in the assessment of integrated listening/speaking. More specifically, the following questions were posed for further scrutiny.

- 1: Is there any significant relationship between Iranian EFL learners' ambiguity tolerance (AT, high vs. low) and integrated listening/speaking performance?
- 2: Is there any significant relationship between Iranian EFL learners' cognitive/ metacognitive strategy use (high vs. low) and integrated listening/speaking performance?
- 3: Is there any significant relationship between Iranian EFL learners' foreign language anxiety (FLA, high vs. low) and integrated listening/speaking performance?

### 3. Method

This study was aimed at finding the role of three individual attributes of AT, FLA, the use of cognitive/metacognitive strategies in performing an integrated speaking/listening test task. The research design most appropriate for our purposes was associational, and this study followed the principles of correlational studies in conducting this research. Such a research design is appropriate when researchers are interested in identifying a link between two or more variables (Author, 2013).

#### 3.1. Participants and Setting

Participants of the study were 70 EFL university students (30 male and 40 female) with the age range of 21-35 ( $M = 20.81$ ,  $SD = 1.24$ ). They were all university students studying English Language Teaching at Islamic Azad University in a northwestern city in Iran, and were selected using convenience sampling. They were native speakers of Turkish (Azeri) who had the experience of English language learning with no opportunity to live in an English-speaking country or be involved in real-life interactions. They took the Oxford Quick Placement Test to measure their proficiency level. Following Ansarin et al. (2021), learners were considered B2 (Upper-Intermediate) with a score of 37-47 and C1 (Advanced) with a score of 48-55. The minimum possible score was 0 and the maximum was 60, where a score of 1-17 could be considered equivalent to Beginner, 18-27 as Elementary, 28-37 as Intermediate, and 56-60 as Very Advanced level. Among the participants who took the test, 10 learners were excluded as outliers: 6 male learners, for having B1 level of proficiency; and 4 female learners, for having C2 level of proficiency. The final number of participants who were qualified for this study were 24 male and 36 female learners. Moreover, the participants were categorized into high vs. low groups of ambiguity tolerance (AT), cognitive/metacognitive strategy use, and foreign language anxiety (FLA) based on their answers to three distinct questionnaires (details below).

#### 3.2. Instrumentation

**3.2.1. Oxford Quick Placement Test.** This study employed 5 instruments for data collection. The first apparatus used in the present study was composed of a standard sample Oxford Quick Placement Test that was used to measure the proficiency levels of the learners. The test included 60 items in two parts incorporating multiple-choice items and cloze tests. The Oxford Quick Placement Test measures knowledge of grammar, vocabulary, reading comprehension as well as appropriate use of language. The time required for taking the test was about 30 minutes. Oxford Quick Placement Test is acceptably valid (Hassaskhah & Roudsari, 2015) and reliable ( $r = .70$ ).

**3.2.2. Ambiguity Tolerance Questionnaire.** Besides, a standardized questionnaire of ambiguity tolerance (AT) was exploited in order to categorize participants into groups of low (119 - 175) vs. high

(176 - 233) AT. The AT questionnaire (Norton, 1975) comprised 61 questions and measured the overall AT level of individuals in the form of a Likert scale. The questionnaire included sub-segments including philosophical, interpersonal communication, public image, job-related, problem-solving, social, habitual, and art-form related aspects of human life. The questionnaire was piloted on 10 similar EFL students who were chosen through convenient sampling from the same institution. Based on the results of the pilot test, it was found that the participants had no difficulty understanding the items in English, so the questionnaire was conducted in English in the main study. Previous research has demonstrated the questionnaire to enjoy appropriate levels of internal reliability ( $r = .85$ ) and criterion validity ( $r = .9$ ).

**3.2.3. Cognitive/Metacognitive Strategy Use Questionnaire.** The second questionnaire was a cognitive/metacognitive strategy use questionnaire by Phakiti (2006). The questionnaire used a 5-point Likert scale (never, occasionally, sometimes, often, always) to uncover the level of cognitive/metacognitive strategy use and divided participants into low (67 - 99) versus high (100 - 132) groups in terms of their cognitive/metacognitive strategy use. The questionnaire collected information on comprehending, retrieval, planning, and monitoring strategies through its 30 items. The test was piloted before administration and was found to be both reliable ( $r = .72$ ) and valid. The validity of the questionnaire was ascertained by checking each item against the measured construct and it was ensured that all items were related to cognitive and metacognitive strategy use rather than any other attribute.

