The effects of modes of test administration on test anxiety and test scores: A study in an Indonesian school

Sahlan, Alberth, Wahyudin Madil
Halu Oleo University, Indonesia
Hutnisyawati
SMPN 12 Konawe Selatan, Indonesia

Online testing is now becoming more prevalent internationally, across different educational levels. However, little is known about the impact of this mode of testing on male and female students' anxiety levels and test performance. Conducted in an Indonesian sociocultural context which is under-represented in the international literature, this paper presents results from a study examining the effect of two different modes of testing (online and paper-and-pencil tests) on test takers' anxiety levels and test scores. A total of 243 (99 males, 144 females) junior high school students were recruited. These students were assigned to either paper-and-pencil or online testing. The PHCC test anxiety questionnaire was administered immediately after the test. A 2 x 2 factorial ANOVA revealed that neither gender, mode of test administration nor the interaction between gender and mode of testing impacted on students' anxiety levels. The main effect of gender and the interaction effect between gender and mode of testing on test scores were significant. Students who experienced high anxiety levels performed significantly poorer than those who had medium or low anxiety levels.

Introduction

Advances in computer and web-based technologies have enabled the administration of online testing (hereafter OT) (Fageeh, 2015), which is becoming generally considered to be superior to conventional on-paper (hereafter OP) tests (DeSouza & Fleming, 2003). For example, web-based quizzes provide immediate feedback and can be administered simultaneously to very many students, with relatively little time required for grading (Sidhu, 2008). Web-based tests are highly efficient as they can be machine scored (Meo & Martí-Ballester, 2020), which is highly economical (Fageeh, 2015). Therfore it is not surprising that computer-assisted testing is now gaining immense popularity across the various educational levels.

In spite of this increasing popularity, relatively few studies have examined the effect of mode of testing on both performance and anxiety, particularly in the context of the present study. Understanding these issues is critical, for test formats could potentially impinge on student motivation to do the test, student anxiety level, and test performance. However, there has been relatively little research on how different modes of assessment (OP versus OT) are related to test takers' anxiety levels, although the interconnectivity between anxiety levels and academic performance is well documented in the literature (Chin, Williams, Taylor & Harvey, 2017).

Whereas a number of studies attempting to compare the viability of OP and OT are available, these studies yielded conflicting results (Choi, Kim & Boo, 2003; Macedo-Rouet,

Ney, Charles & Lallich-Boidin, 2009), blurring the actual merit of OT relative to its counterpart OP assessment. Additionally, where studies comparing OP and OT tests do exist, the majority of these studies focus primarily on the issue of comparability in terms of test scores obtained by test takers (Pommerich, 2004) and perceived difficulty (Way, Davis & Fitzpatrick, 2006). Little research, has endeavoured to examine test takers' levels of anxiety (for both males and females) when sitting tests of different formats. Needless to say, whereas comparability of test scores is crucial in an attempt to understand the merit of OT relative to OP testing, we argue that test takers' levels of anxiety is also critical if we are to understand the merit of computer-assisted testing. The study reported here sought to contribute to the discussion pertaining to the merit of OT relative to conventional OP assessment by providing empirical evidence gathered from a sociocultural context which is under-represented in the international literature.

Literature review

This section examines research comparing OP and OT and identifying discrepancies in this line of research and moves on to a discussion of test anxiety and gender differences.

Online versus offline test

DeSouza and Fleming (2003) compared two groups of students in a conventional classroom, one required to take online mastery quizzes using *Mallard* (an asynchronous web-based assessment program) and the other required to take OP quizzes administered in the classroom. Quiz questions were of different versions, but they were all consistent with materials tested in the examination. The authors claimed that the study design eliminated the effect of mode of instruction, making it possible to examine the effect of online mastery quizzes on student learning. The study found that Mallard group taking online quizzes significantly outperformed students taking OP quizzes, leading to the conclusion that online quizzes enhance student learning more effectively than OP quizzes. However, this difference in student performance could simply be attributed to the fact that Mallard students are exposed more frequently to the quizzes compared to their counterparts, the conventional group, rather than to the quiz format itself. Thus, to examine the real effect of the quiz format, both groups should be provided with the same amount of time and with the same number of quiz questions.

Fynewever (2008) compared the effectiveness of web-based and paper-based homework for general chemistry learning. In this study, students were divided into two cohorts; group one took web-based quizzes and the other group sat OP quizzes. A pre-test was administered to both groups so that improvements in student post-test scores could be analysed. It was reported that the gain scores of students in both cohorts did not differ significantly, which in turn suggested that both quiz formats are equally effective as far as student learning is concerned. However, it was reported that both quiz formats were appreciated by these students. Two advantages of web-based quizzes reported by the students include the provision of immediate feedback and the ability to take the quiz over and over again.

Karadeniz (2009) examined the impacts of three different types of quiz: paper, web and mobile-based quizzes on student learning in Internet-assisted instruction, as well as their perception of these three different formats of quiz. Using a 2*3 factorial design, the study concluded that the learning outcomes of the students taking paper, web and mobile-based quizzes did not differ significantly from the learning outcomes of those taking OP quizzes. Additionally, participants also appreciated mobile and web-based quizzes, primarily owing to their simplicity and to the provision of immediate feedback. Interestingly, when asked to rate the extent of their preference over the three quiz formats, web-based quizzes were the most favoured, followed by mobile and paper-based quizzes. However, this study did not examine students' anxiety level while taking the test. Nor did it compare the anxiety levels of test takers based on gender across different test formats.

Furthermore, Macedo-Rouet et al. (2009) compared the performance of students studying printed lecture notes and taking paper-based quizzes, with the performance of those studying on screen and taking web-based quizzes for an introductory course on mathematical tools for the life sciences. The study found that the former group (the one studying printed materials and taking OP quizzes) outperformed the latter (the one studying on screen and taking web-based quizzes) in terms of their final test scores. Interestingly, students studying printed materials also indicated quite clearly that they would actually prefer web-based tests. However, since differences in the mode of presentation and testing mode are confounded, it is not possible to conclude whether it was mode of presentation, mode of testing, or an interaction between them that caused any observed differences in performance.

In a nutshell, the majority of research studies discussed above focused primarily on students' performance, as indicated by test scores, as a result of sitting a test administered in different formats, not on test takers' anxiety levels when taking such a test. As pointed out earlier, understanding how different modes of testing may impact on test takers' anxiety levels and test scores for both males and females is critical as abundant research evidence demonstrates that a high level of anxiety is detrimental to students' academic performance (e.g. Lowe, 2019; Winke & Lim, 2017) and their social life (Aydin, 2019). The need to examine this issue is becoming more imperative given the omnipresence of OT across different educational levels.

