RESEARCH

Open Access

CrossMark

The relationships between test performance and students' perceptions of learning motivation, test value, and test anxiety in the context of the English benchmark requirement for graduation in Taiwan's universities

Jessica Wu[*] and Matt Chia-Lung Lee

* Correspondence:
jw@lttc.ntu.edu.tw
The Language Training & Testing
Center, Taipei, Taiwan

Abstract

Backgound: Having been influenced by the trend of internationalization of higher education, most universities in Taiwan have implemented an English benchmark requirement for graduation, which requires students to demonstrate their English ability at a specified Common European Framework of Reference for Languages (CEFR) level through taking a standardized English language test (e.g., GEPT, IELTS, TOEFL iBT). This practice has been increasingly criticized for failing to achieve its intended goals of enhancing students' English language proficiency and increasing students' career mobility. Therefore, there is an urgent need to explore the consequences of using standardized tests in support of the policy.

Methods: To this end, this study investigated students' views of the graduation policy in three universities where students are required to take the General English Proficiency Test (GEPT) prior to graduation. Structural equation modeling was employed to find the best fitting model that illustrates the complex interrelationships among test performance, students' perceptions of the requirement, test value, test anxiety, and learning motivation.

Results: The findings show that university students, regardless of English proficiency, generally hold a positive attitude towards the English graduation benchmark policy. Results further reveal that the Intermediate group shows more positive attitudes towards the graduation requirement than the High-Intermediate group. SEM results show that the attitudes of university students towards the English graduation requirement positively impact their perceived test value and their learning motivation. However, there is no significant relationship between the attitudes towards the policy and test performance.

Conclusions: The findings contribute to our understanding of university students as the major stakeholders who defined the context of test use.

Keywords: Learning motivation, Structural equating modeling, Test consequences

Background

In 2005, the Ministry of Education (MoE) began implementing the English graduation benchmark policy in universities across Taiwan, with the aim to encourage college students to pass a credible standardized English test before graduation. Behind the policy lies the belief that taking a standardized test helps students to prove their English competence, prepare for future employment or advanced studies, and also increase their competitive edge in the global environment. Authors like Shohamy (2000) and McNamara (2001) have asserted that language testing serves the function of gate-keeping in many contexts. Similarly, in the context of Taiwan's higher education, language testing has been utilized as a tool to enforce the power of the English graduation benchmark policy.

At present, more than 90% of universities in Taiwan have implemented the English graduation benchmark policy, with each university setting its own benchmark standard. As students from different universities exhibit varying levels of English competence, it is very difficult for the government to formulate uniform standards. However, the MoE's adoption of the Common European Framework of Reference for Languages (CEFR) provided institutions, schools, and the general public with a reference for various levels of language competence and testing requirements (Council of Europe 2001). According to the CEFR, language ability is divided into six levels, specifically A1, A2, B1, B2, C1, and C2, in ascending order from low to high competence. Most of the universities require that their students pass an English test equivalent to the CEFR B1 level upon graduation, whereas the top five universities set their graduation benchmark at the CEFR B2 level. Starting in 2016, National Taiwan University, Taiwan's most prestigious university, has set the benchmark at the CEFR C1 level for its English major students.

Most universities accept either Test of English for International Communication (TOEIC) or General English Proficiency Test (GEPT) scores as proof of their students' English proficiency. Each university sets its own requirement if students fail to pass the graduation benchmark. Some schools require that students retake the test until they pass, some ask students to participate in additional training if they fail the test twice, and others set various "backdoor policies," such as requiring students who fail to sign up for an extra course as a required condition for graduating (Chen 2014). Although the graduation benchmark policy is widely implemented across Taiwan, regulations and problems stemming from the policy have become highly controversial, causing dissatisfaction among students and teachers as they demand to have the policy reexamined (Her et al. 2013; Chang 2005).

The English graduation benchmark policy has attracted the interest of scholars in Taiwan, and some studies have investigated the effectiveness of implementing such a policy. Among existent studies, Chen and Liu (2007), Liao (2010), and Shih (2008) use questionnaires and interviews to compare the learning motivation of students before and after the implementation of the benchmark policy. Their findings show that the policy has helped students develop a more positive attitude towards English learning, with over 50% of the students affirming the positive effect of the policy. In other words, the students surveyed generally agree that implementation of the policy has increased their motivation for learning English. Huang (2010) found that students show relatively low anxiety towards the benchmark policy. Students enrolled in a school without the policy tend to show a higher anxiety than those who are required to pass the benchmark tests.

At present, there are still relatively few studies focusing on the English graduation benchmark policy, learning motivation, and test anxiety, and since the number of studied samples could be larger, more empirical research is needed to show that the policy helps to promote English learning motivation. Data show that college students in Taiwan spend a total of 1 billion new Taiwan dollars (about USD30 million) every 4 years on taking English standardized tests, but the effectiveness of these tests is limited, and there is no evidence that taking English standardized tests helps to increase students' learning motivation or English communication skills (Her et al. 2013).

This study seeks to contribute to this area of research. By sampling a greater number of students, we hope to better understand how students view this policy. Moreover, since learning motivation is a complicated cognitive factor (Cheng et al. 2014), the English graduation benchmark policy is not only related to learning motivation but also reflects students' attitudes towards English standardized tests, and how they feel about their English proficiency. To provide a more comprehensive analysis of students' views on the English graduation benchmark policy, this study will consider relevant factors such as students' learning motivation, their views on the importance of English tests, test anxiety, test performance, and the inter-relationships among them.

Literature review

Motivation and performance

Motivation is a complex psychological trait; it is a hidden force within people that drives them to take action. "Learning motivation" can be defined as an internal thought process that triggers learners to achieve specific physiological or psychological goals and devote their effort voluntarily, while maintaining momentum throughout the learning process (Stipek et al. 1995). In general, motivation is considered one of the most important factors in determining a student's second language performance. However, the actual correlation between learning motivation and performance has not been accurately determined, and a precise definition of this construct is still lacking (Dörnyei 2001). Regarding this, Eccles and Wigfield (2002) have summarized many theories on the relationship between motivation and performance, including intrinsic motivation theory, self-determination theory, flow theory, and goals theory. They also integrated theories of expectation and value construction, such as attribution theory, expectancy-value theory, self-worth theory, and theories integrating motivation and cognition.

Among these theories, a popular topic of research today is self-determination theory (SDT), proposed by Deci and Ryan in 1985. They discussed how when an inherent positive belief persists in an individual, the person will work diligently to achieve their committed goal. In the process, the three inherent needs of human beings—namely autonomy, competence, and psychological relatedness—are fulfilled, which will bring about optimal progress and development for the individual. Ryan and Deci (2000) suggest that both intrinsic and extrinsic motivation can stimulate individual behavior and performance. While intrinsic motivation drives a person to continuously pursue better performance based on his or her interest in the subject and the desire to meet the needs of competence and autonomy, extrinsic motivation is often driven by some actual form of reward, which is considered outside the realm of self-determination, and only through internalization, integration, and regulation does it become part of the self-determination process.

Relationships among attitudes, motivation, test anxiety, and test performance

Although there is still considerable controversy surrounding the SDT theory proposed by Ryan and Deci, Cheng et al. (2014) agreed with Ryan and Brown (2005) that compared to other motivation theories, the SDT principle is more applicable to the discussion of the relationship between learning motivation and high-stakes testing. Ryan and Weinstein (2009) used the SDT theory to explain the effect of high-stakes testing on teachers and learners. In high-stakes testing, the motivation for success in learners is not static, but may vary according to people, time, environment, and other factors. Regardless of whether the test-taker has a set purpose for taking the test, the inherently complex nature between the test-taker and the test may influence motivation as well. Ryan and Brown (2005) argued that the policy of test taking is predicated on the grounds of using reward, punishment, and pressure on self-esteem to motivate learning. Noels (2005) suggested that the social and cultural contexts outside the classroom can greatly influence the motivation for learning a second or foreign language. The effect is even more evident when the purposes of the test or the stakes involved vary (e.g., entrance examination, placement test, immigration test). This can lead to different levels of test anxiety and can possibly become a construct-irrelevant variance.