**3.2.4. Foreign Language Classroom Anxiety Questionnaire.** The last questionnaire was the FL classroom anxiety scale by Horwitz et al. (1986) that used a Likert scale to uncover the level of low (62 - 103) vs. high (104 - 144) individuals' foreign language anxiety (FLA) through 33 questions. The items were related to communication apprehension, test anxiety, and fear of negative evaluation in the foreign language classroom. The scale was pilot tested with 10 students similar to the target population and was found to be appropriate for the study in terms of content and difficulty level. Also, earlier studies have reported high levels of internal reliability ( $r = .78$ ) and criterion validity ( $r = .86$ ) for the questionnaire.

### 3.3. Procedure

The integrated listening/speaking section of a sample TOFEL iBT was administered to all participants in order to assess their performance on an integrated test and find out the possible relationship between the named individual characteristics and the participants' oral performance. The integrated test included 6 listening/speaking tasks. The total required time for taking the test was 20 minutes. The oral performances were recorded and transcribed to facilitate the scoring procedure. The performances were rated against TOFEL-iBT speaking rubrics of general description, delivery, language use, and topic development. Two raters were involved and the inter-rater consistency amounted to 80 percent.

## 4. Results and Discussion

In order to assess the relationship between EFL learners' integrated listening/speaking performance and the variables mentioned earlier (AT, FLA, and Cognitive/Metacognitive Strategy Use), the following analyses were conducted. The descriptive statistics indicate mean score and standard deviations of variables as follows: Integrated listening/speaking (16.9, 4.58), cognitive and metacognitive strategy use (103.6, 15.31), AT (174.9, 28.02), FLA (104.4, 17.99), and placement test (48.5, 5.00).

In order to answer the three research questions, a number of statistical analyses were conducted. Prior to all analyses, the normality of the groups in the study was examined by means of the Kolmogorov-Smirnov test. The obtained data from all measures including integrated listening/speaking scores, AT, FLA, and cognitive and metacognitive strategy use confirmed that the groups were normally distributed ( $\text{sig} > .05$  and skewness and kurtosis statistics were within  $\pm 1$ ).

Additionally, based on the results of Levene Statistics, it can be claimed that the learners of the study were homogenous in terms of language proficiency ( $f(2,59) = 0.17, p = .528$ ).

Also, in order to prevent raters' bias and to have more reliable judgments, inter-rater reliability was computed to make sure there is enough consistency between the two raters' scoring of learners'

final integrated listening/speaking scores. The result of Cronbach's Alpha for inter-rater reliability indicates a strong concordance between raters ( $r = .901$ ).

In order to test whether the students' scores on foreign language anxiety (FLA), AT, and cognitive and metacognitive strategies were significantly related to their performance in an integrated listening/speaking performance, Pearson Product Moment Correlation was employed for each research question. As shown in Table 1, there is a moderate level positive relationship between listening/speaking performance and AT ( $p = .000$ ,  $r = .58$ ). Also, based on Table 2, a higher AT level correlated more with integrative assessment ( $r = .111$ ) than a lower AT level ( $r = .015$ ), which implies that a high AT can influence learners' performance on integrated listening/speaking tests in a positive way, not in a statistically significant different way though.

