International Journal of Language Learning and Applied Linguistics

ISSN: 2835-1924
Volume 2 | No 12 | Dec-2023

Generalizability Theory as a Tool in Examining Classroom Anxiety

Najeba Rashed Mahamed (M.E.)
Northern Technical University. Iraq, Technical Institute. Kirkuk
Email: najebaahmed884@ntu.edu.iq
Phone number: 07701242868

Abstract: Theory by Horwitz et al. (1986): Foreign Language Classroom Anxiety Scale. This hypothesis has been crucial to studying individual differences. Different perspectives may affect second- and foreign-language acquisition. The scale's multiplied organization has been extensively studied, giving validity evidence for Foreign Language Anxiety, a concept of great importance to foreign language specialists and instructors. Despite significant investigation, nothing is known about the scale's dependability. The Foreign Language Classroom Anxiety Scale is a self-report instrument, therefore the theory under study should account both item internal consistency and response consistency across time. The generalizability theory is used to examine the Foreign Language Classroom Anxiety Scale's in terms of classroom anxiety. The theory was applied in an Iraqi higher education institution where English language is taught as a major subject. The theory applied showed a good compound generalizability and dependability. One of the three latent qualities anticipated to appear on the scale, test anxiety, was very unreliable.

Keywords: Generalizability, Examining, Classroom, Anxiety.

Introduction

Given the correlation between education and upward social mobility, it is imperative for the school system to provide pedagogical approaches that facilitate students' proficiency in many languages. This would provide individuals the chance to attain equitable access to educational resources and economic prospects. In 2002, Malaysia implemented English use as a means of instruction at the academia. The implementation of this policy, originally proposed by Mahathir Mohamad, the former Prime Minister of Malaysia, has sparked considerable scholarly discussion. The recognition of the declining proficiency in English on the part of the learners at both the primary and tertiary education levels across Malaysia has prompted the implementation of policy revisions within the state educational system. A decision is made by the authorities to implement the English language in educational institutions, namely in instructing specific courses, has elicited a multitude of divergent perspectives and apprehensions. Notwithstanding these issues, the authorities proceeded to establish English as the language of teaching at the admission level in 2004 and at the academia one in 2005.

Concerns over the ability of students and instructors to meet international standards have prompted a need for modifications to teaching methodologies in Malaysian educational settings. Furthermore, the alteration of the language policy sparked extensive discussions on the well-being and academic performance of students. The recognition of the significant influence of foreign

language learning anxiety on language learners has been well recognised (Ohata, 2005). The recognition of the significance of the English language has been extensively recognised throughout the Asian area, with special emphasis on the People's Republic of China (PRC), subsequent to its accession to the World Trade Organization (WTO). Furthermore, the English language saw a surge in popularity inside the People's Republic of China (PRC) after the hosting rights for the 'Olympic Games' were awarded to Beijing in 2008. According to Bolton (2002), policymakers, educators, and citizens in China see a strong correlation between English proficiency and sustained economic development. In the context of China, Lam (2002) documented an enhancement in the incentive to acquire English language skills, both at the national and individual levels. Additionally, there have been observed improvements in the provision of educational assistance. The phenomenon under consideration has been in existence since at least 1978, a pivotal year marked by enhanced national initiatives aimed at promoting the English language. In order to address the growing needs for English education, it is imperative to enhance teacher training programs and commit more financial resources towards the acquisition of improved learning materials and facilities. an essential component of its modernisation efforts as "necessary for acquiring technological expertise and for fostering international trade" (Adamson & Morris, 1997, p. 3).

Since its development in the mid-1980s by Horwitz, & Cope, 1986, the Foreign Language Classroom Anxiety Scale has been a prominent self-report measure for examining foreign language learners' classroom anxiety. The theory under research helped establish particular area of knowledge and envision of anxiety as originating in the language teaching setting (Teimouri et al., 2019). Many studies have shown that Foreign Language Anxiety is adversely connected with other affecting changings like readiness to interconnect (Liu & Jackson, 2008), and handling comfortableness (Fallah, 2017). This study focuses on dependability, a neglected feature of the Foreign Language Classroom Anxiety Scale, to strengthen its methodological foundation in teaching the English language.