Research confirms that a person's attitude directly influences motivation and behavior (Ajzen 1985; Fishbein and Ajzen 1975). Ajzen (1985) proposed the Theory of Planned Behavior, suggesting that an individual's actions are not only influenced by his or her preference for certain people, things, events, and the subjective norm of society and others but also affected by the individual's desire to complete an action and the perceived behavioral control towards resources and opportunities.

A person's attitude towards specific things and events is also likely to affect motivation and anxiety. Pyun (2013) studied the relationship between task-based language learning and learning motivation and anxiety. The results showed that learner attitude and anxiety show significant negative correlation, while learner attitude and motivation show significant positive correlation. Test anxiety typically surfaces during certain cognitive performances for test takers, such as when they compare themselves with their peers, worry about the consequence of failing a test, experience low self-confidence, or are excessively worried about testing and assessment (In'nami 2006; Liu 2008). Different levels of test anxiety can be attributed to the individual's family background, social environment, and the teaching methods he or she is used to. For example, parents who expect outstanding test results from their children may be putting excessive pressure on them (Bodas and Ollendick 2005). Considering these points of view, test anxiety can negatively impact test performance, impeding test takers from achieving their full potential and strength (Elkhafaifi 2005; Meijer 2001). Many studies have shown a negative correlation between test anxiety and academic performance (Chapell et al. 2005; Ruthig et al. 2004; Putwain 2007).

Test anxiety also affects test takers' performance on foreign language tests, as shown by a study carried out by Tsai (2010). She reported that there was significant negative correlation ($\gamma = -.49$) between test anxiety and performance from the results of the listening section of the GEPT Intermediate Level Test. Her conclusion echoes Kunnan's (1995) view that the complex interaction between social and cognitive factors may influence the test results of students in different testing scenarios, depending on differences in test purpose and the stakes involved in the local context.

The study by Cheng et al. (2014) is one of the few examples using the application of SDT to the area of language learning. It was also the first time that empirical research

has been used to examine the complex relationships among English learners' motivation, test value, test anxiety, and high-stakes test performance (i.e., test scores) from the perspective of social and cognitive factors. The study invited English learners from Taiwan, China, and Canada to take local high-stakes English proficiency tests, and a questionnaire which explored motivation, test anxiety, and perceptions of test importance and purpose to test-takers in each of the three contexts was conducted immediately after the test. The questionnaire responses were then cross-analyzed with the learners' test scores, and the relationships between English learning motivation and test performance under different social and educational contexts were compared. In the study, the high-stakes test used in Taiwan was the GEPT High-Intermediate Level Test, while the students sampled were university students. The results of the study showed a common phenomenon in all three contexts: the purpose of the test and learners' recognition of the importance of the test can affect motivation. Moreover, as Fig. 1 depicts, since motivation and anxiety exhibit a mutually influential relationship, these factors will affect the learners' overall test performance, along with other personal factors such as the test taker's age and gender (Cheng et al. 2014).

However, Cheng et al. pointed out that these inferences need to be confirmed by further research and different statistical methods. Moreover, the study's "test purpose" was filled in by respondents who completed the questionnaire and not defined specifically as "English graduation benchmark." Authors (2016) used the theoretical framework of Cheng et al., but included questions related to "English graduation benchmark policy" in the questionnaire. The study explored the views of 1620 students from two universities in northern Taiwan on the English graduation benchmark policy, as well as the relationships among learning motivation, test value, test anxiety, and test performance. The structural equation model (SEM) was applied to test the complex relationships among variables. The findings of
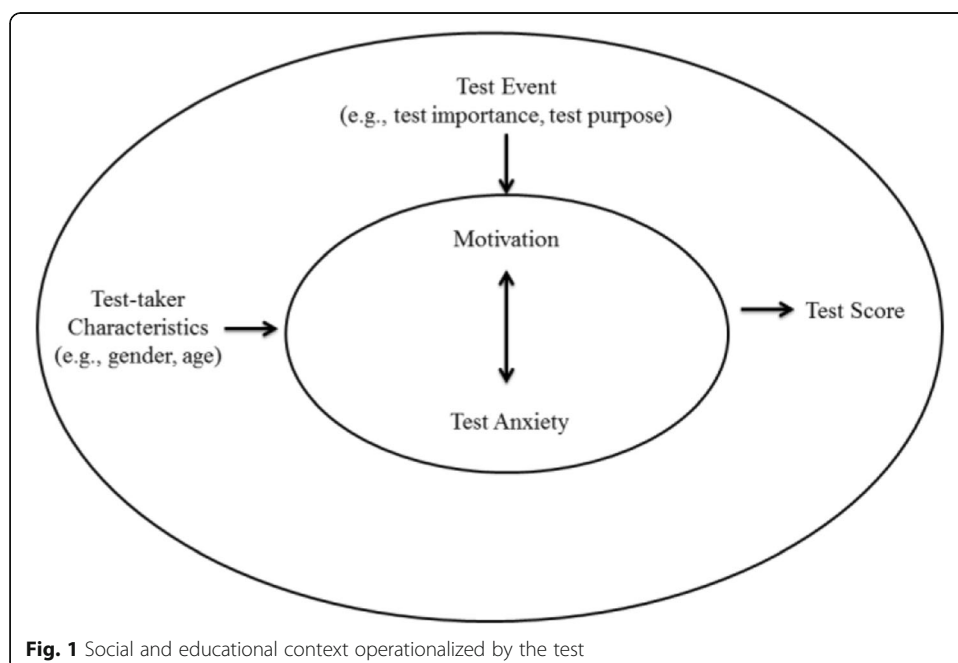

Fig. 1 Social and educational context operationalized by the test

their study show that university students generally hold a positive attitude towards the English graduation benchmark policy. SEM results show that the attitudes of university students towards the English graduation benchmark policy positively impact their perceived test value and their learning motivation.

Based on the preliminary results, authors (2016) suggest that their study serves the purpose of examining the consequences of using the GEPT as a benchmark test for graduation in the context of Taiwan's higher education. Yet, with regard to one limitation acknowledged by the researchers themselves, their study only used data from students of two universities that have established passing the GEPT High-Intermediate Level Test (equivalent to the CEFR B2) as their English graduation benchmark requirement. They speculated that the sampled students tend to show a stronger learning motivation and more autonomous learning behavior than the university students who have lower levels of English proficiency; thus, their attitudes towards the graduation benchmark policy are also more positive and optimistic. Therefore, they suggest that to confirm the findings of their study, more data needs to be gathered from other universities that use different levels of the GEPT as their graduation benchmark requirement. Although the High-Intermediate group data could be divided into passers and non-passers in order to examine whether results might vary due to students' English proficiency, we considered it to be more effective if the results between two different GEPT levels could be compared directly because over 30% of the High-Intermediate non-passers are very likely to pass the GEPT Intermediate based on the results of the GEPT vertical scaling studies (e.g., Wu and Liao 2010). Therefore, following the procedures employed in the previous study which used the GEPT High Intermediate data, the present study collected new data from one university that uses the Intermediate Level of the GEPT as its graduation benchmark requirement. By comparing the results from these two studies that used different levels of the GEPT and involved students who have different levels of English proficiency, we hope to obtain more insights into the context of using the GEPT scores to fulfill the graduation requirement and to justify the consequences of using the GEPT to encourage university students to learn more English. Given that the two studies were conducted by the same researchers and the latter one was carried out on the basis of the former with respect to the validation of the research instrument and the model to explain the relationship among students' attitudes' towards the GEPT, learning motivation, test value, test anxiety, and test performance, this paper will report these two studies as one single study.