**Table 1**  
*Correlation between Integrated Listening/Speaking and AT*

		I	AT
I	Pearson Correlation	1	.582
	Sig. (2-tailed)		.000
	N	60	60
AT	Pearson Correlation	.582	1
	Sig. (2-tailed)	.000	
	N	60	60

Note. I = Integrated Listening/Speaking; AT = Ambiguity Tolerance

**Table 2**  
*Correlation between Integrated Listening/Speaking and Low and High AT*

		I	Low AT	High AT
I	Pearson Correlation	1	.015	.111
	Sig. (2-tailed)		.532	.734
	N	60	36	24
High AT	Pearson Correlation	.111	.944	1
	Sig. (2-tailed)	.734	.000	
	N	24	24	24
Low AT	Pearson Correlation	.015	1	.944
	Sig. (2-tailed)	.532		.000
	N	36	36	24

Note. I = Integrated Listening/Speaking; AT = Ambiguity Tolerance

To find out if cognitive/metacognitive strategy use has any relationship with integrated listening/speaking test, another Pearson Correlation was employed, shown in Table 3. Similar to the AT story, there is a positive and moderate relationship between integrated listening/speaking performance and cognitive and metacognitive strategy use ( $p = .000$ ,  $r = .47$ ). Additionally, based on Table 4, performance on the integrated test was more correlated with high strategy use ( $r = .080$ ) than with low strategy usage ( $r = .020$ ,  $\text{sig} > .05$ ). In other words, cognitive and metacognitive strategy use can influence learners' performance on integrated listening/speaking tests in a positive way although there is a no significant difference between low user and high users of the strategies.



**Table 3***Correlation between Integrated Listening/Speaking and Cognitive and Metacognitive Strategy Use*

		I	C
I	Pearson Correlation	1	.472
	Sig. (2-tailed)		.000
	N	60	60
C	Pearson Correlation	.472	1
	Sig. (2-tailed)	.000	
	N	60	60

Note. I = Integrated Listening/Speaking; C = Cognitive and Metacognitive Strategy Use

**Table 4***Correlation between Integrated Listening/Speaking and Low and High Cognitive and Metacognitive Strategy Use*

		I	Low C	High C
I	Pearson Correlation	1	.020	.080
	Sig. (2-tailed)		.927	.643
	N	60	24	36
Low C	Pearson Correlation	.020	1	.929
	Sig. (2-tailed)	.927		.000
	N	24	24	24
High C	Pearson Correlation	.080	.929	1
	Sig. (2-tailed)	.643	.000	
	N	36	24	36

Note. I = Integrated Listening/Speaking; C = Cognitive and Metacognitive Strategy Use

Another Pearson correlation was run to investigate the connection between FLA and integrated listening/speaking performance. Based on the results of correlational analysis shown in Table 5, it can be concluded that there is a significant but negative relationship between integrated listening/speaking performance and FLA ( $p = .000$ ,  $r = .84$ ). According to statistics in Table 6, performance on the integrated test was correlated almost equally with lower levels of FLA ( $r = -.109$ ) and higher levels of FLA ( $r = .115$ ,  $\text{sig} > .05$ ), with opposite directions though.

**Table 5***Correlation between Integrated Listening/Speaking and FLA*

		I	FLA
I	Pearson Correlation	1	-.843
	Sig. (2-tailed)		.000
	N	60	60
FLA	Pearson Correlation	-.843	1
	Sig. (2-tailed)	.000	
	N	60	60

Note. I = Integrated Listening/Speaking; FLA = Foreign Language Anxiety



**Table 6***Correlation between Integrated Listening/Speaking and Low and High FLA*

		I	Low FLA	High FLA
I	Pearson Correlation	1	-.109	.115
	Sig. (2-tailed)		.613	.503
	N	60	24	36
Low FLA	Pearson Correlation	-.109	1	.932
	Sig. (2-tailed)	.613		.000
	N	24	24	24
High FLA	Pearson Correlation	.115	.932	1
	Sig. (2-tailed)	.503	.000	
	N	36	24	36

*Note.* I = Integrated Listening/Speaking; FLA = Foreign Language Anxiety

The above analyses indicated that individual features were more or less related to performance on an integrated listening/speaking test. Among these features, FLA had the most significant relationship and was negatively related to performance on integrated listening/speaking assessment. In addition, integrated listening/speaking scores were overall positively related to AT and cognitive and metacognitive strategy use. These findings imply that performance on integrated listening/speaking assessments is not solely influenced by learners' linguistic or language knowledge; it is also affected by individual features such as FLA, AT, and cognitive and metacognitive strategy use. However, the categorization of individuals to high versus low possessors of each attribute did not reveal significant differences between the groups.