Test anxiety

Research on the interconnectivity between emotion and test performance is well documented (Pekrun et al., 2004). Among the widely investigated emotional constructs in testing-related research is anxiety. Defined as "a set of cognitive, physiological, and behavioral responses related to concerns about possible failure or a poor performance on a test or a similar evaluative situation…" (Bodas, Ollendick & Sovani, 2008 p. 387), test anxiety is considered to be a critical psychological construct. Test anxiety, also referred to as "exam anxiety, exam stress or test stress", is considered to be different from other types of anxiety as it relates to testing situations (von der Embse, Jester, Roy & Post, 2018,

p. 483). Test anxiety has long been reported to be a serious problem in education as it may hamper academic performance (Lowe, 2019).

Test anxiety levels increase when students are required to take a test or when their abilities are evaluated (von der Embse et al., 2018). It may also occur when test takers are worried about the negative consequences resulted from test results (von der Embse et al., 2018), such as failing the test. In other words, test anxiety results from a physical and psychological reaction to potential failure on test (Ahmad, Hussain & Khan, 2018). Evidence suggests that test anxiety may impact on not only academic outcomes and motivation of the students, but also their social life (Aydin, 2019). Thus, understanding students' anxiety levels while sitting a test is of immense importance (Soares & Woods, 2020).

Although a high anxiety level, also referred to as debilitating anxiety, is detrimental to learning (Lowe, 2019; Aydin, 2019; Alberth, 2019) it is important to acknowledge the existence of positive anxiety referred to as facilitating anxiety. Research studies have suggested that facilitating anxiety could motivate students to learn and work harder (Park & French, 2013; Arifin, Mursalim & Sahlan, 2020). Thus, in a sense, anxiety could serve as a booster that motivates learners to exert more effort and energy in learning, thus achieving higher performance.

Little research has examined test anxiety with respect to individual differences (Aydin, 2019). In particular, how mode of test administration may actually impact on test anxiety levels and test performance of both male and female students remains an open question. Whereas some work has been conducted for conventional OP testing, there is relatively scarce research comparing anxiety levels of students sitting a test administered in two different modes of testing, taking gender and learning outcomes into account. Woldeab and Brothen (2019) examined test anxiety in an online proctored environment and found that low performing students exhibited a high anxiety level. In fact, research studies show that students who have low anxiety levels tend to focus more on the task at hand, whereas students with high anxiety levels are less focused, thus performing poorer than those who are less anxious (Aydin, 2019). However, test anxiety is in part a function of expected performance. So low-performing students are expected to be more anxious, irrespective of the mode of administration. Thus, in order to examine the effect of mode of test administration on anxiety, student prior achievement needs to be controlled so that the effect of mode can be demonstrated clearly.

Wren and Benson (2004) identified three important components or dimensions associated with children's test anxiety namely (a) thoughts, (b) autonomic reactions, and off-task behaviours (c). The first dimension, 'thoughts', is related to cognitive anxiety, that is, test-takers' worry and concern about their own performance on the test and this worry involves internal dialogue and evaluation of one's own performance. The second dimension, 'autonomic reactions', concerns somatic responses such as sweating and stomach problems. The third dimension, 'off-task behaviour', refers to behavioural anxiety which can be observed by such instances as students playing with hair (referred to

as auto-manipulation), playing with pen (referred to as object manipulation) and distracted/ unfocused attention.

Similarly, echoing the literature, Stöber (2004) identified four dimensions of test anxiety referred to as worry, emotionality, interference, and lack of confidence. The term 'worry' refers to feeling anxious due to possible failure along with its negative consequences, whereas 'emotionality' refers to involuntary reactions brought about by test situation. The term 'interference', refers to distracting cognitions which interrupt performance, whereas 'lack of confidence' refers to the minimum confidence of test takers.

To sum up, debilitating test anxiety (a very high anxiety level), which may also indicate low expectations for performance, is detrimental to students' test performance. Moderate anxiety level, however, may have a positive effect on test performance. It is, therefore, critical to scrutinise the impact of different modes of test administration on male and female students' anxiety levels, in addition to their test performance.

The advent of online testing, also known as web-based testing or computer-assisted testing, necessitates meticulous investigations into the effect of the new mode of test administration, relative to conventional OP testing, on male and female test anxiety levels and their learning outcomes.

Gender differences

Research in the realm of test anxiety has been extended to investigating the extent to which male and female students feel anxious when sitting OP test. In fact, there is a plethora of research studies which have examined anxiety levels by gender in a face-to-face classroom setting, but these studies have reported conflicting findings, resulting in the three different conflicting research findings. The first kind of finding suggests that female students are more anxious than male students (Cotner, Jeno, Walker, Jørgensen & Vandvik, 2020; Lowe & Lee, 2008; Stöber, 2004). The second kind, however, concludes that male students are more anxious than female students (Celik, 2019), whereas the third kind reveals insignificant differences (Ahmad et al., 2018).

Research studies have also compared the academic achievement of male and female students taking a conventional OP test and suggested that females outperformed male students (Kamphorst, Adriaan Hofman, Jansen & Terlouw, 2015; Workman & Heyder, 2020; Duckworth & Seligman, 2006; Alfarhan & Dauletova, 2019; Conger & Long, 2010). A meta-analysis conducted by Voyer and Voyer (2014) concluded that, whereas female students generally (not always) outperform their male counterparts academically in various courses such as science and math, this trend is more strongly pronounced in language-related courses (Workman & Heyder, 2020).

However, a number of shortcomings are found in previous studies. First, research comparing male and female learning outcomes is conducted in a single mode of testing, either in a conventional OP (Gorjian & Javadifar, 2013) or in an OT test (Kaarakainen, Kivinen & Kaarakainen, 2017). Whilst a few research studies have concerned both modes

of testing (Macedo-Rouet, Ney, Charles & Lallich-Boidin, 2009; DeSouza and Fleming, 2003; Fynewever, 2008), these focused almost exclusively on students' learning outcomes as indicated by test scores. Little research has explored possible differences in male and female students' anxiety levels, in addition to learning outcomes, when the test is administered in two different test formats, OP and OT. The present study sought to fill this gap in research by examining the effect of different modes of test administration on male and female students' anxiety levels and test performance.