Research questions and hypotheses

Based on the literatures previously discussed, two main research questions were addressed:

1. What are students' attitudes towards the English benchmark policy for graduation in Taiwan's universities? Are there differences between the High Intermediate group and the Intermediate group?
2. What are the relationships between students' attitudes towards the graduation benchmark policy, their test performance (i.e., raw scores), and their perceptions of learning motivation, test value, and test anxiety? Are there differences between the High Intermediate group and the Intermediate group?

To help explore the complex relationships between the variables, the following research hypotheses and research model were proposed (Fig. 2). Hypotheses 1–4: The attitudes towards the English graduation benchmark policy have a significant positive effect on test value, learning motivation, and test performance; hypothesis 5: The attitudes towards the English graduation benchmark policy have a significant negative effect on test anxiety; hypotheses 6–9: Test value has a significant positive effect on learning motivation, test performance, and test anxiety; hypotheses 10–11: Learning motivation has a significant positive effect on test performance; hypotheses 12–13: Test anxiety and learning motivation show an interaction effect; hypotheses 14: Test anxiety has a significant negative effect on test performance.

Methods

As mentioned earlier, the study was carried out on the basis of the previous research (Authors 2016) with respect to the validation of the research instrument and the model to explain the relationship among the variables. Although the previous research based on the GEPT High-Intermediate data has been published (in Chinese), for the sake of clarity and coherence, a summary of the research procedures which introduces the research instruments and explains how the best-fit model was established is provided below. As for the other details of the previous study, they are not included here due to text length limitation. Having said so, the key findings of the previous study are discussed and compared with those of the new study which used the GEPT Intermediate data. By doing so, we hope that we can provide more insights when addressing the research questions.

Measurement instruments

Two measurement instruments were used: the General English Proficiency Test (GEPT) and the questionnaire on learning motivation, test value, and test anxiety.
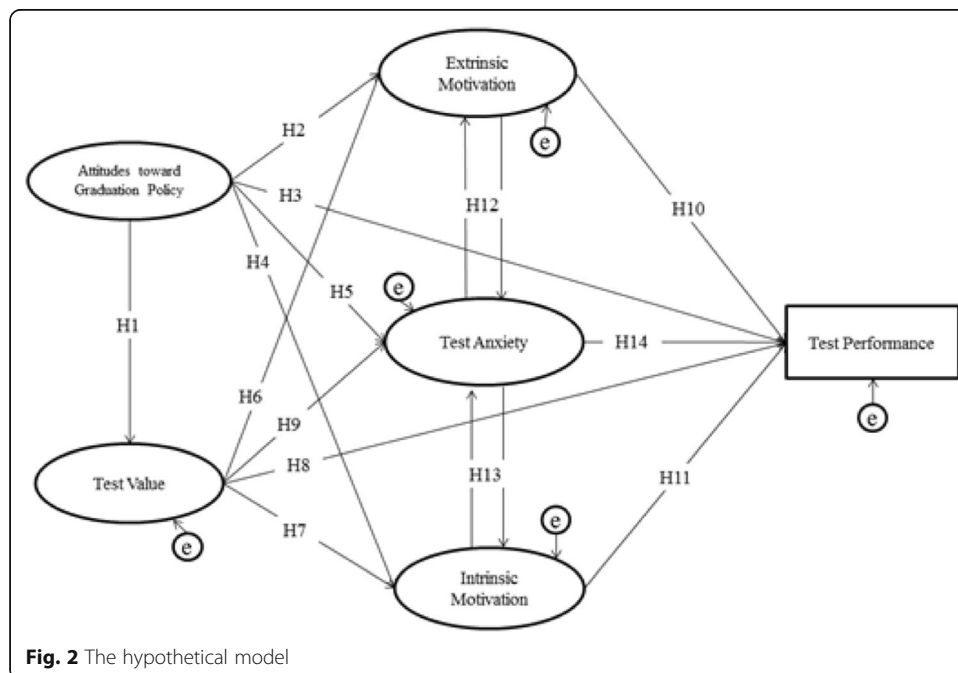

Fig. 2 The hypothetical model

GEPT

The General English Proficiency Test (GEPT) is a five-level criterion-referenced EFL testing system, which targets English learners in Taiwan at all levels, from junior high school upwards. The development of the GEPT was started as an in-house project of the Language Training and Testing Center (LTTC). Later, it was partially funded by Taiwan's Ministry of Education with the aims of promoting life-long learning and introducing positive washback effect on the learning and teaching of English. Since its launch in 2000, the GEPT has been administered independently by the LTTC. Currently, the GEPT is the largest standardized English language test in Taiwan, which is taken by approximately 500,000 test takers at over 100 test sites around the country each year. Numerous evidence of validity has been demonstrated to support the use of the GEPT as a valid indicator of learners' English language proficiency (e.g., Chan et al. 2014; Liao 2016; Weir et al. 2013; Wu 2016). Currently, GEPT scores are not only considered as proof of English ability by government offices, schools, and employers domestically, but also increasingly recognized by universities around the world, including prestigious institutions in Hong Kong, Japan, France, Germany, the UK, and the US, as a means of measuring the English language ability of Taiwanese learners who are interested in pursuing further study overseas.

The test content of the GEPT is not only linked to the local English curriculum but also takes account of local cultural and social references. The levels of the GEPT, which are also linked with the CEFR empirically, are roughly equivalent to CEFR A2–C1 (e.g., Wu and Wu 2010; Wu 2011).

The GEPT was designed as a skill-based test battery assessing both receptive (listening and reading) and productive (speaking and writing) skills. The test places equal weight on each of the four test components and has general level descriptors and skill-area level descriptors. Each GEPT level is administered in two stages. Test-takers who pass listening and reading (160 out of 240 score points) are allowed to register for speaking and writing (80 out of 100 score points). Those who pass both stages will automatically receive both a score report and a Certificate of General English Proficiency. More details about the GEPT and associated research are available at http://www.gept.org.tw.

The questionnaire

This questionnaire was adapted from the questionnaire used in the study by Cheng et al. (2014). However, with the aim to avoid negative associations students might have with "graduation benchmark" and "test anxiety," the study dispersed questions related to these two dimensions in the section related to test value. Cheng et al. in 2009 studied a sample of 538 test takers of the GEPT High-Intermediate Level Test, and the questionnaire proved to have good reliability and validity. Some of their questionnaire items were included in the questionnaire for this study, for example, "I study English for the satisfaction I gain from learning new things," "I study English for the purpose of finding an ideal job," and "I am under a lot of pressure to get good scores on this test." With a focus on attitudes towards the English graduation benchmark, the questionnaire in the study aimed to reflect students' views on the implementation of the benchmark policy; thus, items such as "I think it is reasonable to require college students to pass the GEPT High-Intermediate Level Test" and "I understand that my university encourages students to take the GEPT in order to enhance their English proficiency" were created. After piloting, the final version of the questionnaire contains 43 questions, of

which 11 are related to test value, 8 related to the benchmark policy, 4 related to test anxiety, and 20 related to learning motivation. The questionnaire uses the six-point Likert-type scale. The higher the number on the scale, the more important the students think the test is, the more positive the attitude towards the benchmark policy, the higher the test anxiety, and the stronger the learning motivation. The questionnaire was conducted in Chinese.