As one of the foremost aims of language learners is to be able to communicate more efficiently, most research focus has so far been devoted to investigating the impact of various factors on language learning. Language assessment has only been second to language teaching/learning and that is why there has been comparatively scarce research attention in this domain. More specifically, while due attention has been devoted to learners' personal characteristics in their language learning success, the role of such individual attributes in assessment and specially in integrated assessment tasks has largely remained unknown and unexplored.

The present study was accordingly intended to examine the relationship between individual features of ambiguity tolerance (AT), cognitive/metacognitive strategy use, foreign language anxiety (FLA), and performance on integrated listening/speaking tests by Iranian EFL learners. Results indicated that individual differences have the potential to affect performance on integrated listening/speaking test in both positive and negative ways. This is due to the observation that overall, all the individual attributes were significantly related to integrated speaking/listening assessment either positively (AT and strategy use) or negatively (FLA). However, no significant differences were found when sub-groups were compared. It was observed that students with higher AT levels and more cognitive/metacognitive strategy use performed better (not significantly though) on an integrated listening/speaking test than those with lower AT levels or less frequent use of cognitive/metacognitive strategies. Understandably, higher levels of AT facilitate the test-taker's task in bridging the gap between a listening activity and the subsequent speaking, relieving him from the fear of forgetting old information when more new information flows in. Also, more use of cognitive/metacognitive strategies can fill in some of the loopholes in students' language knowledge, leading to a better performance.

The findings also suggest that learners with lower FLA levels perform differently from those with higher levels of FLA; however, our study did not find any significant differences among study groups despite the expectation that higher level FLA group should have done much more differently from lower-level FLA group. Although further research is called for to substantiate the direction and extent of these differences, one thing is clear. These findings lend support to the hypothesis that performance on integrated listening/speaking tests can be affected by individual characteristics and that the resulting score is a function of both relevant and irrelevant abilities. Only when considering their impact can language teachers and assessors have reliable evaluations of learners' oral task performance. Our observations are in line with, and can be interpreted according to, the major assumptions of

dominant language assessment theories such as Classical True Score model, where the observed score is a function of true score and error score. As such, by controlling the contribution of individual attributes to test performance, test makers and assessment researchers can play a significant role in minimizing the influence of construct irrelevant factors on the assessment process (construct) and product (test score).

Results of the present research are in line with a number of other relatively similar studies, which have found significant correlations between individual characteristics including AT, cognitive and metacognitive strategy use and FLA, and language learning. Regarding the relationship between AT and language performance, the findings of this study support those of Chapelle and Roberts (1986), for instance, who studied the role of AT in learners' performance on the TOEFL test, and found that AT was a significant predictor of the learners' performance. Similarly, Başöz (2015) found a statistically significant relationship between AT and achievement in a foreign language. Both these studies imply that the observed score on a TOEFL test and achievement tests is a reflection of the candidates' true abilities (their language-related knowledge) and their other language-irrelevant, personal attributes (AT here). However, to the best of our knowledge, almost no research has been conducted on the link between AT and performance on the integrated listening/speaking assessment, and further research is required to tease out the contribution of not only AT but numerous other individual attributes in test performance (whether in isolated or integrated formats). Preliminary data from this study indicates that there is a link between AT and integrative listening/speaking assessment, such that higher and lower levels of AT contribute differently to the observed score. In any sense, however, the resulting observed score is an inaccurate reflection of the true score, which should be a function of only the integrated listening/speaking test, nothing less or more.

Regarding the relationship between cognitive/metacognitive strategy use, the findings of this study are in conformity with studies which found a similar positive relationship between cognitive and metacognitive strategy use and language learning. Phakiti (2003) found that using cognitive and metacognitive strategies has an expediting role in learners' performance. Similarly, Zhang and Seepho (2013) indicated the positive relation between cognitive and metacognitive strategy use and achievement of EFL Chinese learners. Within the context of language assessment, Phakiti (2006) reported that highly successful test takers utilize more cognitive and metacognitive strategies in language assessment and that these two variables are positively correlated. Given that our study is never enough to make any generalizations about the findings and that no previous study exists on the relationship between cognitive/metacognitive strategy use and integrated listening/speaking assessment, further studies are required to explore this link to determine how much of the observed score is due to the specific construct and how to minimize the effect of irrelevant variables in order to produce more valid test scores.