Research questions

The present study was guided by the following research questions:

1. Does gender have a significant effect on test anxiety and test scores?
2. Does mode of test (OT versus OP) have a significant effect on test anxiety level and test scores?
3. Does interaction between gender and mode of assessment have a significant effect on test anxiety level and test scores?
4. Do significant differences exist in test scores for students who experience low, mild and high anxiety levels?

Method

Participants

This study was conducted with 243 high school students (99 males, 144 females) in Kendari (SMPN 2 Kendari), South-east Sulawesi, Indonesia. The ages of the participants ranged from 11-16 years of age. These students were recruited using convenience sampling and consent from these participants was sought prior to the study. Participants were free to transfer to their preferred mode of testing, but none requested so. The number of participants in each mode along with their genders is presented in Table 1.

Table 1: Participants

Mode of testing	Gender	N	%
OP (on paper)	Male	39	36.1
	Female	69	63.9
OT (online testing)	Male	60	44
	Female	75	56
Total		243	100

Note that there was an imbalance in the propotions of males and females assigned to the two modes owing to the fact that females in this particular school far outnumbered males. This trend was reflected in the proportions of males and females assigned in these groups.

Participants of the present study were very familiar with the new technology and had access to the Internet both from their own mobile phones and from the school computer

lab. Thus, there is no question about technology literacy, as well as access to computers and mobile phone devices. This is understandable since the school involved in this study is located in an urban area where technology becomes part of people's lives and this is particularly true for the digital natives like those participating in the present study.

Instrumentation

The present study adopted the *PHCC Test Anxiety Questionnaire* developed by Nist and Diehl (1990). Encompassing 10 items, the questionnaire was developed to examine the extent to which a student might experience test anxiety. Student indicate how often each statement describes him/her: 1 - Never; 2 - Rarely; 3 - Sometimes; 4 - Often; 5 - Always. Thus, the minimum total score would be 10 and the maximum would be 50. A score within a range of 10-19 indicates that the respondent does not suffer from test anxiety. A score ranging from 20-35 suggests mild, but healthy, anxiety, whereas a score of higher than 35 indicates the presence of debilitating anxiety (Nist & Diehl, 1990). The instrument was reported to be both valid and reliable. The Cronbach's alpha value was 0.9 (Ogundokun, 2011). For the sake of convenience and clarity, all questionnaire items were translated into Indonesian.

Additionally, the academic achievement test was developed by the teacher. The test was designed to measure general English language proficiency of the third semester of junior high school students. Encompassing 40 multiple choice items, the test measured the language components listening comprehension, reading comprehension, expressions and vocabulary and grammar. The weight of each test item varied from 1-3 points with more difficult items having more weight. Note that this test was not validated in the present study. However, it was previously validated using a split-half method involving 50 students from a similar population. The reliability of the test, as indicated by the correlation between test scores (correlation between forms) equaled 0.75, which indicated an acceptable level of reliability.

The validity of the test was examined by a pannel of three English teachers who rated the 40 test items as either relevant or irrelevant (1 point for relevant and 0 for irrelevant items). The relevance of each of these test items was judged against the objective of the test. *Item Content Validity Index* (I-CVI) and *Scale Content Validity Index* (S-CVI) were then calculated based on the scores provided by the three teachers acting as the judges (DeVellis, 2003). I-CVI was obtained by adding total scores for a given item divided by three. So, if all teachers regarded test item #1 to be relevant, for example, then the I-CVI for that item was 1 (1+1+1/3). Items whose I-CVIs were below 0.6 were removed. To calculate S-CVI, total I-CVI scores were divided by total number of items.

The lowest I-CVI was found to be 0.67 (when one out of the three teachers considered a given item to be irrelevant) and the highest was 1 (when all three teachers agreed that a given test item was relevant). The S-CVI, which indicated the overall validity of all test items, was reported to be 0.8, which indicted an acceptable level of validity. Thus, there was some evidence that the academic achievement test developed by the teacher was both valid and reliable and its use was, therefore, justified.

Research procedure

Students sitting the test were divided into two cohorts. Group one sat OP test, whereas group two sat OT test. To this end, 108 students sat OP (39 males, 69 females) and 135 sat OT (60 males, 75 females) tests. Test items were identical and the same amount of time (60 minutes) was provided for both groups.

Data collection

The PHCC Test Anxiety Questionnaire, designed using *Google Forms*, was administered to both groups of students immediately after the summative English test. These students were given 10 – 30 minutes to complete the questionnaire from their mobile phones. Test results for the OT group became available immediately after the test, whereas for the OP group, it took some time since the marking was done by hand by the third author and the class teacher.

Data analysis

Results of the questionnaire and test scores were subsequently entered into the *SPSS* computer software program version 20 for data analysis by the fourth author under the auspices of the second and the first author. The present study employed a 2 x 2 between-subjects factorial ANOVA design. All classical assumptions were carefully checked prior to the analysis to ensure that none of the assumptions was violated. Inspection of the classical assumptions of factorial ANOVA suggested that all assumptions were met.

Findings

Anxiety mean scores by gender and mode of test administration are presented in a 2x2 table (Table 2) along with the standard deviations and number of cases:

Table 2: Anxiety Scores by gender and mode of testing

Gender	Mode of testing	Mean	SD	n
Male	OP (on paper)	2.04	.62	39
	OT (online testing)	2.22	.66	60
	Total	2.15	.65	99
Female	OP (on paper)	2.23	.67	69
	OT (online testing)	2.22	.49	75
	Total	2.22	.59	144
Total	OP (on paper)	2.15	.66	108
	OT (online testing)	2.22	.57	135
	Total N	2.19	.61	243

As stated earlier, the present study was aimed at examining (a) whether gender has a significant effect on test anxiety and test scores; (b) whether mode of testing (OT versus OP) has a significant effect on test anxiety level and test scores; (c) whether the interaction

between gender and mode of assessment has a significant effect on test anxiety level and test scores; and (d) whether test scores differ according to the levels of anxiety. To answer questions (a), (b) and (c), a 2 (gender) x 2 (mode) factorial ANOVA was performed on students' anxiety scores. However, since factorial ANOVA belongs to parametric statistics, inspection of the classical assumptions of the test is required. This includes the assumption of independence of observation, normal distribution of standardised residual and homogeneity of variance. None of these assumptions was violated and the use of a 2 x 2 factorial ANOVA was therefore justified.