Many studies have examined the Foreign Language Classroom Anxiety Scale's structure. Despite the Foreign Language Classroom Anxiety Scale's construct validity, most research has relied on conventional measures of test-retest coefficient. These standards only examine a few measurement error causes, giving a limited view of mark steadiness. One could ask how reliable research studies' FLA interpretations and conclusions are. As the Foreign Language Classroom Anxiety Scale is a self-report instrument, scoring consistency is crucial to understanding the construct and helping teachers and researchers make informed decisions about teaching practices and research findings.

As mentioned previously, Generalizability Theory addresses the limits of traditional dependability estimations, that "permits a multifaceted perspective on measurement error and its components" (Brennan, 2001a, p. viii) with self-determining assessment of data gathering equipment capacity fault causes. Thus, the chief aim the current study works for, is to use Generalizability Theory to expand evidence for the Foreign Language Classroom Anxiety Scale and assess whether its common uses are reliable.

Review of Literature

Any measuring technique should aim for a goal, which is directly related to the significance of implications, explanations, and activities concerning assessment objects.

Many applied linguists and their subfields are concerned with research investigations (Norris & Ortega, 2000). Failure to disclose dependability estimates is a common complaint (Plonsky & Derrick, 2016). Reliable variables underpin validity, therefore without examining, study results and claims are difficult to evaluate. Noting that research occasionally gives results from previous studies.

Reporting data is vital, according to Purpura, Brown, & Schoonen (2015). Therefore, dependability must be experimentally evaluated through studies that use quantitative and statistical processes to validate a test or instrument (Woodrow, 2014). Therefore, robust research checks are essential to the iterative, developing process of test/instrument validation.

Generalizability Theory: An Overview

Generalizability (G) theory is a psychometric theory that use a numerical sample strategy to split marks into their respective sources of variance. The concept of G theory was first presented by Cronbach and others (Cronbach, Rajaratnam, & Gleser, 1963) as an expansion of Cronbach's seminal work on coherent of the text. Subsequently, several scholars have made significant contributions to the development of G theory (Brennan, 2000, 2001; Webb, Shavelson, & Haertel, 2006). This study aims to assess the reliability and validity of score interpretations by acknowledging and quantifying the extent of measurement error originating from various factors.

The present discourse elucidates the theoretical underpinnings and practical implementations of G theory, tailored specifically for a readership including clinical psychologists rather than psychometricians. The present discourse introduces an illustrative instance to explicate the fundamental principles of G theory, providing a comprehensive review of its theoretical underpinnings and practical applications. The following section provides a description of one-facet and multi-facet designs, including a comprehensive presentation of the fundamental principles covered, accompanied by illustrative examples. The passage finishes by providing a concise overview of significant subjects pertaining to the application of G theory, while also highlighting the availability of supplementary materials for further exploration.

Conceptual Foundations of G Theory

In the context of clinical psychology, it is customary for practitioners to employ psychological or behavioral instruments to assess various phenomena of interest, such as bullying, anxiety, racial identity, belonging, or learning disability. The primary objective of these assessments is to obtain a reliable and consistent estimation of an individual's level on the respective scale. This information is crucial for diagnosing the severity of the condition and determining appropriate treatment strategies. The essence of G theory is rooted in the concept that an instrument is often limited to a small subset of objects or activities from a wider pool of potential items or tasks that may have been selected instead. The primary objective of test administration is to get sufficient data from a restricted sample in order to make generalizations that may be applied to the broader population. The extent to which we may draw conclusions about a wide population based on a restricted sample is contingent upon the characteristics and extent of the errors that arise from our capacity to only sample a tiny portion of the intended population. G theory offers a framework for identifying the levels of measurement error that pose a risk to the generalization from a sample to a population. Additionally, it allows for an assessment of the reliability of such generalizations by analyzing the degree to which scores fluctuate as a result of individual differences and different sources of measurement error.