FLA has often been reported to influence language learning negatively (e.g., Cubukcu 2007; Dewaele et al., 2023; Horwitz et al., 1986; Liu 2007; MacIntyre and Gardner 1991). MacIntyre and Gardner (1994) claimed that language anxiety was the best single correlate of achievement. In a similar vein, Ganschow et al. (1994) demonstrated that students with high levels of anxiety exhibited poorer language skills. Bagheri and Ghanizadeh (2016) explored the interrelationship between different variables including task value, effort-regulation, AT, test anxiety, learning strategies, and language achievement. Their study with 180 EFL learners found a negative influence of test anxiety on deep learning. Ganschow and Sparks (1996) also reported that students with low anxiety levels outperformed those with high anxiety levels overall. However, Babin, et al. (2012) reported no significant relationship between anxiety and performance. Evidently, research on FLA has concentrated on its role in language learning and achievement. Our research data did not show any strong and conclusive links between different levels of FLA and integrated listening/speaking assessment. In light of this, and given that test-taking is already a stressful condition, examining the link between FLA and test performance is highly warranted. After establishing the exact link between FLA and test performance, the next important task would be to identify anxiety-provoking factors and try to alleviate them to boost the observed score, making it a closer representation of the true score. Few studies have explored the effect of individual and psychological features such as FLA on the students' integrated listening/speaking performance in the literature (Crossley & Kim, 2019). This study lends further support to the general

observation of previous research that FLA has a detrimental effect on language performance, even in an integrated context.

## 5. Conclusion

This study investigated the relationship between three individual features of ambiguity tolerance (AT), cognitive and metacognitive strategies, and foreign language anxiety (FLA) and integrated listening/speaking TOEFL-iBT task performance. EFL learners' oral performance was assessed in terms of general description, delivery, language use, and topic development. The study results demonstrated a strong inverse relationship between overall FLA and integrated listening/speaking test scores, as well as a considerable positive relationship between overall AT and employment of more cognitive and metacognitive strategies and integrated listening/speaking test scores. However, no significant differences were observed between high versus low groups, possibly due to operationalization method. Using scores at the extreme ends as high versus low groups could have yielded different results.

This investigation has some important implications for EFL English language instructors and learners, material developers and syllabus designers, test developers, test users, assessment researchers, and the educational system. On the basis of the findings reported here, it is recommended that EFL instructors pay more attention to ongoing assessment rather than relying on one-shot final exams to minimize the impact of individual differences and other uncontrolled variables on test outcomes, thereby providing more accurate accounts of learners' actual skills and abilities. Teachers should accordingly focus more on a balanced summative and formative assessment during the course of instruction. L2 assessment experts as well as users of test results including educational authorities and employers are advised to interpret score outcomes with care as well as not relying in their decisions on data coming only from one assessment source: in making high-stakes decisions, teachers, tester and other decisions makers should make an attempt to tease out the function of numerous other irrelevant variables that could have polluted the students' test performance.

Although the researchers tried to control as many variables as possible, the study, nonetheless, suffered from certain limitations that might have affected the results. Access was limited to small number of learners who could fit well with the purpose of the study. Consequently, a relatively small sample size of 60 participants makes the findings cautiously generalizable. Future studies should consider bigger sample sizes as well as adding diverse age groups, proficiency levels, gender, and learners studying different majors or perhaps learners of other languages. Also, researchers can explore other individual features (e.g., motivation, working memory capacity, and field dependence that can presumably exert an influence on the result of integrated tests. By addressing these limitations and expanding the scope of research, future studies can provide a more comprehensive understanding of the factors affecting integrated listening/speaking test performance and develop more effective assessment strategies.

## Declaration of Conflicting Interests

The authors have no conflicting interests.

## Funding

There was no funding for this study.