Having passed through a set of rigorous validation procedures, we were able to confirm the relevance between the questionnaire items and the latent variables previously established. The factor extraction method with the principal axis factor and the oblimin rotation was used to extract 7 factors with eigenvalues greater than 1, and the cumulative explained variance was 55.53%. But after deleting items that did not meet qualification (a factor loading lower than .4 or items with multiple factor loadings), factor analysis was performed and in the end, 5 factors and 24 items were retained, among which 5 were related to test value, with factor loadings of .74~.89, Cronbach's α of .90; 5 were related to English graduation benchmark, with factor loadings of .54~.72, Cronbach's α of .73; 3 were related to test anxiety, with factor loadings of .78~.86, Cronbach's α of .86; and learning motivation was divided into 6 items on intrinsic motivation, with factor loadings of .52~.84, Cronbach's α of .89 and 5 questions on extrinsic motivation, with factor loadings of .55~.90, Cronbach's α of .84, with the cumulative explained variance at 64.72% (see Table 1).

Next, we used SEM to conduct reliability and validity analysis and to explore possible relationships between the latent variables. SEM can be used to test the theoretical assumptions between observed variables and latent variables and to verify whether hypothetical relationships between observed variables can be supported by empirical data. In general, a model can be established based on previous research, theory, general knowledge, or a hypothesis proposed by the researcher. The correlation-covariance matrix of the estimated model and that of the observed data are then compared, using the chi-squared test to test the difference between the two. The smaller the chi-squared value is, the greater the indication that the data fit the model, meaning the model reflects the observed data more accurately.

Parameter calibration and model-fit analysis

By using statistical software AMOS 23.0, the relationships among attitudes towards the graduation benchmark, test value, test anxiety, learning motivation, and test performance were analyzed. The maximum likelihood estimation (MLE) was applied to calibrate the parameters, and after a series of calibration, model evaluation, and theoretical modification of the model, the final model was established (Fig. 3). Before exploring the relationships among latent variables, we needed to assess whether the model fits the data, which can be determined according to the following indicators. The first indicator is the chi-squared value; if the chi-squared value is not significant, then the model fits the data. However, because the chi-squared value may increase due to the number of samples and the complexity of the model, we also need to examine the goodness-of-fit index (GFI), comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). When an optimal model needs to be chosen from numerous competitive models, this can be determined by comparing the Akaike Information Criterion (AIC) with the Bayesian Information Criterion (BIC): the smaller the value, the more parsimonious the model.

Table 1 Items, factor loadings, explained variance, and Cronbach's *a*

Factors	TV	AGP	TA	IM	EM
Items					
1. How important is my GEPT score for applying to graduate schools.	.76				
2. How important is my GEPT score for gaining professional certification.	.89				
3. How important is my GEPT score for obtaining an academic award/ scholarship.	.80				
4. How important is my GEPT score for obtaining a job.	.83				
5. How important is my GEPT score for studying/working abroad.	.74				
1. I understand that my university encourages students to take the GEPT in order to enhance their English proficiency.		.62			
2. I think it is reasonable to require college students to pass the GEPT High-Intermediate Level Test.		.72			
3. I think the GEPT is fair and reliable.		.54			
4. I agree that Taiwan should develop its own English proficiency tests, for instance the GEPT.		.54			
5. I think my university should not require students to take the GEPT.		.56			
1. I am under a lot of pressure to get good scores on this GEPT.			.82		
2. The GEPT listening section makes me feel anxious.			.86		
3. The GEPT reading section makes me feel anxious.			.78		
I study English…					
1. Because I enjoy reading English literature.				.73	
2. For the satisfaction I gain from learning new things.				.65	
3. Because I want to know more about English communities and their way of life.				.52	
4. Because I like to understand difficult topics in English.				.84	
5. For the satisfaction I feel when I am completing a difficult task in English.				.68	
6. For the excitement I feel when hearing English.				.73	
I study English…					
1. Because I want to speak more than one language.					.70
2. Because I think it helps my personal development.					.90
3. Because I want to be able to speak English.					.86
4. Because I would like to perform better at school or at work.					.60
5. Because I want to get a better job.					.55
Explained variance	26.42	5.86	7.04	14.89	10.51
Cumulative explained variance	26.42	32.28	39.32	54.21	64.72
Cronbach's *a*	.90	.73	.86	.89	.84

Note: *TV* test value, *AGP* attitudes towards graduation policy, *TA* test anxiety, *IM* intrinsic motivation, *EM* extrinsic motivation

To achieve a parsimonious model, we removed the paths that did not reach statistical significance from the hypothetical model, and according to the modification index (MI), added a correlation path between the residuals of intrinsic motivation and extrinsic motivation. Intrinsic motivation and extrinsic motivation were originally two sublatent variables of learning motivation; thus, one way of modifying the model was to add a higher-level latent variable (learning motivation) to the two variables, while another possibility was to add a new correlation coefficient between the residuals of latent variables. In an attempt to better understand the impact of the graduation benchmark

Fig. 3 The modified model for the High-Intermediate group
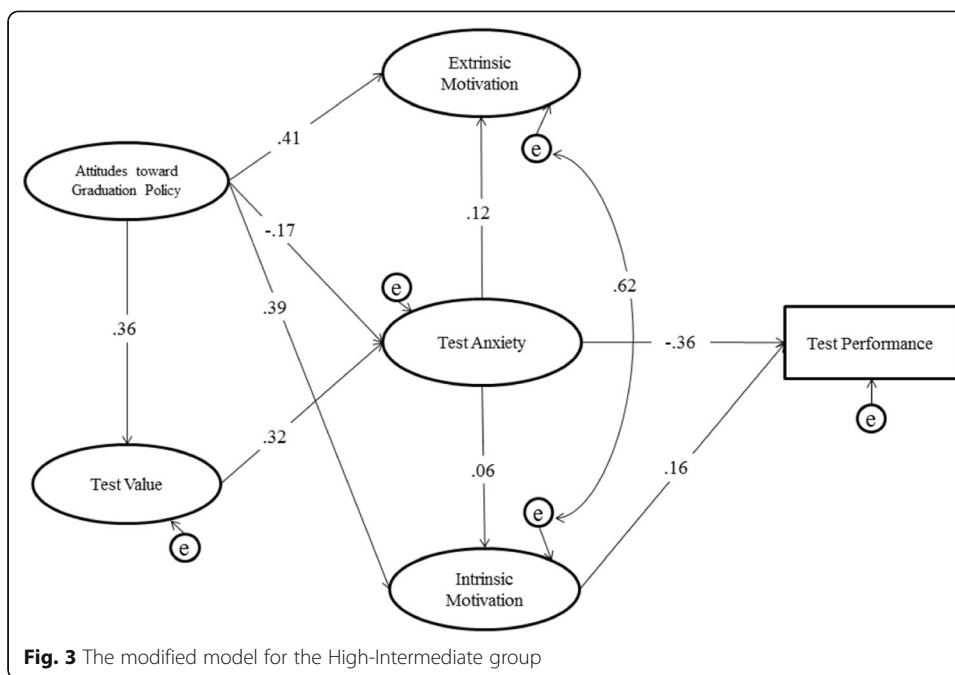
on intrinsic and extrinsic motivation, we decided on the latter and modified the model accordingly (Fig. 3).

As can be seen from Table 2, the chi-squared value in both the hypothetical model and the modified model fail to reach the standard of "absolute fit." This could be due to the exceedingly large sample size ($n = 1620$) or the large number of parameters to be calibrated. The various indicators of the modified model ($x^2 = 2011.83$, GFI = .91, CFI = .91, RMSEA = .06, SRMR = .05, AIC = 2131.83, BIC = 2133.79) show an improvement from those of the hypothetical model ($x^2 = 2504.20$, GFI = .89, CFI = .88, RMSEA = .07, SRMR = .09, AIC = 2624.20, BIC = 2626.16). We then used the chi-square difference test to test which model has a better fit. The results of chi-square difference test also showed that the modified model was superior to the original model ($\Delta x^2 = 492.37$, df = 6, $p < .01$) and further confirmed that adding a correlation coefficient between intrinsic motivation and extrinsic motivation residuals significantly improved the model fit.