To get a comprehensive grasp of G theory, it is essential to consider four interconnected concepts: (a) the delineation of a universe based on one or more facets, (b) the division of measurement error, (c) the distinction between relative and absolute judgments, and (d) the examination of generalizability studies and decision studies. The subsequent discussion begins by contextualizing these notions inside a research investigation on the assessment of psychological distress experienced by individuals affected by disasters (Gleser, Green, & Winget, 1978). A total of twenty adult survivors who experienced a dam tragedy were subjected to independent interviews conducted by two interviewers. Subsequently, three raters independently assessed each interview for

the degree of long-term psychiatric impairment across three subdimensions (namely, anxiety, depression-isolation, and hostility-belligerence) as well as an overall scale, utilizing the Psychiatric Evaluation Form. The scale ranged from 1 (indicating no impairment) to 6 (indicating extreme impairment).

The World is Distinguished by Aspect(s)

G theory tackles the idea that marks vary both because of someone's in personality, behaviors, symptoms, capabilities, or talents. It measures errors connected with "facets" of the measurement. The variance due to individual differences is regarded to represent "wanted" or "expected" mark. Universe scores are the anticipated rate of all remarks on a person over the world of conceivable possibilities; they are comparable to true scores in classical test theory. The world is defined by identifying permissible measurement facets lacking modifying the concept of concern and exchangeable or parallel levels for each facets. Facets include test item, survey form, rater, and occasion, which determine the universe of observations. Each aspect has levels or circumstances that differ yet are similar.

Thus, a measuring technique combines aspect conditions from such a universe. Three raters rated 20 survivors' interviews by two interviewers in the survivor study (Gleser et al., 1978). This research uses a double-standard design with survivors, interviewers, and raters, designated s×i×r. The anticipated assessment of all acceptable remarks of a survivor from the universe of generalization is their universe score. The universe measurement has two interviewer and three rater levels. The universe's interviewers and raters are not random. Instead, the universe of examiners should be compared to the two study examiners in terms of training and interview cues. The raters should have similar scoring experience, training, and knowledge with the assessed construct to the three raters in the research.

Partitioning of Measurement Error

G theory conceptualises and estimates multiple score variance sources. Applying G theory requires partitioning observed score variation into person-related variance and measurement error from many sources' main and interaction effects.

Measurement inaccuracy may result from interviewer and rater errors in the survivor research. Sampling mistakes with interviewers, raters, their interactions, and other undefined causes cause measurement inaccuracies. One interviewer tended to elicit information showing consistently larger impairment levels than the other throughout all 20 interviews, which might be measurement error. G theory, supplemented by statistical techniques, estimates each source's variance component simultaneously.

Decisions of Relativity and Absoluteness

G theory defines reliable factor as the fraction of mark variation related to someone compared to a full alteration owing to someone and measurement error. G theory understands that psychologists, doctors, researchers, parents, policymakers, and managers may use scores to make relative and absolute judgments. Different G coefficients may be calculated for these two choices.

Norm-referenced score interpretations affect relative judgements. The choices include scoring consistency when evaluating or categorizing people by personality, conduct, knowledge, and/or talents. As different items were examined to find out their capability of matching familiar figures or whether marks from different interviewers judged by different raters' survivors consistently for psychiatric impairment, relative decision is involved. Absolute judgments use criterion- or domain-referenced score interpretations. They assess the precise or absolute degree of people' concept for a

field independent of others' performance, like as driving exams in daily life. Measurement judgments are based on absolute scores from different things, not relative ranking among measured people.

Anxiety in Foreign Language Learning

Since 1970, researchers have examined how anxiety affects foreign language acquisition. Language acquisition may cause anxiety during intake, processing, and output. Researching language anxiety is complicated by as they conflict indication from mechanisms used in diverse languages to measure diverse sorts of anxiety, language skills, and learning level (Onwuegbuzie, Bailey, and Daley, 2000).

Apprehension persists in university foreign language courses despite teaching methods and techniques updates. Researchers have shown that language anxiety impacts foreign language acquisition (MacIntyre, 1995; Daly, 1991; Horwitz, 1986). Anxiety affects foreign language learning. Its interactions with other emotional variables including self-esteem, inhibition, and risk-taking are complicated and hard to assess. Anxiety hinders foreign language output and success, according to many studies. About 50% of language learners suffer anxiety (Campbell & Ortiz, 1991).