A 2 (gender) x 2 (mode) between-subjects factorial ANOVA was calculated comparing test anxiety scores for participants sitting one of two modes and who were either male or female. There was a non-significant main effect of gender on anxiety, $F(1,239) = 0.82$, $p = .37$, partial $\eta^2 = .003$, observed power = 0.15. A non-significant main effect for modes of testing was also found $F(1,239) = 0.99$, $p = 0.32$, partial $\eta^2 = .004$, observed power = .67. This, in turn, suggests that students who sat OT had a similar anxiety level (M= 2.22, SD = .57240) with that of those who sat OP test (M = 2.16, SD = .66). Finally, the interaction effect between gender and mode of testing on test anxiety was also not significant $F(1,239) = 0.90$, $p = 0.34$, partial $\eta^2 = 0.004$, observed power = 0.16. The 95% confidence intervals of anxiety scores for OP and OT are presented in Table 3.

Table 3: Confidence intervals of anxiety scores for both OP and OT

Variable	Levels	Mean	95% confidence interval	
			Lower bound	Upper bound
Anxiety scores	OP	2.16	2.03	2.29
	OT	2.22	2.12	2.32

The interaction effect between gender and mode of testing on anxiety scores is depicted in the interaction plot in Figure 1.

The above findings led to the conclusion that students' anxiety levels were affected by neither gender, modes of testing, nor by the interaction between gender and modes of testing. Thus, it can be concluded that the anxiety level of all participants, regardless of mode of testing (OP or OT) and gender (Male or Female), is similar.

The next question to be examined is whether mode of testing, gender, or the interaction between mode of testing and gender impact on student test scores. Test scores by gender and mode are presented in the 2x2 table (Table 4) along with the standard deviation and number of cases.

In order to understand the effect of gender, modes of testing and interaction between gender and modes of testing on test scores, a 2 (gender) x 2 (mode) between-subjects factorial ANOVA was calculated comparing test scores for participants who sat either OT or OP and who were either male or female. There was a significant main effect of gender on test scores, $F(1,236) = 30.077$, $p = 0.000$, partial $\eta^2 = 0.113$, observed power = 1.000.
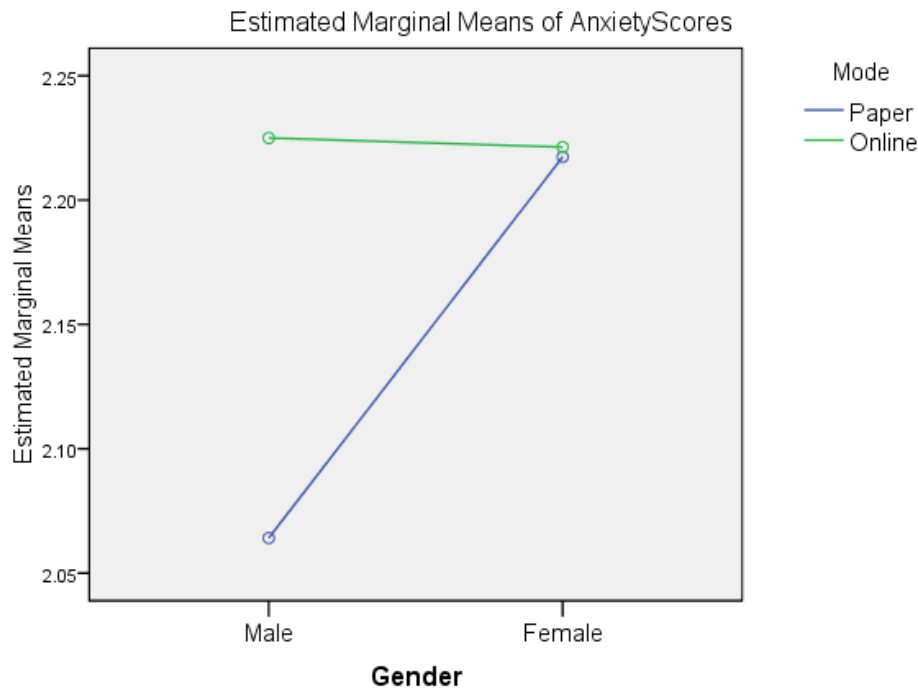
Figure 1: Interaction plot for Gender*Mode>Anxiety

Table 4: Test Scores by gender and mode

Gender	Mode	Mean	Std. dev.	No. cases
Male	Paper	77.59	6.42	39
	Online	74.37	5.46	60
	Total	75.64	6.04	99
Female	Paper	79.91	6.40	67
	Online	80.73	5.65	74
	Total	80.34	6.01	141
Total	Paper	79.06	6.48	106
	Online	77.88	6.39	134
	Total	78.40	6.44	240

Female test scores OT (M=80.73, SD=5.65) and OP (M=79.91, SD=6.40) were higher than those of males OT regardless of mode of testing (M=74.37, SD=5.46) and OP (M=77.59, SD=6.42). A non-significant main effect for the mode of testing was found $F(1,236) = 2.305$, $p = 0.130$, partial $\eta^2 = 0.010$, observed power = 0.327. The interaction effect between gender and mode of testing on test scores was significant $F(1,239) = 6.518$, $p = 0.011$, partial $\eta^2 = 0.027$, observed power = 0.720. However, this effect size is very small; so, even though the finding is statistically significant, it is of little practical significance. The 95% confidence intervals of test scores for both OP and OT are presented in Table 5.

Table 5: Confidence intervals of test scores gender*mode

Variable	Gender	Mode	Mean	95% confidence interval	
				Lower bound	Upper bound
Test scores	Male	Paper	77.59	75.71	79.47
		Online	74.37	72.85	75.88
	Female	Paper	79.91	78.48	81.34
		Online	80.73	79.37	82.09

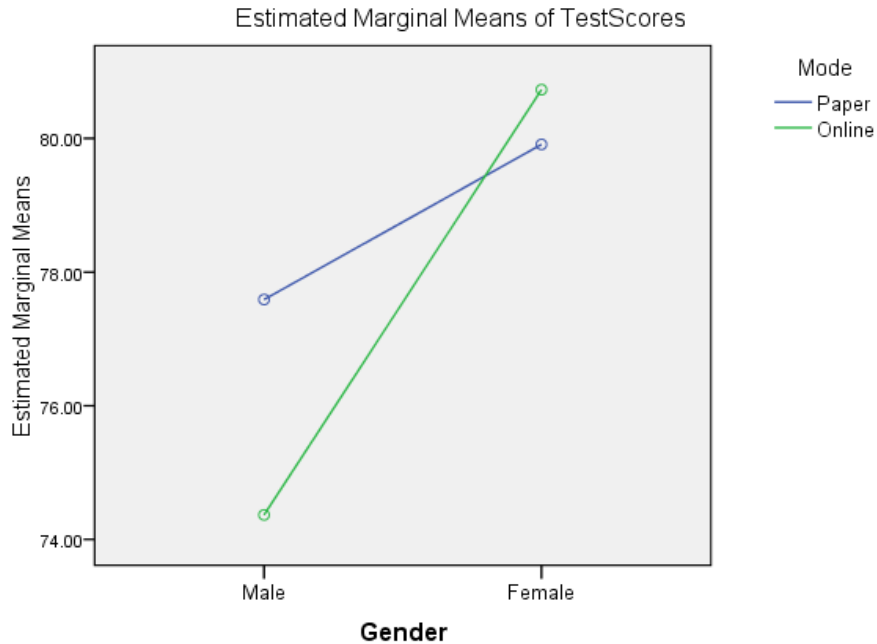The interaction effect between gender and mode of testing on test scores is depicted in an interaction plot (Figure 2).