Table 2 Fit results for competing models

Indexes	Criterion	Hypothetical model	Modified model for High-Intermediate group
x^2		2504.20	2011.83
df		261	267
p	>.05	<.01	<.01
GFI	>.9	.89	.91
CFI	>.9	.88	.91
RMSEA	<.1	.07	.06
SRMR	<.1	.09	.05
AIC	The smaller the better	2624.20	2131.83
BIC	The smaller the better	2626.16	2133.79

Reliability and validity analysis

After the modified model was established, construct reliability (CR) and average variance extracted (AVE) were calculated using the factor loadings between the latent variables and observed variables. The CR refers to the reliability of the latent variables, while AVE refers to the validity of the latent variables. Fornell and Larcker (1981) suggested that the CR of the latent variables should reach .60, while the AVE should reach .50. It can be seen from Table 3 that the CRs of all the latent variables are above .60, indicating the modified model is acceptable. In terms of AVE, with the exception of the AVE of the attitudes towards the policy being less than .50, the AVEs of all other latent variables are above .50. In practice, it is difficult for the AVE of all latent variables to pass the .50 threshold, as this would mean the factor loadings need to be higher than .71 (.71 squared = .50). For this reason, Fornell and Larcker have proposed that a model with five latent variables can be considered acceptable if among the five latent variables, the AVEs of three to four latent variables are higher than .50, while the AVEs of the other variables are higher than .30 or .40. This indicates that both the questionnaire and the model show good reliability and validity.

Participants

The GEPT Intermediate data were collected in 2016 to confirm the model which has been established based on the results obtained from the study with the High Intermediate data (Authors, 2016). The new data was collected from one university in northern Taiwan that uses the GEPT Intermediate Level Test as its English graduation benchmark. A total of 624 respondents answered the same questionnaire used by the High Intermediate group, of which 570 were valid samples, giving the questionnaire an effective return rate of 91% (282 male, 50%; 288 female, 50%).

Results

Questionnaire responses

Overall, the participants of the Intermediate group showed positive attitudes towards test value (M = 4.21, SD = .1.21), English graduation benchmark (M = 4.17, SD = .86), intrinsic motivation (M = 3.83, SD = 1.07), and extrinsic motivation (M = 4.88, SD = .92). The test created a certain level of test anxiety (M = 4.09, SD = 1.13) in test takers. The Appendix reports the analyses of both ability groups. To compare the two groups, in "test value," the Intermediate group's perception of the GEPT was significantly more positive than the other group. The same tendency was evident in

Table 3 CR (construct reliability) and AVE (average variance extracted)

Latent variables	CR	AVE
TV	.90	.65
AGP	.74	.36
TA	.85	.67
IM	.89	.60
EM	.88	.55

Note: *TV* test value, *AGP* attitudes towards graduation policy, *TA* test anxiety, *IM* intrinsic motivation, *EM* extrinsic motivation

"attitudes towards graduation policy." Yet, a reverse pattern was observed in "test anxiety," which means that the Intermediate group felt more pressure than the High-Intermediate group from taking the GEPT. As for "learning motivation," both ability groups were strong, particularly in the aspect of "extrinsic motivation." The following table summarizes the comparison by each variable between the two ability groups (Table 4).

GEPT test scores

In terms of test performance, the total mean score were 154.07 for the Intermediate group. Together with the Intermediate group's test score analysis, the High-Intermediate's is included in the following table. The mean scores of both groups resemble those of the regular GEPT administrations during 2014–2016 (Table 5).

Parameter calibration and model-fit analysis

Given the large sample size and that the data of questionnaire responses and test scores were approximately normally distributed, the maximum likelihood estimation (MLE) was considered appropriate for the study. Figures in Table 6 and 7 show that the modified model which was established with the GEPT High-Intermediate data fits the GEPT Intermediate data well ($x^2 = 1055.17$, GFI = .90, CFI = .91, RMSEA = .07, SRMR = .05).

Examining the effect of the students' attitudes towards the policy on the various latent variables for the Intermediate group, it can be seen that attitudes towards the policy have a significant positive effect on the perceived test value, with a path coefficient of .53; attitudes towards the policy also have a positive effect on both the extrinsic and intrinsic motivation, with a path coefficient of .62 and .56 respectively. Among the two, the effect on extrinsic motivation is slightly higher than that on intrinsic motivation. There is no significant correlation between attitudes towards the policy and test performance. In terms of test anxiety, results show attitudes towards the policy have a positive effect on test anxiety ($\beta = .21$), meaning that students with a more positive attitude towards the policy tend to experience a increased level of anxiety when taking the test.

The interaction effect between test value and other latent variables was also examined: Test value has a positive effect on test anxiety ($\beta = .22$), but shows no significant relationship with intrinsic motivation and extrinsic motivation. As for the effect of

Table 4 Comparison summary of questionnaire responses between groups

Latent variables	Intermediate group		High-Intermediate group		Sig.
	M	SD	M	SD	
TV	4.21	1.21	3.64	1.20	.00**
AGP	4.17	.86	4.08	.86	.03*
TA	4.09	1.13	3.71	1.29	.00**
IM	3.83	1.07	3.94	.97	.03*
EM	4.88	.92	5.02	.78	.00**

Note: *TV* test value, *AGP* attitudes towards graduation policy, *TA* test anxiety, *IM* intrinsic motivation, *EM* extrinsic motivation
*$p < .05$; **$p < .001$

Table 5 Test performance

	Intermediate group				High-Intermediate group			
	M	SD	Skew	Kurt.	M	SD	Skew	Kurt.
Test scores	154.07	33.27	−.22	−.40	166.93	35.06	−.21	−.44

learning motivation on test performance, results show that only extrinsic motivation has a positive effect ($\beta = .17$) on test performance, while there is no significant relationship between intrinsic motivation and test performance. In other words, in order to improve students' performance on English tests, it is important to increase their extrinsic learning motivation. In addition, there is no relationship between test anxiety and learning motivation. Results also show there is a significant negative effect ($\beta = −.29$) of test anxiety on test performance. The various relationships among learning motivation, test anxiety, and test performance in this study echo those reported by Cheng et al. in their findings.

The path coefficients for both the High-Intermediate and the Intermediate group are shown in the modified model, with the information about the Intermediate group presented in a box (see Fig. 4). The results show that like the GEPT High-Intermediate group, the GEPT Intermediate group perceived the graduation requirement positively and their perception had a positive relationship with test value and learning motivation (both intrinsic and extrinsic). However, four differences between the two sample groups were found. First, the relationships among the variables for the Intermediate group are stronger than those for the High-Intermediate group. Second, unlike the High-Intermediate group, who demonstrated a negative relationship between attitudes towards the graduation requirement and test anxiety, the Intermediate group demonstrated a positive relationship between the two variables. In other words, for students who are less proficient in English, the more positively they perceive the graduation requirement, the greater test anxiety they will feel. Third, an indirect effect of the graduation benchmark on test performance via a different path in each group was observed: intrinsic motivation for the higher group and extrinsic motivation for the lower group.

Conclusions and discussion
This section discusses the results drawn from the two studies corresponding to the research questions.