## References

- Ansarin, A.A., Karafkan, M.A., & Hadidi, Y. (2021). The effects of task type on Iranian EFL learners' use of lexical diversity and sophistication. *Applied Research on English Language*, 10(4), 39-70. <https://doi.org/10.22108/are.2021.126660.1673>
- Ariyanti, A. (2016). Psychological factors affecting EFL students' speaking performance. *ASIAN TEFL Journal of Language Teaching and Applied Linguistics*, 1(1), 77-88. <http://dx.doi.org/10.21462/asianteftl.v1i1.14>
- Babin, E. A., Palazzolo, K. E., & Rivera, K. D. (2012). Communication skills, social support, and burnout among advocates in a domestic violence agency. *Journal of Applied Communication Research*, 40(2), 147-166. <https://doi.org/10.1080/00909882.2012.670257>



- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bagheri, F., & Ghanizadeh, A. (2016). Critical thinking and gender differences in academic self-regulation in higher education. *Journal of Applied Linguistics and Language Research*, 3(3), 133-145.
- Barkaoui, K., Brooks, L. A., Swain, M., & Lapkin, S. (2013). Test-takers' strategic behaviors in independent and integrated speaking test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics*, 34(3), 304-324. <https://doi.org/10.1093/applin/ams046>
- Başöz, T. (2015). Exploring the relationship between tolerance of ambiguity of EFL learners and their vocabulary knowledge. *Journal of Language and Linguistic Studies*, 11(2), 53-66.
- Brown, A., Iwashita, N., McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks. *ETS Research Report Series*, 2005(1), i-157. <https://doi.org/10.1002/j.2333-8504.2005.tb01982.x>
- Budner, N. S. (1962). Intolerance of ambiguity as a personality variable. *Journal of Personality*, 30(1), 29-50. <https://doi.org/10.1111/j.1467-6494.1962.tb02303.x>
- Butler, F. A., Eignor, D., Jones, S., McNamara, T., & Suomi, B. K. (2000). *TOEFL 2000 speaking framework A working paper (Vol. 20)*. Educational Testing Service.
- Chapelle, C., & Roberts, C. (1986). Ambiguity tolerance and field independence as predictors of proficiency in English as a second language. *Language Learning*, 36(1), 27-45. <https://doi.org/https://doi.org/10.1111/j.1467-1770.1986.tb00367.x>
- Crossley, S. A., Clevinger, A., Kim, Y. (2014). The role of lexical properties and cohesive devices in text integration and their effect on human ratings of speaking proficiency. *Language Assessment Quarterly*, 11(3), 250-270. <https://doi.org/10.1080/15434303.2014.926905>
- Crossley, S. A., & Kim, Y. (2019). Text integration and speaking proficiency: Linguistic, individual differences, and strategy use considerations. *Language Assessment Quarterly*, 16(2), 217-235. <https://doi.org/10.1080/15434303.2019.1628239>
- Cubilo, J., & Winke, P. (2013). Redefining the L2 listening construct within an integrated writing task: Considering the impacts of visual-cue interpretation and note-taking. *Language Assessment Quarterly*, 10(4), 371-397. [10.1080/15434303.2013.824972](https://doi.org/10.1080/15434303.2013.824972)
- Cubukcu, F. (2007). Foreign language anxiety. *Iranian Journal of Language Studies*, 1(2), 133-142.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10(1), 1-75. <https://doi.org/10.1016/j.asw.2005.02.001>
- Dewaele, J. M., Botes, E., & Meftah, R. (2023). A Three-Body Problem: The effects of foreign language anxiety, enjoyment, and boredom on academic achievement. *Annual Review of Applied Linguistics*, 1-16. <https://doi.org/10.1017/S0267190523000016>
- Doubet, K., & Southall, G. (2017). Integrating reading and writing instruction in middle and high school: The role of professional development in shaping teacher perceptions and practices. *Literacy Research and Instruction*, 57(1), 59-79. <https://doi.org/10.1080/19388071.2017.1366607>
- Fitriani, D. A., & Apriliawati, R. (2015). A study on student's English speaking problems in speaking performance. *Jurnal Pendidikan dan Pembelajaran Khatulistiwa*, 4(9), 1-13. <https://dx.doi.org/10.26418/jppk.v4i9.11345>
- Frost, K., Elder, C., Wigglesworth, G. (2012). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test takers' oral performances. *Language Testing*, 29(3), 345-370. <https://doi.org/10.1177/0265532211424479>
- Frost, K., Clothier, J., Huisman, A., & Wigglesworth, G. (2020). Responding to a TOEFL iBT integrated speaking task: Mapping task demands and test takers' use of stimulus content. *Language Testing*, 37(1), 133-155. <https://doi.org/10.1177/0265532219860750>
- Frost, K., Wigglesworth, G., & Clothier, J. (2021). Relationships between comprehension, strategic behaviours and content-related aspects of test performances in integrated speaking tasks. *Language Assessment Quarterly*, 18(2), 133-153. <https://doi.org/10.1080/15434303.2020.1835918>