Figure 2: Interaction plot for gender*mode>test

Finally, a one-way between subjects ANOVA was conducted to compare the effect of different anxiety levels (low, medium, and high) proposed by Nist and Diehl (1990) on test scores. ANOVA assumptions were carefully checked to ensure that none was violated. There was a significant effect of level of anxiety on test scores at the $p < .05$ for the three different anxiety levels, $F(2,237)=3.911$, $p = 0.021$. Post hoc comparisons using Bonferroni indicated that test scores of students who experienced a medium anxiety level (M= 78.83, SD = 6.64) was significantly higher than test scores of students who experienced a high anxiety level (M= 71.50, SD = 4.04), p (0.019) < 0.05. Test scores of students who experienced a low anxiety level (M=78.17, SD = 6.006) was also significantly higher than test scores of students who experienced a high anxiety level (M=71.50, SD=4.04), p (0.041 < 0.05). Although test scores of students who experienced

a medium anxiety level was slightly higher than test scores of students who experienced a low anxiety level, the difference was not statistically significant.

The 95% confidence intervals of test scores based on the three levels of anxiety (low, medium, high) are presented in Table 6.

Table 6: Confidence intervals of test scores based on anxiety levels

Variable	Levels	Mean	95% confidence interval	
			Lower bound	Upper bound
Test scores	Low	78.17	76.90	79.43
	Medium	78.83	77.74	79.92
	High	71.50	67.26	75.74

Discussion

The present study suggested that neither gender, mode of testing (OT vs OP), nor interaction between gender and mode of testing affected students' anxiety level. Males and females were equally anxious regardless of mode of testing. This finding was somehow surprising as literature generally suggests that females were inclined to be more anxious than males (Stöber, 2004; Cotner, Jeno, Walker, Jørgensen & Vandvik, 2020; Lowe & Lee, 2008), primarily due to females feeling that males are better than them; they also are more fearful of failing compared to males (Celik, 2019).

However, in the present study, the majority of female students had always been performing better than their male counterparts in the past. Interviews with the teacher revealed that female students were, by and large, academically superior to males. Female students took part in various English debate competitions and the like representing their school. Whilst some male students were also involved in those activities, the number was relatively small compared to that of females. Perhaps it is this sense of being 'better' academically (Numan & Hasan, 2017) and of previous positive experience that makes them feel more secure and more confident, leading to an anxiety level comparable to that of male students (Çimen & Yilmaz, 2015).

Unfortunately, no quantitative data on students' past performance were available to the researchers, apart from the teacher's general qualitative information about students' past performance. Were this quantitative information available, a further statistical analysis taking students' past performance as both the covariate and predictor of both anxiety and test scores could be performed. Understanding students' past performance and experiences is critical because negative experiences in the past could be an important determinant of anxiety (MacIntyre, 1999). This could be another avenue for further research.

Note that, in the present study, female students sitting OP exhibited a slightly higher mild anxiety level than males taking OP, although the difference was not significant. Park and French (2013) referred to this type of anxiety as 'facilitating anxiety' which is considered

to be indispensable to learning. This 'facilitating anxiety' might have led female students to exert more effort and resources to test preparation, to be better prepared for the test and to work harder, thus performing better in the test. This speculation was confirmed by the fact that females outperformed males in the test and they were slightly more anxious than males.

Of course, preparation for the test is just one potentially influential factor. Other factors such as self-efficacy, motivation, parents' and peers' expectations, to name a few, may also come into play and should be carefully scrutinised in future studies to better understand the nature of students' anxiety regardless of gender. Since anxiety is affected by a wide range of factors, both student-internal and student-external, Nahavandi and Mukundan (2013) suggested that classroom anxiety and gender have no relation. While acknowledging the 'inherent' nature of females being prone to anxiety and "men tend to be more confident with regard to almost everything" (Cotner, Jeno, Walker, Jørgensen & Vandvik, 2020 p. 6), further research could be geared towards understanding factors contributing to test anxiety. Only then can necessary steps be taken and informed decision be made to help highly anxious students regain their focus regardless of their genders.

Furthermore, the present study revealed that mode of testing did not affect students' anxiety levels. Consistent with previous studies, the anxiety level of students sitting OP or OT was comparable (Stowell & Bennett, 2010). It goes without saying, technology is increasingly becoming a part of people's everyday business (Suardika, Alberth, Mursalim, Siam, Suhartini & Pasassung, 2020; Alberth, Wiramihardja & Uden, 2020). It is hard to imagine doing things wihout technology these days. Doing online quizzes for many students has now become a part of their daily routines and, for some students, computer-assisted testing is even perceived to be more convenient than its counterpart OP quizzes, primarily owing to immediate feedback provided (Tanduklangi, Lio & Alberth, 2019). However, this might not be the case with early integration of technology in testing where technology was perceived to cause some negative effects (Abidin, Ismail & Ramlan, 2011). As students are becoming more acclimatised to using technology, however, and as technology becomes so ubiquitous, it simply becomes a part of people's lives (Alberth, Mursalim, Siam, Suardika & Ino, 2018). The fact that no significant differences were found in anxiety levels of students in both OP and OT suggests that these students might feel equally comfortable with both modes of testing.