Table 6 Fit results for GEPT Intermediate data

Indexes	Criterion	Modified model for Intermediate group
x^2		1055.17
df		245
p	>.05	<.01
GFI	>.9	.90
CFI	>.9	.91
RMSEA	<.1	.07
SRMR	<.1	.05

Table 7 Hypotheses verification

Hyp.	Path	β	
Students' attitudes towards graduation policy as a predictor		I	HI
H$_1$	Have a significant positive effect on test value	.53**	.36**
H$_2$	Have a significant positive effect on extrinsic motivation	.62**	.41**
H$_3$	Have a significant positive effect on test performance	ns.	ns.
H$_4$	Have a significant positive effect on intrinsic motivation	.56**	.39**
H$_5$	Have a significant negative effect on test anxiety	.21**	−.17**
Students' perception of test value as a predictor			
H$_6$	Has a significant positive effect on extrinsic motivation	ns.	ns.
H$_7$	Has a significant positive effect on intrinsic motivation	ns.	ns.
H$_8$	Has a significant positive effect on test performance	ns.	ns.
H$_9$	Has a significant positive effect on test anxiety	.22**	.32**
Students' learning motivation as a predictor			
H$_{10}$	Extrinsic motivation predicts test performance positively	.17**	ns.
H$_{11}$	Intrinsic motivation predicts test performance positively	ns.	.16**
Interrelationship between test anxiety and learning motivation			
H$_{12}$	Test anxiety influences extrinsic motivation	ns.	.12**
	Extrinsic motivation influences test anxiety	ns.	ns.
H$_{13}$	Test anxiety influences intrinsic motivation	ns.	.06*
	Intrinsic motivation influences test anxiety	ns.	ns.
Test anxiety			
H$_{14}$	Has a significant negative effect on test performance	−.29**	−.36**

*$p < .05$; **$p < .001$; ns. $p > .05$

RQ1: What are students' attitudes towards the English benchmark policy for graduation in Taiwan's universities? Are there differences between the High-Intermediate group and the Intermediate group?

The findings show that university students, regardless of whether they belonged to the High-Intermediate group or the Intermediate group, generally hold a positive attitude towards the English graduation benchmark policy, which is consistent with the results reported by Chen and Liu (2007), Liao (2010), and Shih (2008). Results further reveal that the Intermediate group shows more positive attitudes towards the graduation requirement than the High-Intermediate group. In addition, students' attitudes towards the graduation requirement have affected students' perceptions of test value, test performance, and learning motivation in a positive manner. In the aspect of learning motivation, the results of our study show that students' extrinsic motivation is stronger than intrinsic motivation for learning English. A study by Wu and Lin (2009) discussing college students' motivation for English learning showed that "instrumental motivation" scored the highest among different types of motivation. This finding echoes the sociological viewpoint of instrumental motivation proposed by Gardner (1985), which refers to how learners study a language for the purpose of progress and development, such as career advancement or further studies. Items in this study that are related to extrinsic

Fig. 4 The modified model for both groups
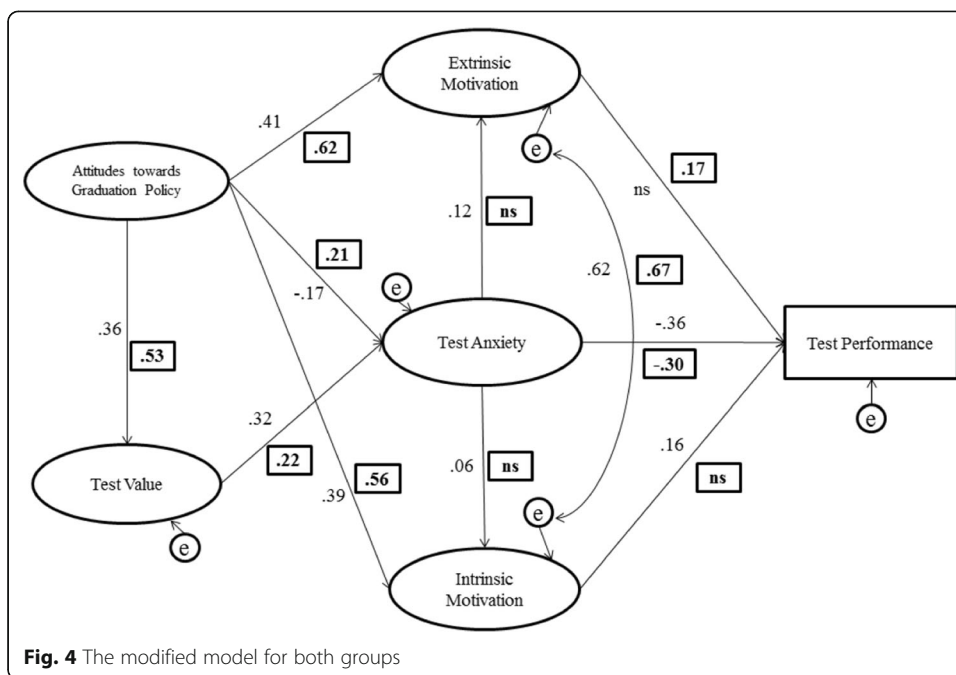
motivation include "I study English for the purpose of finding a better job" and "because I would like to perform better at school or at work."

Despite the similarities in students' attitudes towards the graduation requirement between the two groups, results show the Intermediate group's attitudes towards the policy have a positive effect on test anxiety ($\beta = .21$), whereas the High-Intermediate group's attitudes towards the policy have a negative effect on test anxiety ($\beta = -.17$). This suggests that students with a lower level of English proficiency are more sensitive to the pressure from the graduation requirement and more likely to experience an increased level of anxiety when taking the test. In view of this, schools and teachers should be aware of the risk of increasing students' anxiety when implementing the graduation benchmark requirement.

RQ 2: What are the relationships between students' attitudes towards the graduation benchmark policy, their test performance, and their perceptions of learning motivation, test value, and test anxiety? Are there differences between the High-Intermediate group and the Intermediate group?

SEM results show that the attitudes of university students towards the English graduation benchmark policy positively impact their perceived test value and their learning motivation. However, there is no significant relationship between the attitudes towards the policy and test performance. The attitudes of students towards the English benchmark policy have a positive effect on their motivation for learning English, and this motivation is in turn reflected in their behavior. According to the theory of planned behavior (Ajzen 1985), in order for students to take the initiative to study English and improve their English ability, it is important to change their attitudes towards learning English, making it fun and interesting rather than for the sole purpose of passing the GEPT and meeting the graduation requirement.

In addition, the results show that test value has a positive effect on test anxiety. Students who feel that the test results are more important for the purpose of obtaining a degree, a scholarship, or a job are more likely to feel anxious when facing a test. However, the results concerning the relationship between students' attitudes towards the graduation requirement and test anxiety are more complicated. Among the High-Intermediate group, students who had a positive attitude towards the English graduation benchmark were less likely to feel anxious in the face of the test. On the contrary, a positive relationship between students' attitudes towards the graduation requirement and test anxiety was evident among the Intermediate group. In other words, among the Intermediate group, those who perceived the graduation requirement more positively felt more anxious about the test. Regardless of test groups, test anxiety has a negative effect on test performance; students who were more anxious about the test showed poorer performance (Cassady and Johnson 2002). In this regard, we suggest that schools provide relevant resources to help students, especially those students with lower levels of English proficiency, to reduce their test anxiety, for example, inviting the testing organization behind the GEPT to provide more information about how to prepare for the test or give mock tests.

Another noticeable difference between the two ability groups lies in the aspect of motivation. While the Intermediate group students were inclined to be driven by extrinsic motivation, the High-Intermediate group students tended to be more intrinsically motivated. The finding is consistent with Maslow's hierarchy of needs (1943 and 1954) in that before learners can be intrinsically motivated, they must first be provided with external rewards such as good grades. In view of this, the English graduation benchmark requirement can be considered a useful measure to motivate students to learn, especially for those who are at a lower level of English proficiency. But, it is absolutely insufficient for schools to simply impose the graduation requirement upon their students. More importantly, schools must provide necessary remedial course programs to help the less proficient students to build their confidence in improving their English proficiency and at the same time in coping with the graduation benchmark requirement. Different from the Intermediate group students, the High-Intermediate group students may not need a graduation benchmark requirement to encourage them to learn more English because they are willing and eager to learn new material for self-actualization. However, their schools should ensure that students' English learning experience is more meaningful, and they can learn deeper to fully understand it.