- Fulcher, G. (2015). Assessing second language speaking. *Language Teaching*, 48(2), 198–216. <https://doi.org/10.1017/S0261444814000391>
- Ganschow, L., & Sparks, R. (1996). Anxiety about foreign language learning among high school women. *The Modern Language Journal*, 80(2), 199–212. <https://doi.org/10.1111/j.1540-4781.1996.tb01160.x>
- Ganschow, L., Sparks, R. L., Anderson, R., Javorshy, J., Skinner, S., & Patton, J. (1994). Differences in language performance among high-, average-, and low-anxious college foreign language learners. *The Modern Language Journal*, 78(1), 41–55. <https://doi.org/10.2307/329251>
- Goh, C. C. M. (1998). How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research*, 2(2), 124–147. <https://doi.org/10.1177%2F136216889800200203>
- Green, J. M., & Oxford, R. (1995). A closer look at learning strategies, L2 proficiency, and gender. *TESOL Quarterly*, 29(2), 261–297. <https://doi.org/10.2307/3587625>
- Hassaskhah, J., & Roudsari, F. H. (2015). The relationship between learners' academic self-concept and the consequential validity of quick placement Test. *International Journal of English and Education*, 4(4), 240–252.
- Hewitt, E., & Stephenson, J. (2012). Foreign language anxiety and oral exam performance: A replication of Phillips's MLJ study. *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. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.2307/327317>
- Huang, H. T. D., & Hung, S. T. A. (2010). Examining the practice of a reading-to-speak test task: Anxiety and experience of EFL students. *Asia Pacific Education Review*, 11, 235–242. <https://doi.org/10.1007/s12564-010-9072-6>
- Huang, H. T. D., Hung, S. T. A., & Plakans, L. (2016). Topical knowledge in L2 speaking assessment: Comparing independent and integrated speaking test tasks. *Language Testing*, 35(1). <https://doi.org/10.1177%2F0265532216677106>
- Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds), *Handbook of Second Language Assessment* (pp. 131–146). DeGruyter Mouton. <https://doi.org/10.1515/9781614513827-011>
- Kamran, S. K. (2011). Effect of gender on ambiguity tolerance of Iranian English language learners. *Journal of Education and Practice*, 2(11), 25–32.
- Kamran, S. K., & Maftoon, P. (2015). An analysis of the associations between ambiguity tolerance and EFL an analysis of the associations between ambiguity tolerance and EFL reading strategy awareness. *English Language Teaching*, 5(3), 188–196. <https://doi.org/10.5539/elt.v5n3p188>
- Kunnan, A. J. (1995). *Test taker characteristics and test performance: A structural modeling approach*. Cambridge University Press.
- Kunnan, A. J. (1998). *Validation in Language Assessment* (1st ed.). Routledge.
- Lee, Y. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing*, 23(2), 131–166. <https://doi.org/10.1191%2F0265532206lt325oa>
- Lee, K., & Ye, Y. (2021). The underlying structure of foreign language anxiety in integrated speaking assessment: A mixed methods study. *Language Education and Assessment*, 4(3), 105–121. <https://doi.org/10.29140/lea.v4n3.554>
- Lewkowicz, J. A. (1997). The integrated testing of a second language. *Encyclopedia of Language and Education*, 7, 121–130.
- Liu, M. (2007). Language anxiety in EFL situations. *International Journal of Applied Linguistics*, 153, 53–76. <https://doi.org/10.2143/ITL.153.0.2022821>
- MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Languages Learning*, 41(4), 513–534. <https://doi.org/10.1111/j.1467-1770.1991.tb00691.x>
- MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, 44(2), 283–300. <https://doi.org/10.1111/j.1467-1770.1994.tb01103.x>