Although the effect of mode of testing on test scores was not significant, which confirmed findings reported by previous studies (e.g. Stowell & Bennett, 2010), there was evidence indicating that gender and interaction between gender and modes of testing had a significant effect on test scores. Females significantly outperformed males in online testing. Test score differences for male and female were much bigger in OT than in OP. What is more, there was a greater difference between OP and OT for male than there was for female. Female participants performed about equally well in both OP and OT, but male participants performed better in OP than in OT. That females outperformed males in this particular study confirmed findings reported in previous studies (Kamphorst, Adriaan Hofman, Jansen & Terlouw, 2015; Workman & Heyder, 2020; Duckworth &

Seligman, 2006; Alfarhan & Dauletova, 2019; Conger & Long, 2010; Alfarhan & Dauletova, 2019).

A meta-analysis conducted by Voyer and Voyer (2014) concluded that, whereas female students generally (not always) outperformed their male counterparts academically in various courses such as science and maths, this trend was more strongly pronounced in language-related courses (Workman & Heyder, 2020) and the present study provided further support to this meta-analysis. Female students' higher test scores were attributed to the fact that they spent more time on independent study (Kamphorst, Adriaan Hofman, Jansen & Terlouw, 2015), exerted greater effort than their male counterparts (Duckworth & Seligman, 2006), worked harder in school (Lam et al., 2012), exerted more effort on their homework compared to male students (Trautwein, Lüdtke, Kastens & Köller, 2006), valued hard work more than males did (McCrea, Hirt & Milner, 2008), were more self-disciplined (Duckworth & Seligman, 2006), put more emphasis upon test preparation than male students (Pomerantz, Altermatt & Saxon, 2002), possessed non-cognitive skills such as help seeking behaviour, self-discipline, dependability, atteniveness and organisation better than male students (Conger & Long, 2010), and were more hard-working, responsible, and organised (Alfarhan & Dauletova, 2019).

Note, however, that in the present study, whilst the effect of gender on test scores was significant and the observed power was high, its effect size as indicated by the partial eta squared statistic was very small. A similar case was also observed for the interaction effect between gender and mode of testing on test scores, in which case the interaction effect was significant and the observed power was relatively high, but the effect size was very low even when outliers in the data had all been removed. It can, therefore, be concluded that despite the significant effect and high observed power, it is of little practical significance.

Interestingly, levels of anxiety (low, medium, high) appeared to impact on test scores differently. Students with medium anxiety levels (facilitating anxiety) and those with low anxiety levels significantly outperformed students with high anxiety levels (debilitating anxiety). Although students with medium anxiety also outperformed those with low anxiety levels, differences in test scores proved to be non-significant. This finding is consistent with the widely cited research evidence indicating that, whereas debilitating anxiety may hamper learning (Lowe, 2019; Aydin, 2019), facilitating anxiety could motivate students to learn and work even harder towards reaching their learning goals (Park & French, 2013; Arifin, Mursalim & Sahlan, 2020).

Limitations of the study

Limitations to the present study need to be recognised. To begin with, the number of male and female students in both cohorts was unequal; female far outnumbered male students in both OP and OT sections. This difference in number might have introduced some bias in the results. Secondly, no quantitative data on students' past performance were available to the researchers. Were this quantitative information available, a further statistical analysis taking students' past performance as both the covariate and predictor of

both anxiety and test scores could be performed, thus confirming whether or not and to what extent past performance can predict both test scores and anxiety levels. What is more, the achievement test developed by the class teacher was not rigorously checked for validity and reliability in the present study, though the test was previously validated with some evidence of validity and reliability. Consequently, the findings reported in the present study should be treated with caution and should be considered suggestive, rather than definitive.

Conclusion

The present study indicated that neither gender, mode of test administration, nor the interaction between gender and mode of testing impacted on students' anxiety levels. Regardless of mode of testing, no significant difference was found in the anxiety levels of male and female students. However, the main effect of gender and the interaction between gender and mode of testing on test scores proved to be significant and the observed power was relatively high, but its effect size was very small, thus lacking practical significance. Female students outperformed their male counterparts in both modes of testing (OP and OT). Test score differences for male and female were much bigger in OT than in OP. There was a greater difference between OP and OT for male than there was for female. Female students performed about equally well in both OP and OT, but male participants performed better in OP than in OT.

Test scores appeared to vary with anxiety levels. Test scores of students who experienced a high anxiety level was significantly lower than test scores of those who experienced a low or medium anxiety level. Students with a medium anxiety level outperformed those with a low anxiety level, but the difference was not significant. This finding confirmed the widely cited undesirable effect of debilitating anxiety and, at the same time, substantiated the positive effect of facilitating anxiety on performance. We have also argued that future research could be geared towards understanding determinants of anxiety with respect to gender across different modes of testing. Only then can necessary steps be taken and informed decision be made in an attempt to help learners maintain their anxiety levels at an ideal level, thus boosting their academic performance.

Acknowledgement

References

Abidin, S. A. Z., Ismail, O. & Ramlan, Z. S. (2011). E-assessment in motion. *Journal of Language Studies,* 7(1), 63-71. https://www.researchgate.net/publication/328913487_E-Assessment_in_Motion

Ahmad, N., Hussain, S. & Khan, F. N. (2018). Test anxiety: Gender and academic achievements of university students. *Journal of Postgraduate Medical Institute,* 32(3). https://jpmi.org.pk/index.php/jpmi/article/view/2251

Alberth, Wiramihardja, E. & Uden, L. (2020). WhatsApp with English language teaching? Some practical ideas and strategies. *International Journal of Technology Enhanced Learning,* 12(3), 262-274. https://doi.org/10.1504/IJTEL.2020.107974

Alberth (2019). Use of Facebook, students' intrinsic motivation to study writing, writing self-efficacy and writing performance. *Technology, Pedagogy and Education,* 28(1), 21-36. https://doi.org/10.1080/1475939X.2018.1552892

Alberth, Mursalim, Siam, Suardika, I. K. & Ino, L. (2018). Social media as a conduit for teacher professional development in the digital era: Myths, promises or realities? *TEFLIN,* 29(2). https://doi.org/10.15639/teflinjournal.v29i2/293-306

Alfarhan, U. F. & Dauletova, V. (2019). Revisiting the gender academic achievement gap: Evidence from a unique environment. *Gender and Education,* 31(7), 827-848. https://doi.org/10.1080/09540253.2017.1324129

Arifin, N., Mursalim, M. & Sahlan, S. (2020). Enhancing speaking performance and reducing speaking anxiety using Ted Talks. *Journal of Language Education and Educational Technology,* 5(1), 41-58. http://ojs.uho.ac.id/index.php/JLEET/article/view/6953

Aydin, U. (2019). Test anxiety: Gender differences in elementary school students. *European Journal of Educational Research,* 8(1), 21-30. https://eric.ed.gov/?id=EJ1203109