Finally, schools should be aware that implementing the English graduation benchmark policy is only one of many ways of assessing students' English ability and not the sole reason to set up English courses. At the same time, we believe universities should consider the differences in students' background, learning objectives, and learning status when establishing the English graduation benchmark policy. Since students possess different qualifications, personal traits, and some may not have kept up with the school's curriculum in their learning process, they should be allowed to set their own expectations and English learning goals according to their ability, resources, and opportunities. In short, although the implementation of the English graduation benchmark policy is based on good intentions, as shown by related studies (Her et al. 2014),

applying a fixed standard established by the school or the department may not be suitable for every student.

The questionnaire used in this study was adapted from the questionnaire "Learning Motivation, Test Anxiety and English Test Performance" designed by Cheng et al. (2014). We then compared the result of this study with that of Cheng et al. and attempted to discuss possible reasons for various similarities and differences. Both studies show that the test purpose is positively correlated with learning motivation and test anxiety and that test value is positively correlated with learning motivation. However, the result of our study showed that test value is positively correlated with test anxiety, while Cheng et al. showed that there was no significant correlation between test value and anxiety. This difference may be due to the fact that the latter study analyzed the results of all three tests, namely Taiwan's GEPT, China's CET, Canada's CAEL, as a whole instead of individually. Another reason may be due to the distinct traits of the sampled body. The GEPT samples gathered by Cheng et al. were from students who passed the listening and reading tests and formed a homogeneous ability group, while the present study did not limit the samples to those who passed the listening and reading tests and can be considered a more heterogeneous group.

Limitations

Although the two studies are the first attempt that draws on a large sample of university students to explore the complex relationships among the English graduation benchmark policy, test value, learning motivation, test anxiety, and test performance, it has a number of limitations. First, these studies only used data of students from three universities that have established passing the GEPT High-Intermediate Level Test (equivalent to the CEFR B2) and the GEPT Intermediate Level Test (equivalent to the CEFR B1), as their English graduation benchmark policy. The sampled students, especially the High-Intermediate group, tend to show a strong learning motivation and autonomous learning behavior; thus, their attitudes towards the graduation benchmark policy are also positive and optimistic. Consequently, more data are needed to support whether the findings of this study can be applied to universities that use other forms of English proficiency testing tools as their graduation benchmark requirement. In view of this, we suggest that future researchers can gather data from a large sample of universities that use the other two levels of the GEPT, i.e., elementary (equivalent to the CEFR A2) and advanced (equivalent to the CEFR C1), or other tests (e.g., TOEIC, IELTS) as their graduation benchmark requirement. Second, in addition to using SEM to analyze the relationships among latent variables, hierarchical linear modeling (HLM) can also be applied when there is a larger amount of data. Factors such as the student's department and school can be integrated into the analysis, as these factors are important elements that are likely to influence one's attitude towards the English graduation benchmark policy. Third, open-ended questions should be included in the questionnaire to gain a more comprehensive understanding of the views of university students towards the English graduation benchmark policy, as well as how their English learning process is influenced by the implementation of this policy.

Appendix

Table 8 Descriptive statistics of questionnaire responses

Test value	Intermediate group				High-Intermediate group			
	M	SD	Skew	Kurt.	M	SD	Skew	Kurt.
1. How important is my GEPT score for applying to graduate schools.	3.82	1.59	−0.33	−0.85	3.13	1.59	0.11	−1.11
2. How important is my GEPT score for gaining professional certification.	4.40	1.33	−0.61	−0.20	3.78	1.42	−0.30	−0.63
3. How important is my GEPT score for obtaining an academic award/scholarship.	4.22	1.38	−0.45	−0.48	3.72	1.44	−0.21	−0.75
4. How important is my GEPT score for obtaining a job.	4.43	1.38	−0.73	−0.17	3.97	1.39	−0.43	−0.54
5. How important is my GEPT score for studying/working abroad.	4.21	1.51	−0.52	−0.67	3.59	1.62	−0.10	−1.11
Attitudes towards graduation policy	**Intermediate group**				**High-Intermediate group**			
	M	SD	Skew	Kurt.	M	SD	Skew	Kurt.
1. I understand that my university encourages students to take the GEPT in order to enhance their English proficiency.	4.48	1.14	−0.68	0.53	4.32	1.21	−0.60	0.11
2. I think it is reasonable to require college students to pass the GEPT High-Intermediate Level Test.	4.55	1.22	−0.80	0.51	4.45	1.26	−0.70	0.12
3. I think the GEPT is fair and reliable.	4.40	1.06	−0.46	0.37	4.27	1.09	−0.33	−0.15
4. I agree that Taiwan should develop its own English proficiency tests, for instance the GEPT.	4.00	1.26	−0.41	0.03	3.80	1.29	−0.24	−0.44
5. I think my university should not require students to take the GEPT.	3.43	1.47	−0.22	−0.87	3.54	1.36	−0.16	−0.66
Test anxiety	**Intermediate group**				**High-Intermediate group**			
	M	SD	Skew	Kurt.	M	SD	Skew	Kurt.
1. I am under a lot of pressure to get good scores on this GEPT.	4.16	1.29	−0.36	−0.45	3.66	1.53	−0.07	−0.98
2. The GEPT listening section makes me feel anxious.	4.10	1.33	−0.36	−0.50	4.01	1.46	−0.31	−0.81
3. The GEPT reading section makes me feel anxious.	4.01	1.28	−0.28	−0.45	3.45	1.41	0.11	−0.74
Intrinsic motivation	**Intermediate group**				**High-Intermediate group**			
	M	SD	Skew	Kurt.	M	SD	Skew	Kurt.
I study English… 1. Because I enjoy reading English literature.	4.73	1.17	−0.79	1.08	4.93	1.08	−0.86	0.30
2. For the satisfaction I gain from learning new things.	5.03	1.00	−0.98	0.89	5.23	0.89	−1.02	1.07
3. Because I want to know more about English communities and their way of life.	4.95	1.07	−0.95	0.98	5.12	0.98	−0.94	0.75
4. Because I like to understand difficult topics in English.	4.78	1.04	−0.81	0.96	4.91	0.96	−0.65	−0.01
5. For the satisfaction I feel when I am completing a difficult task in English.	4.90	1.07	−1.01	1.08	4.90	1.08	−0.95	1.03
6. For the excitement I feel when hearing English.	3.41	1.27	0.10	0.31	3.41	1.31	0.11	−0.43
Extrinsic motivation	**Intermediate group**				**High-Intermediate group**			
	M	SD	Skew	Kurt.	M	SD	Skew	Kurt.
I study English… 1. Because I want to speak more than one language.	4.20	1.18	−0.37	−0.29	4.28	1.17	−0.27	−0.48
2. Because I think it helps my personal development.	4.20	1.31	−0.43	−0.46	4.45	1.21	−0.45	−0.51

Table 8 Descriptive statistics of questionnaire responses (Continued)

3. Because I want to be able to speak English.	3.36	1.28	0.22	−0.27	3.52	1.29	0.06	−0.46
4. Because I would like to perform better at school or at work.	4.01	1.32	−0.30	−0.48	4.07	1.27	−0.24	−0.59
5. Because I want to get a better job.	3.78	1.28	−0.04	−0.56	3.87	1.27	−0.06	−0.53

Authors' contributions

Both authors were responsible for the research design, literature review, and data interpretation. The first author wrote the manuscript; the co-author analyzed the data and checked and edited the manuscript. Both authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Ajzen, I. (1985). From intentions to actions: a theory of planned behavior. In J. Kuhl & J. Beckman (Eds.), *Action-control: from cognition to behavior* (pp. 11–39). Heidelberg: Springer.