- Norton, R. W. (1975). Measurement of ambiguity tolerance. *Journal of Personality Assessment*, 39(6), 607–619. [https://doi.org/10.1207/s15327752jpa3906\\_11](https://doi.org/10.1207/s15327752jpa3906_11)
- O'Malley, J. M., & Chamot, A. U. (1989). *Learning strategies in second language acquisition*. Cambridge University Press.
- Oya, T., Manalo, E., & Greenwood, J. (2004). The influence of personality and anxiety on the oral performance of Japanese speakers of English. *Applied Cognitive Psychology*, 18(7), 841–855. <https://doi.org/10.1002/acp.1063>
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26–56. <https://doi.org/10.1191/0265532203lt243oa>
- Phakiti, A. (2006). Modeling cognitive and metacognitive strategies and their relationships to EFL reading test performance. *Language Testing*, 11(1), 53–94. <https://search.informit.org/doi/10.3316/aeipt.165763>
- Plakans, L., Gebriel, A. (2012). A close investigation of source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18–34. <https://doi.org/10.1016/j.asw.2011.09.002>
- Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test Takers: A structural equation modelling approach. *Language Testing*, 15(3), 333–379. <https://doi.org/10.1177/026553229801500303>
- Resnik, P., Dewaele, J. M., & Knechtelsdorfer, E. (2023). Differences in the intensity and the nature of foreign language anxiety in in-person and online EFL classes during the pandemic: A mixed-methods study. *TESOL Quarterly*, 57(2), 618–642. <https://doi.org/10.1002/tesq.3177>
- Rukthong, A. (2021). MC listening questions vs. integrated listening-to-summarize tasks: What listening abilities do they assess? *System*, 97, <https://doi.org/10.1016/j.system.2020.102439>
- Rukthong, A., & Brunfaut, T. (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. *Language Testing*, 37(1), 31–53. <https://doi.org/10.1177/0265532219871470>
- Sadeghi, K. (2013). Doubts on the validity of correlation as a validation tool in second language testing research: the case of cloze testing. *Language Testing in Asia*, 3(15), 1–7. <https://doi.org/10.1186/2229-0443-3-15>
- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26(1), 113–125. <https://doi.org/10.1023/A:1003044231033>
- Souzandehfar, M. (2024). New perspectives on IELTS authenticity: An evaluation of the speaking module. *International Journal of Language Testing*, 14(1), 34–55. <https://doi.org/10.22034/IJLT.2023.409599.1272>
- Swain, M., Huang, L. S., Barkaoui, K., Brooks, L., & Lapkin, S. (2009). The speaking section of the TOEFL iBT™ (SSTiBT): Test-takers' reported strategic behaviors. *ETS Research Report Series*, 2009(2), i–118. <https://doi.org/10.1002/j.2333-8504.2009.tb02187.x>
- Wesche, M. B. (1987). Second language performance testing: The Ontario test of ESL as an example. *Language Testing*, 4(1), 28–47. <https://doi.org/https://doi.org/10.1177%2F026553228700400103>
- Weigle, S. C., Parker K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, 21(1), 118–133. <https://doi.org/10.1016/j.jslw.2012.03.004>
- Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly*, 10(1), 110–114. <https://doi.org/10.1080/15434303.2013.766744>
- Zhang, L., & Seepho, S. (2013). Metacognitive strategy use and academic reading achievement: Insights from a Chinese context. *Electronic Journal of Foreign Language Teaching*, 10(1), 54–69.
- Zhang, W., & Wilson, A. (2023). From self-regulated learning to computer-delivered integrated speaking testing: Does monitoring always monitor? *Frontiers in Psychology*, 14, <https://doi.org/10.3389/fpsyg.2023.1028754>