MacIntyre and Gardner (1991) argue that worry "can interfere with the acquisition, retention, and production of the new language" (p. 86), which can pose many issues for foreign language students. Casado and Dereshiwsky (2001) found that foreign language anxiety persists independent of development. They said that university students' initial and previous language learning anxiety 'does not always fade or diminish'. (pp. 86). Some studies on anxiety and language acquisition may have been affected by preconceptions about anxiety. As trait and state techniques fail to capture and illustrate the core of foreign language anxiety, many researchers turn to situation-specific language environments (MacIntyre and Gardner, 1991). Trait, mood, and situational anxiety contribute to language learning anxiety. Foreign language learners often worry about how others see them in everyday situations. They also worry about speaking the target language in different situations. Another reason is fear of doing well on foreign language examinations.

Anxiety and Learning a Language

Foreign language anxiety is 'a complicated psychological construct' that is hard to define. Trylong (1987) suggests that the variables' unclear ordering may explain this. This review will concentrate on anxiety of learning a language, which Horwitz, Horwitz, and Cope (1991) say affects at least one student out of ten.

Learners of foreign languages regularly evaluate their progress (MacIntyre, Noels and Clement, 1997, p. 266). The three researchers reviewed MacIntyre and Gardener (1989) and found that extremely worried learners would feel pained and demotivated when faced with their professed foreign language difficulties. Thus, MacIntyre, Noels, and Clements (1997) stated that learners should be able to properly appraise their own skills using the right assessment instruments.

Language anxiety may hinder language being acquired, retained, and produced in foreign language learners (MacIntyre & Gardner, 1991). Ellis (1994) advises against seeing anxiety as a key factor in language acquisition success. Ellis recommended seeing anxiety as a variable that affects learners differently.

Language learners struggle to assimilate material and acquire language due to anxiety (Krashen, 1985a, 1985b; MacIntyre et al. 1997). Krashen believes that nervous pupils may learn less and be unable to display their knowledge if worry inhibits cognitive function. Thus, further failure may increase their anxiousness.

How Anxiety can be Measured?

Researchers have produced several language learning anxiety measures. MacIntyre and Gardner (1991) reviewed early language learning anxiety measurement approaches. Horwitz et al. (ed. Horwitz & Young, 1991) established the (FLCAS) in 1986 based on a 1983 study of University of Texas "Support Group for Foreign Language Learning" learners at the start of their language sessions. All items had significant corrected item-total scale correlations, indicating internal reliability of.93 for the FLCAS. Additionally, the scales had r =.83 test-retest reliability (p. 32). The FLCAS was given to 75 university students from four intact introductory Spanish classes, and the results showed that many students feel significant foreign language anxiety about foreign language learning factors. Horwitz and her colleagues found that (FLA) may cause poor emotional reactions to language acquisition.

Casado and Dereshiwsky (2001) used the FLCAS developed by Horwitz et al. to compare the perceived anxiety levels of a randomly selected sample of university students at the start of their first semester with those of an alike sample at the end of their second semester in learning Spanish. The two researchers want to see whether pupils' anxiety levels decrease as they learn the language. Despite not feeling anxious in most circumstances, starting foreign language students had stronger confidence than second semester students. Language learners' anxiety may not reduce as they learn the target language.

The Importance of Generalizability along with Classroom Anxiety

MacIntyre & Gardner (1994) describe the Foreign Language Classroom Anxiety Scale as "the worry and negative emotional reaction aroused when learning or using a second language" (p. 27). Language learners' classroom anxiety was described by Horwitz, Horwitz, & Cope, 1986 as Foreign Language Classroom Anxiety "a distinct complex of self-perceptions, beliefs, feelings and behaviors related to classroom learning arising from the uniqueness of the language learning process" (p. 128). Horwitz et al. created the 33-item Foreign Language Classroom Anxiety Scale to assess 1–5. In general, Horwitz et al. said the scale was meant to "assess the degree of anxiety, as evidenced by negative performance expectancies and social comparisons, psychological symptoms, and avoidance behaviors" (p. 559). The poll asks whether respondents feel uncomfortable in various language classroom circumstances. Since its development over 30 years ago, the Foreign Language Classroom Anxiety Scale has been the most widely used self-report instrument for classroom evaluation and research.