Authors. (2016). 探究大學英語能力畢業門檻與學習動機、測驗焦慮、測驗表現之關係 [An exploratory study on the relationships among student attitudes towards the English graduation benchmark policy, learning motivation, test anxiety, and test performance]. *English Teaching & Learning, 40*(3, (Fall 2016)), 61–86.

Bodas, J., & Ollendick, T. H. (2005). Test anxiety: a cross-cultural perspective. *Clinical Child and Family Psychology Review, 8*(1), 65–88.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27*, 270–295.

Chan, S. H. C., Wu, R. Y. F., & Weir, C. J. (2014). *Examining the context and cognitive validity of the GEPT advanced writing task 1: a comparison with real-life academic writing tasks. LTTC-GEPT Research Report No. RG-03*. Taipei: LTTC.

Chang, W. C. (2005). 台灣的英語教育:現況與　思 [Taiwan's English language education: present and future]. *Educational Resources and Research, 69*, 129–144.

Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal of Educational Psychology, 97*, 268–274.

Chen, S. C. (2014). 全球化下的臺灣英文教育:政策、教學及成果 [English education in Taiwan in the age of globalization: policy, practice and results]. *Educators and Professional Development, 31*(2), 7–20.

Chen, S. W., & Liu, G. Z. (2007). *The impact of English exit policy on motivation and motivated learning behaviors* (Proceedings of the twenty-fourth international conference on English teaching and learning in the Republic of China, pp. 104–110). Taipei: Crane.

Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: the CAEL, CET and GEPT. *TESOL Quarterly, 48*, 300–330.

Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment.* Cambridge: Cambridge University Press.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior.* New York: Plenum.

Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics, 21*, 43–59.

Eccles, J. S., & Wigfield, A. (2002). Motivational belief, values, and goals. *Annual Review of Psychology, 53*, 109–132.

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *Modern Language Journal, 89*, 206–222.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: an introduction to theory and research.* Reading, MA: Addison-Wesley.

Fornell, C. R., & Larcker, F. F. (1981). Structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39–51.

Gardner, R. C. (1985). *Social psychological and second language learning: the role of attitudes and motivation.* London, Ontario, Canada: Edward Arnold.

Her, O. S., Chou, C., Su, S. W., Chiang, K. H., & Chen, Y. H. (2013). 我國大學英語畢業門檻政策之檢討 [A critical review of the English benchmark policy for graduation in Taiwan's universities]. *Education Policy Forum, 86*(3), 1–30.

Her, O. S., Liao, Y. H., & Chiang, K. H. (2014). 論現行大學英語畢業門檻的適法性─以政大法規為實例的論證 [The appropriateness of the English benchmark requirement for graduation in Taiwan's universities––the case of National Chengchi University]. *NCCU Law Review, 139*, 1–64.

Huang, L. H. (2010). 英檢畢業門檻對大學生的英語焦慮、英語學習動機、英語學習策略之影響研究 *[English graduation threshold among the college students' EFL anxiety, motivation and learning strategies]*. Hualien, Taiwan: Tzu Chi University.

In'nami, Y. (2006). The effects of test anxiety on listening performance. *System, 34*, 317–340.

Kunnan, A. J. (1995). *Test taker characteristics and test performance*. Cambridge, England: Cambridge University Press.

Liao, Y. H. (2010). 技專校院學生英語畢業門檻之態　初探 [An exploratory investigation of attitude towards the English benchmark policy for graduation among technological and vocational university students]. *Journal of National Huwei University of Science & Technology, 29*(3), 41–60.

Liao, Y. F. (2016). Investigating the score dependability and decision dependability of the GEPT listening test: a multivariate generalizability theory approach. *English Teaching & Learning, 40*(1), 79–111.

Liu, M. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *Modern Language Journal, 92*, 71–86.

Maslow, A. (1943). A theory of human motivation. *Psychological Review, 50*, 370–396.

Maslow, A. (1954). *Motivation and personality*. New York: Harper.

McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing, 18*(4), 333–349.

Meijer, J. (2001). Learning potential and anxious tendency: test anxiety as a bias factor in educational testing. *Anxiety, Stress and Coping, 14*, 337–362.

Noels, K. A. (2005). Orientations to learning German: heritage language learning and motivational substrates. *Canadian Modern Language Review, 62*, 285–312.

Putwain, D. W. (2007). Test anxiety in UK schoolchildren: prevalence and demographic patterns. *British Journal of Educational Psychology, 77*, 579.

Pyun, D. O. (2013). Attitude toward task-based language learning: a study of College Korean language learners. *Foreign Language Annals, 46*, 108–121.

Ruthig, J. C., Perry, R. P., Hall, N. C., & Hladkyj, S. (2004). Optimism and attributional retraining: longitudinal effects on academic achievement, test anxiety, and voluntary course withdrawal in college students. *Journal of Applied Social Psychology, 34*, 709–730.

Ryan, R. M., & Brown, K. W. (2005). Legislating competence: high-stakes testing policies and their relations with psychological theories and research. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 354–372). New York, NY: Guilford Press.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology, 25*, 54–67.

Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: a self-determination theory perspective on high-stakes testing. *Theory and Research in Education, 7*, 224–233.

Shih, C. M. (2008). Critical language testing: a case study of the General English Proficiency Test. *English Teaching & Learning, 32*(3), 1–34.

Shohamy, E. (2000). *The power of tests: a critical perspective on the uses of tests*. London: Longman.

Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development, 66*, 209–223.

Tsai, C. R. (2010). 英語聽力表現與英語能力、學習動機、聽力焦慮以及後設認知之關係研究 *[EFL learners' listening performance: its relationship to English proficiency, motivation, listening anxiety, and metacognitive awareness]* (Unpublished master dissertation). Taipei, Taiwan: The University of Taipei.

Weir, C., Chan, S. H., & Nakatsuhara, F. (2013). *Examining the criterion-related validity of the GEPT advanced reading and writing tests: comparing GEPT with IELTS and real-life academic performance. LTTC-GEPT Research Report No. RG-01.* Taipei: LTTC.

Wu, R.Y.F. (2011). Establishing the validity of the General English Proficiency Test reading component through a critical evaluation on alignment with the Common European Framework of Reference. Unpublished doctoral thesis, University of Bedfordshire.

Wu, J. (2016). A socio-cognitive approach to assessing speaking and writing: the GEPT experience. *JLTA (Japan Language Testing Association) Journal, 19*, 3–11.

Wu, R. Y. F., & Liao, C. H. Y. (2010). Establishing a common score scale for the GEPT Elementary, Intermediate, and High-Intermediate Level listening and reading tests. In T. Kao & Y. Lin (Eds.), *A new look at language teaching and testing: English as subject and vehicle—selected papers from the 2009 LTTC International Conference on English Language Teaching and Testing* (pp. 309–329). Taipei: Language Training and Testing Center.

Wu, Y. S., & Lin, C. P. (2009). 大學生英語學習環境、學習動機與學習策略的關係之研究 [A study of the relationship among English learning environment, learning motivation and learning strategies of college students]. *Journal of Taipei Municipal University of Education: Education, 40*(2), 181–222.

Wu, J. R. W., & Wu, R. Y. F. (2010). Relating the GEPT reading comprehension tests to the CEFR. In W. Martyniuk (Ed.), *Aligning tests with the CEFR, Studies in Language Testing 33* (pp. 204–224). Cambridge: Cambridge University Press.