Bodas, J., Ollendick, T. H. & Sovani, A. V. (2008). Test anxiety in Indian children: A cross-cultural perspective. *Anxiety, Stress & Coping,* 21(4), 387-404. https://doi.org/10.1080/10615800701849902

Celik, B. (2019). Relationship between foreign language learning and exam stress with gender: A study on Tshik International University Preparatory School students, Iraq. *International Journal of Social Sciences & Educational Studies,* 5(3), 311-322. https://ijsses.tiu.edu.iq/index.php/volume-5-issue-3-article-23/

Chin, E. C. H., Williams, M. W., Taylor, J. E. & Harvey, S. T. (2017). The influence of negative affect on test anxiety and academic performance: An examination of the tripartite model of emotions. *Learning and Individual Differences,* 54, 1-8. https://doi.org/10.1016/j.lindif.2017.01.002

Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing,* 20(3), 295-320. https://doi.org/10.1191/0265532203lt258oa

Çimen, O. & Yilmaz, M. (2015). Evaluating high school students anxiety and self-efficacy towards biology. *Educational Research and Reviews,* 10(7), 987-993. https://doi.org/10.5897/ERR2015.2111

Conger, D. & Long, M. C. (2010). Why are men falling behind? Gender gaps in college performance and persistence. *The Annals of the American Academy of Political and Social Science,* 627(1), 184-214. https://www.jstor.org/stable/40607412

Cotner, S., Jeno, L. M., Walker, J. D., Jørgensen, C. & Vandvik, V. (2020). Gender gaps in the performance of Norwegian biology students: The roles of test anxiety and science confidence. *International Journal of STEM Education,* 7, article 55. https://doi.org/10.1186/s40594-020-00252-1

DeSouza, E. & Fleming, M. (2003). A comparison of in-class and online quizzes on student exam performance. *Journal of Computing in Higher Education,* 14(2), 121-134. https://doi.org/10.1007/BF02940941

DeVellis, R. F. (2003). *Scale development: Theory and applications.* SAGE. [5th ed.] https://au.sagepub.com/en-gb/oce/scale-development/book269114

Di Meo, F. & Martí-Ballester, C.-P. (2020). Effects of the perceptions of online quizzes and electronic devices on student performance. *Australasian Journal of Educational Technology,* 36(1), 111-125. https://doi.org/10.14742/ajet.4888

Duckworth, A. L. & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology,* 98(1), 198-208. https://psycnet.apa.org/doi/10.1037/0022-0663.98.1.198

Fageeh, A. I. (2015). EFL student and faculty perceptions of and attitudes towards online testing in the medium of Blackboard: Promises and challenges. *JALT CALL Journal,* 11(1), 41-62. https://eric.ed.gov/?id=EJ1107991

Fynewever, H. (2008). A comparison of the effectiveness of web-based and paper-based homework for general chemistry. *The Chemical Educator,* 13(4), 264-269. http://chemeducator.org/bibs/0013004/13080264hf.htm

Gorjian, B. & Javadifar, M. (2013). Effects of gender and passage content on multiple-choice reading comprehension test. *Procedia - Social and Behavioral Sciences,* 84, 723-727. https://doi.org/10.1016/j.sbspro.2013.06.634

Kaarakainen, M.-T., Kivinen, A. & Kaarakainen, S.-S. (2017). Differences between the genders in ICT skills for Finnish upper comprehensive school students: Does gender matter? S*eminar.net,* 13(2). https://journals.oslomet.no/index.php/seminar/article/view/2304

Kamphorst, J. C., Adriaan Hofman, W. H., Jansen, E. P. W. A. & Terlouw, C. (2015). Explaining academic success in engineering degree programs: Do female and male students differ? *Journal of Engineering Education,* 104(2), 189-211. https://doi.org/10.1002/jee.20071

Karadeniz, S. (2009). The impacts of paper, web and mobile based assessment on students' achievement and perceptions. *Scientific Research and Essay,* 4(10), 984-991. https://academicjournals.org/article/article1380547300_Karadeniz.pdf

Lam, S.-F., Jimerson, S., Kikas, E., Cefai, C., Veiga, F. H., Nelson, B., et al. (2012). Do girls and boys perceive themselves as equally engaged in school? The results of an international study from 12 countries. *Journal of School Psychology,* 50(1), 77-94. https://doi.org/10.1016/j.jsp.2011.07.004

Lowe, P. A. (2019). Exploring cross-cultural and gender differences in test anxiety among U.S. and Canadian college students. *Journal of Psychoeducational Assessment,* 37(1), 112-118. https://doi.org/10.1177/0734282917724904

Lowe, P. A. & Lee, S. W. (2008). Factor structure of the Test Anxiety Inventory for Children and Adolescents (TAICA) scores across gender among students in elementary and secondary school settings. *Journal of Psychoeducational Assessment,* 26(3), 231-246. https://doi.org/10.1177/0734282907303773

Macedo-Rouet, M., Ney, M., Charles, S. & Lallich-Boidin, G. (2009). Students' performance and satisfaction with Web vs. paper-based practice quizzes and lecture notes. *Computers & Education,* 53(2), 375-384. https://doi.org/10.1016/j.compedu.2009.02.013

MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere,* Ch. 3.

McCrea, S. M., Hirt, E. R. & Milner, B. J. (2008). She works hard for the money: Valuing effort underlies gender differences in behavioral self-handicapping. *Journal of Experimental Social Psychology,* 44(2), 292-311. https://doi.org/10.1016/j.jesp.2007.05.006

Nahavandi, N. & Mukundan, J. (2013). Foreign language learning anxiety among Iranian EFL learners along gender and different proficiency levels. *Language in India,* 13(1), 133-161. https://www.academia.edu/2402309/Foreign_Language_Learning_Anxiety_among_Irani an_EFL_learners_Along_Gender_and_Different_Proficiency_Levels

Nist, P. & Diehl, M. (1990). *PHCC Test Anxiety Questionnaire.* [retrieved 20 August 2010 from http://phcc.edu/ods/questionnaire.html; not found 5 July 2021, see alternative sources, e.g. https://nursing.lsuhsc.edu/apps/AcademicSuccessProgram/AnxietyQuestionnaire/]

Numan, A. & Hasan, S. S. (2017). Effect of study habits on test anxiety and academic achievement of undergraduate students. *Journal of Research & Reflections in Education,* 11(1), 1-14. https://tehqeeqat.com/downloadpdf/2559