Most studies that administered the Foreign Language Classroom Anxiety Scale to second/foreign language learners indicated good findings. The Foreign Language Classroom Anxiety Scale's component structure has been studied mostly using Reliability of instruments (i.e., internal consistency). Though self-reported, the Foreign Language Classroom Anxiety Scale requires more scoring consistency to account for individual differences. Even when using factor analysis, few researches have addressed factor-level dependability. In order to address these deficiencies in the existing body of research and enhance the validity of the Foreign Language Classroom Anxiety Scale, this study uses Generalizability Theory to examine various sources of measurement error that may affect its reliability. This study should help assess the reliability of the Foreign Language Classroom Anxiety Scale in research and practice.

Research Questions

This research has posed the following questions that can be stated as follows:

1. How does a teacher deal with the event and item aspects to the Foreign Language Classroom Anxiety Scale?

2. What is the examining of each of the hidden qualities of the Foreign Language Classroom Anxiety Scale?

Methodology

Generalizability Theory, based on Classical Test Theory and Analysis of Variance, quantifies numerous measurement error sources independently (Brennan, 2001a). One guiding principle is that a teacher observes scoring in the sum of their real score and measurement error. Generalizability Theory assumes error causes may be deconstructed and quantified independently, unlike Classical Test Theory. Multivariate G theory analysis also provides a more detailed knowledge of measuring instrument component dependability. Due to different mistake and real score estimations.

Thus, Generalizability Theory can identify and compute error sources in a measurement procedure. To the extent that Generalizability Theory "is particularly well suited to evaluating assessments that are based on ratings of human performance" (Brennan, 2001a, p. 117), This analytical framework is widely used in psychological and educational research to study rating-based measuring processes. The Generalizability Theory framework has been used in language testing and assessment research to analyze the generalizability of diverse speaking tasks (Han, 2018), offer reliability indication for rating scales (Ohta et al., 2018).

Generalizability Theory basically uses two quantitative methods. Generalizability studies focus on acceptable observation universes. Persons or test-takers are the entity of dimension, an entity of interest to the researcher, and a set of measuring circumstances is an aspect. Generalizability Theory analyses. Thus, choice studies help practitioners and academics make educated test building, development, and validation decisions.

Data Collection and Analysis

The author of this study teaches English to non-English majors in Iraqi higher education as part of their general education. He gave his English reading pupils the Foreign Language Classroom Anxiety Scale. He gave pupils an internet access to the Foreign Language Classroom Anxiety Scale to examine how they may complete it freely. To minimize confusion, some words and expressions (e.g., singled out, at ease) were translated to Arabic. Therefore, generalizing across items and contexts is crucial to self-reported data score consistency. Survey data, like testing data, seeks substantial person-score variation, while minor variance components owing to other factors imply scale reliability.

Findings

For fear of undesirably articulated items creating a single factor, the psych package is used (Revelle, 2020) to do an exploratory factor analysis. Positively phrased things were the first factor and undesirably articulated ones were the second factor. In the light of this assumption, all nine negatively phrased questions for Generalizability Theory analysis are removed to reduce anxieties regarding scale reliability. Given the emergence of a putative, fake factor entirely consisting of undesirably articulated items. For Generalizability Theory analysis as originally intended, which examines communication apprehension, test anxiety, and fear of negative evaluation.

Univariate Generalizability Theory Analyses

Variance components are shown for the two univariate Generalizability Theory analyses. The variability distributions were similar on both instances. According to Shavelson & Webb (1991), this

study examines the relative difference between construct-related items. The variability of mean item scores relative to real score variation was larger. Foreign Language Classroom Anxiety Scale at multiple times. Remember that generalizability. Theory analysis using two data sets from distinct occasions. Like separate univariate studies, the combined analysis indicated that the largest interaction term and residual errors contributed most to measurement error. Large three-way interaction term undifferentiated with unexplained error is expected "the largest estimated variance component is often the one associated with the highest order interaction" (Brennan, 2001a, p. 83).

Three univariate Generalizability Theory investigations showed how analytic design affects data reliability conclusions. When the event facet was used for assessing Foreign Language Classroom Anxiety Scale reliability, generalizability and dependability coefficients declined significantly. Thus, analytic model choice considerably influenced FLCAS reliability coefficient estimations. Adding the event component significantly reduced generalizability and dependability coefficients. Traditional reliability measures like Cronbach's alpha may exaggerate scale dependability. Generalizability Theory studies do not use significance tests (Brennan, 2001a), but standard acceptability cutoffs should help explain these findings.

Generalizability and the Components of the FLCAS

The appropriate item and occasion count for the Foreign Language Classroom Anxiety Scale were determined by computing generalizability and dependability coefficients in follow-up Decision trials. The default Decision study, which employed same quantities of items and situations as the Generalizability Theory study, calculated the generalizability of composite scores. Estimating past estimations of Foreign Language Classroom Anxiety Scale. A detailed look at each latent attribute showed varied score consistency distributions. Communication apprehension and fear of bad appraisal had reliable. Based on the Generalizability research results, we set the Foreign Language Classroom Anxiety Scale's total number of items at 24 for practicality in using it in diverse situations.

Discussion

Generalizability Theory decomposes undifferentiated error into multiple variance components, making it a more cautious reliability estimate than standard estimates. This research examined the "what extent" question using FLCAS data from college-level EFL students with L1 Arabic backgrounds. The study's main goals were to determine how much Generalizability Theory overestimates the scale's classroom anxiety and how many items are needed to obtain a sufficient degree of reliability for each latent characteristic.

Difference between the Foreign Language Classroom Anxiety Scales

The Foreign Language Classroom Anxiety Scale in the multivariate Generalizability Theory study. A detailed look at each latent feature showed that communication apprehension scored substantially complex on reliabilities than fear of negative assessment and exam anxiety.

Test anxiety was considerably lower than the commonly used reliability for generalizability and dependability. These estimations are substantially lesser than Aydin et al. (2021), Huang & Hung (2013), and Salehi and Marefat (2014) reliability coefficients. The Foreign Language Test Anxiety Scale (FLTAS). The study's poor reliability-like of test anxiety may be related to a small sample size and survey administration. Huang & Hung (2013) and Salehi and Marefat (2014) used their measures with test tasks to evaluate test anxiety and score relationships. In contrast, the current research used the Foreign Language Classroom Anxiety Scale alone for data collection. Given this,

increasing the amount of questions in the scale and administering it with an evaluation to better elicit test-related worries may help improve its reliability.

However, the current study showed that for communication apprehension, exam anxiety, and fear of unfavorable assessment may be most reliable. This pattern seems to be a major change from the original elements per sub-scale. Test anxiety reliability increased and this sub-scale improved in reliability as the number of items rose.

Theoretical and Practical Implications

The outcomes of this study highlight a key qualitative research concept: validity requires reliability. In the four decades since the Foreign Language Classroom Anxiety Scale was created, several validity claims have been made, and their reliability estimates. A self-reporting instrument like the Foreign Language Classroom Anxiety Scale has more measurement error than Cronbach's alpha can capture, thus this research used Generalizability Theory to get more accurate reliability coefficients. The Foreign Language Classroom Anxiety Scale may be less reliable than previous studies, so researchers should be cautious when making validity claims about its predictive ability and underlying constructs.

Results provide light on the Foreign Language Classroom Anxiety Scale's dependability, although multi-occasion study methods may not be suitable for most second/foreign language educational environments. Teachers utilize their limited resources wisely, and time restrictions may prevent them from administering the Foreign Language Classroom Anxiety Scale many times. Based on our results, instructors should be aware that a learner's answer to the Foreign Language Classroom Anxiety Scale may vary and may have been influenced by classroom activities that cause anxiety. Thus, classroom FLA interpretations should be evaluated often. Thus, FLCAS replies should not diagnose FLA in language learners in the classroom.