Ogundokun, M. (2011). Learning style, school environment and test anxiety as correlates of learning outcomes among secondary school students. *IFE PsychologIA,* 19(2), 321-336. https://doi.org/10.4314/ifep.v19i2.69555

Park, G.-P. & French, B. F. (2013). Gender differences in the foreign language classroom anxiety scale. *System,* 41(2), 462-471. https://doi.org/10.1016/j.system.2013.04.001

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M. & Molfenter, S. (2004). Beyond test anxiety: Development and validation of the test emotions questionnaire (TEQ). *Anxiety, Stress & Coping,* 17(3), 287-316. https://doi.org/10.1080/10615800412331303847

Pomerantz, E. M., Altermatt, E. R. & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology,* 94(2), 396-404. https://doi.org/10.1037/0022-0663.94.2.396

Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests. *Journal of Technology, Learning, and Assessment,* 2(6). https://ejournals.bc.edu/index.php/jtla/article/view/1666

Sidhu, S. M. (2008). Computer-based assessment. In L. A. Tomei (Ed.), *Encyclopedia of information technology curriculum integration* (pp. 131-137). IGI Global. https://doi.org/10.4018/978-1-59904-881-9

Soares, D. & Woods, K. (2020). An international systematic literature review of test anxiety interventions 2011-2018. *Pastoral Care in Education,* 38(4), 311-334. https://doi.org/10.1080/02643944.2020.1725909

Stöber, J. (2004). Dimensions of test anxiety: Relations to ways of coping with pre-exam anxiety and uncertainty. *Anxiety, Stress & Coping,* 17(3), 213-226. https://doi.org/10.1080/10615800412331292615

Stowell, J. R. & Bennett, D. (2010). Effects of online testing on student exam performance and test anxiety. *Journal of Educational Computing Research,* 42(2), 161-171. https://doi.org/10.2190/EC.42.2.b

Suardika, I. K., Alberth, Mursalim, Siam, Suhartini, L. & Pasassung, N. (2020). Using WhatsApp for teaching a course on the education profession: Presence, community and learning. *International Journal of Mobile and Blended Learning,* 12(1), 17-32. https://doi.org/10.4018/IJMBL.2020010102

Tanduklangi, A., Lio, A. & Alberth, A. (2019). Classroom action research in teaching English for senior high school students through blended learning in Kendari of Indonesia. *Journal of e-Learning and Knowledge Society,* 15(1). https://doi.org/10.20368/1971-8829/1579

Trautwein, U., Lüdtke, O., Kastens, C. & Köller, O. (2006). Effort on homework in grades 5-9: Development, motivational antecedents, and the association with effort on classwork. *Child Development,* 77(4), 1094-1111. https://www.jstor.org/stable/3878417

Von der Embse, N., Jester, D., Roy, D. & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders,* 227, 483-493. https://doi.org/10.1016/j.jad.2017.11.048

Voyer, D. & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin,* 140(4), 1174-1204. https://www.apa.org/pubs/journals/releases/bul-a0036620.pdf

Way, W. D., Davis, L. L. & Fitzpatrick, S. (2006). Score comparability of online and paper administrations of the Texas Assessment of Knowledge and Skills. *PEM Research report 06-01*, April 2006. http://images.pearsonassessments.com/images/tmrs/Score_Comparability_of_Online_and_Paper_Administrations_of_TAKS_03_26_06_final.pdf

Winke, P. & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly,* 14(4), 380-397. https://doi.org/10.1080/15434303.2017.1399396

Woldeab, D. & Brothen, T. (2019). 21st Century assessment: Online proctoring, test anxiety, and student performance. *International Journal of E-Learning & Distance Education,* 34(1), 1-10. http://www.ijede.ca/index.php/jde/article/view/1106

Workman, J. & Heyder, A. (2020). Gender achievement gaps: The role of social costs to trying hard in high school. *Social Psychology of Education,* 23(6), 1407-1427. https://doi.org/10.1007/s11218-020-09588-6

Wren, D. G. & Benson, J. (2004). Measuring test anxiety in children: Scale development and internal construct validation. *Anxiety, Stress & Coping,* 17(3), 227-240. https://doi.org/10.1080/10615800412331292606

Appendix A: The PHCC Test Anxiety Questionnaire

Item	Statement
1.	I have visible signs of nervousness such as sweaty palms, shaky hands, and so on right before a test.
2.	I have "butterflies" in my stomach before a test
3.	I feel nauseated before a test
4.	I read through the test and feel that I do not know any of the answers.
5.	I panic before and during a test.
6.	My mind goes blank during a test.
7.	I remember the information that I blanked on once I get out of the testing situation.
8.	I have trouble sleeping the night before a test.
9.	I make mistakes on easy questions or put answers in the wrong places.
10.	I have difficulty choosing answers.

Response scale: Never=1; Rarely=2; Sometimes=3; Often=4; Always=5

Dr Sahlan is a senior lecturer currently working at the Department of Indonesian Language and Literature Education, Faculty of Teacher Training and Education, Universitas Halu Oleo, Indonesia. He has taught a range of courses, including language testing and evaluation, education profession, literary theory, and language teaching research. His research interests include language testing and integration of local wisdom in education.
Email: sahlan.kdi@gmail.com

Dr Alberth holds a PhD in language and intercultural education from Curtin University, Australia, and is currently a senior lecturer attached to the Department of English Language Education, Universitas Halu Oleo. Alberth's research interests encompass computer-assisted language learning, applied linguistics and teaching English as a foreign language. Alberth has published extensively in a wide array of internationally refereed journals, including book chapters, and has presented at various national and international conferences.
ORCID: https://orcid.org/0000-0002-0846-1712
Email: alberth@uho.ac.id

Wahyudin Madil is a lecturer at the Department of English Language Education, Universitas Halu Oleo, Indonesia, and is presently pursuing a doctorate in educational technology at State Universitas Negeri Malang, Indonesia. He has taught a range of courses, including web-based learning, critical reading, and computer assisted language learning. His research interests include mobile assisted language learning, teaching English as a foreign language and media in language teaching.
ORCID: https://orcid.org/0000-0002-9964-2921
Email: wahyu.md@uho.ac.id, wahyuddin.uho@gmail.com

Hutnisyawati is an English teacher at SMPN 12 Konawe Selatan, Southeast Sulawesi, Indonesia. Holding a masters degree in language education, Hutnisyawati's research interests include teaching strategies, teaching techniques, and critical thinking skills.
Email: nischajuddah@gmail.com, hutnisyawati.kdi@gmail.com