Limitations and Future Research

One dataset in this investigation was susceptible to technique bias, as is typical with self-reporting data. That is, FLCAS items are distinguished by word valence rather than theoretical and substantive substance. Therefore, we had to eliminate all negatively stated questions from further analysis. This reduced the amount of items for G theory analysis, which may have impacted reliability for one of the latent qualities, test anxiety, as test duration is widely thought to affect reliability (Thorndike & Thorndike-Christ, 2009). Future study should identify respondents' cognitive processes while using self-report instruments like the FLCAS, especially when processing reverse coded items. The significant value from the three-way interaction suggests that respondents gave inconsistent ratings or that unexplained error influenced reliability measurements.

Scale delivery changed from out-of-class to in-class, affecting the study design. Concerns concerning greater response set incidence than expected in the first data set validated our conclusion. However, this may have raised event variance components like item-event two-way interaction. That is, minimizing the negative impacts of answer sets on scale reliability resulted in a rather large variance component associated with occasions. However, the relative impacts of lowering response sets vs. change in occasion circumstances on scale reliability are unclear.

Conclusion

The findings of the current research indicate that the Foreign Language Classroom Anxiety Scale exhibited variability across two different occasions. Furthermore, when the occasion facet was accounted for, there was a notable decrease in the reliability coefficients. In light of the aforementioned, the present research posited the significance of meticulous reliability evaluations on

assessments, specifically focusing on self-report measures. The results also indicated good stages of complex reliability, irrespective of the lesser levels seen in relation to one of the latent qualities, namely test anxiety. It has been proposed that in order to improve the reliability evidence for the Foreign Language Classroom Anxiety Scale, it is required to consider two steps: increasing the amount of items related to the latent trait and/or conducting the Foreign Language Classroom Anxiety Scale alongside a specific test task. In conclusion, it is important to thoroughly examine the dependability of the Foreign Language Classroom Anxiety Scale while assessing the broadening range of validity evidence.

References

1. Adamson, B. & Morris, P. (1997). Focus on curriculum change in China and Hong Kong: The English Curriculum in the People's Republic of China. Comparative Education Review, 41, 3-26.
2. Aydın, S., Denkci-Akka ş, Türnük, T., Baş türk-Beydilli, A., & Saydam, İ (2021). A descriptive study on text anxiety among foreign language learners. Futuristic Implementations of Research in Education (FIRE), 2 (2), 43–55.
3. Bolton, K. & Tong, Q. S. (2002). Introduction: Interdisciplinary perspectives on English in China. World Englishes, 21 (2), 177- 180.
4. Brennan, R. L. (2000). (Mis)conceptions about generalizability theory. Educational Measurement: Issues and Practice, 19(1), 5–10.
5. Brennan, R. L. (2001a). Generalizability Theory. New York: Springer-Verlag Publishing.
6. Brennan, R. L. (2001b). Manual for mGENOVA (Ver. 2. 1). Iowa City: Iowa Testing Programs. University of Iowa.
7. Campbell, C. M., & Shaw, V. M. (1994). Language anxiety and gender differences in adult second language learners: Exploring the relationship.
8. Casado, M. A., & Dereshiwsky, M. I. (2001). Foreign language anxiety of university students. College Student Journal, 35(4), 539-552.
9. Cheng, Y. S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. Language learning, 49(3), 417-446.
10. 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.
11. Crick, J. E., & Brennan, R. L. (1983). Manual for GENOVA (Ver. 2. 1) . Iowa City: Iowa Testing Programs, University of Iowa.
12. Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. British Journal of Statistical Psychology, 16, 137–163.
13. Ellis, R. (1994). The Study of Second Language Acquisition. Oxford, ngland: Oxford University Press.
14. Fallah, N. (2017). Mindfulness, coping self-efficacy and foreign language anxiety: A mediation analysis. Educational Psychology, 37 (6), 745–756.
15. Gleser, G. C., Green, B. L., & Winget, C. N. (1978). Quantifying interview data on psychic impairment of disaster survivors. Journal of Nervous and Mental Disease, 166(3), 209–216.
16. Han, C. (2018). A generalizability theory study of optimal measurement design for a summative assessment of English/Chinese consecutive interpreting. Language Testing, 36 (3), 1–20.

17. Horwitz, E. K., Horwitz, M. B. & Cope, J. A. (1991). Foreign Language Classroom Anxiety. In E. K. Horwitz & D. J. Young (Eds.), Language Anxiety: From Theory and Research to Classroom Implications. 27-36. Englewood Cliffs, NJ: Prentice Hall.
18. Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. The Modern language journal, 70(2), 125-132.
19. 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.
20. Krashen, S. D. (1985a). Applications of psycholinguistic research to the classroom. In C. James (Ed.), Practical applications of research in foreign language teaching. 51-66. National Textbook Co: Lincolnwood, Illinois.
21. Krashen, S. D. (1985b). The Input Hypothesis: Issues and Implications. Longman: New York.
22. Lam, W. (2005). Successful knowledge management requires a knowledge culture: a case study. Knowledge management research & practice, 3, 206-217.
23. Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. The Modern Language Journal, 92 (1), 71–86.
24. MacIntyre, P. D, & Gardner, R. C. (1991). Methods and results in the study of anxiety and language learning: A review of the literature. Language Learning, 41, 85-117.
25. MacIntyre, P. D. (1995). How does anxiety affect second language learning? A reply to Sparks and Ganschow. The modern language journal, 79(1), 90-99.
26. MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second- language learning: Toward a theoretical clarification. Language learning, 39(2), 251-275.
27. MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. Language Learning, 44, 283–305.
28. MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self- ratings of second language proficiency: The role of language anxiety. Language learning, 47(2), 265-287.
29. Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. Language Learning, 50, 417–528.
30. Ohata, K. (2005). Potential sources of anxiety for Japanese learners of English: Preliminary case interviews with five Japanese college students in the US. TESL-EJ, 9(3), n3.
31. Ohta, R., Plakans, L., & Gebril, A. (2018). Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. Assessing Writing, 38, 21–36. 10.1016/j.asw.2018.08.001.
32. Onwuegbuzie, A. J., Bailey, P. & Daley, C. E. (2000). The validation of three scales measuring anxiety at different stages of the foreign language learning process: The input anxiety scale, the processing anxiety scale, and the output anxiety scale. Language Learning, 50 (1), 87-117.
33. Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. The Modern Language Journal, 100 (2), 538–553.
34. Purpura, J. E., Brown, J. D., & Schoonen, R. (2015). Improving the validity of quantitative measures in applied linguistics research. Language Learning, 65 (1), 37–75.
35. Revelle, W. (2020). Psych: procedures for personality and psychological research. Evanston: Northwestern University https://CRAN.r-project.org/package = psych R package version 2.0.8.

36. Salehi, M., & Marefat, F. (2014). The effects of foreign language anxiety and test anxiety on foreign language test performance. Theory and Practice in Language Studies, 4 (5), 931–940.
37. Shavelson, R. J., & Webb, N. M. (1991). Generalizability theory: A primer. Thousand Oaks, CA: Sage.
38. Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. Studies in Second Language Acquisition, 41, 363–387.
39. Thorndike, R. M., & Thorndike-Christ, T (2009). Measurement and evaluation in psychology and education. Boston: Prentice Hall.
40. Trylong, V. L. (1987). Aptitude, attitudes, and anxiety: A study of their relationships to achievement in the foreign language classroom. Purdue University.
41. von Worde, R. (2003). Students' perspective of foreign language anxiety. Inquiry, 8 (1). Retrieved on April 16, 2006, from http:// www.vccaedu.org/inquiry/inquiry-spring2003/i-81-worde.html.
42. Webb, N. M., Shavelson, R. J., & Haertel, E. H. (2006). Reliability coefficients and generalizability theory. In C. R. Rao & S. Sinharay (Eds.), Handbook of statistics (Vol. 2, pp. 81–124). Amsterdam: Elsevier.
43. Woodrow, L. (2014). Writing about quantitative research in applied linguistics. New York, NY: Palgrave Macmillan.