GEPT and English Language Teaching and Testing in Taiwan

The General English Proficiency Test (GEPT) is a five-level criterion-referenced EFL testing system implemented in Taiwan to assess the general English proficiency of EFL learners. In 1999, with the aim of encouraging the general study of English and to result in beneficial washback effects on the teaching and learning of English, the Ministry of Education lent its support to the Language Training and Testing Center (LTTC) in the development of the GEPT. Throughout a decade of efforts, the GEPT has won popular recognition in Taiwan. To date, more than 3.7 million Taiwanese have taken the test. The paper first documents the evolution of the GEPT from the perspectives of test development and validation. The paper then provides an overview of how GEPT scores are used in both educational and professional domains and discusses several key issues and problems that have emerged due to the new context introduced by the GEPT. Finally, the paper outlines how the GEPT will address the challenges it faces in pursuit of continuity and innovation in the years to come.

INTRODUCTION

Ten years ago, the history of EFL tests in Taiwan changed due to the introduction of a new testing system, the General English Proficiency Test (GEPT). The history of the GEPT over the past decade demonstrates ongoing collective efforts to provide a fair test for EFL learners as well as encourage English language learning in Taiwan. In this paper, I will first introduce the GEPT from the perspectives of its development process and test validation efforts. I will then provide an overview of how GEPT scores are being used in both the educational and professional domains. I will also discuss a number of salient issues and problems that have emerged due to the new context introduced by the GEPT. Finally, as to the way forward, I will conclude the paper by suggesting how the GEPT can address the challenges it faces in pursuit of continuity and innovation in the years to come.

DEVELOPMENT OF THE GEPT

The GEPT started as an in-house research project at the Language Training and Testing Center (LTTC) in 1997. The LTTC, a non-profit cultural and educational foundation, was established in 1951 under a remit to meet the needs of social and

economic development with work on research, development, and services in the language teaching and testing fields. With the aim to develop a public language test that could bring beneficial washback effects to the EFL classroom in Taiwan, the LTTC invited a number of well-established EFL educators from different parts of the country to form the GEPT Advisory Board and the GEPT Research Committee. Two years later, the Ministry of Education (MoE) recognized that these efforts were in accord with its promotion of lifelong learning and therefore decided to sponsor the development of the GEPT, a five-level criterion-referenced testing system. The first GEPT (intermediate level) was made available to the public in 2000, followed by the elementary and high-intermediate levels in 2001, and the top two levels (advanced and superior) in 2002. The objective to bring beneficial washback effects to EFL education in Taiwan was addressed from the outset of the test development process, and it was later reflected in the test design and delivery system. Several key features of the GEPT are as follows.

Creating a Five-level Criterion Referenced EFL Testing System

The GEPT is a five-level criterion-referenced EFL testing system. The table in the appendix provides a general description of the GEPT levels and the test tasks. The levels also correspond to the major stages in English competency in the educational system in Taiwan, with the hope that the junior high graduates can have the English competence expected at the elementary level; senior high graduates at the intermediate level; non-English major college/university graduates at the high-intermediate level; English major college/university graduates at the advanced level. The highest level is the superior level, which approaches the competency of well-educated native speakers of English. For the first two levels of the GEPT, test content is guided by the national curriculum objectives of junior high schools and senior high schools, respectively. The upper three levels of the GEPT, for which no national curriculum exits, were developed based on the expectations of stakeholders in English education in Taiwan as identified through textbook analysis, needs analysis, and teachers' forums. Each level consists of four components: listening, reading, writing, and speaking. Skill-specific level descriptors can be found at the GEPT website (http://www.lttc.ntu.edu.tw).

The GEPT is intended for students and other individuals from all walks of life. Test-takers may register for whichever level they feel is appropriate for them except for the Superior level, which is only offered on an institutional basis. The scoring of the GEPT is non-compensatory, thus test-takers are required to pass all four of the components in order to obtain a passing certificate. However, influenced by the fact

that Taiwanese learners are less familiar with the assessment of productive skills (speaking and writing) that are not included in schools exams or large-scale entrance exams, the GEPT is divided into two stages. That is, at all levels below the superior level (an integrated test), test-takers must pass the listening and reading tests (the first stage) in order to qualify to register for the writing and speaking tests (the second stage). Test-takers must score 160 or above, with a minimum part score of 72 on both listening and reading (60% of 120 score points) to pass the first stage. To pass the second stage, test-takers at the first three levels must score Band 4 or above on both speaking and writing, while at the two higher levels Band 3 is the minimum passing standard. Only those who pass all four test components at the same level are awarded a certificate indicating their English proficiency has reached that level. The two-stage design of the test has been criticized for providing speaking and writing assessments only to those test-takers who have passed the listening and reading assessments. However, the restriction may be justified because most of the test-takers who fail the listening and reading papers can not pass the speaking and writing tests either (LTTC, 1999). Moreover, test-takers actually benefit from the two-stage design because they do not pay to take the speaking and writing tests until they have passed the first-stage tests.

Promoting Life-Long Learning

Since the first administration of the GEPT in 2000, approximately four million Taiwanese people have taken it to date. The following table gives the profile of the GEPT test-takers at each level except the Superior Level. The figures show that students make up a great proportion of the total population of GEPT test-takers, a trend which is more noticeable at the lower levels. However, it is significant that the percentage of non-student test-takers increases dramatically at High-Intermediate Level and Advanced Level. This confirms that the GEPT is being used not only by students but also non-students as intended. Moreover, the average age of test-takers in the table shows an increase as the level advances. This, as Roever and Pan (2008) suggested, implies a trend towards more people taking the GEPT across their lifetime and presumably studying more English. As described earlier, each GEPT level is targeted at the learners whose English proficiency corresponds to that expected at each major educational stage in Taiwan. Such links between GEPT levels and educational stages are supported in the most recent GEPT report (LTTC, 2009), which shows that the majority of the student test-takers at each test level were learners studying at the target educational stage.

Assessing Listening Comprehension and Oral Proficiency as Compulsory Components

The GEPT was designed as a skill-based test battery assessing both receptive (listening and reading) and productive (writing and speaking) skills. The design was chosen in response to the concerns of educators and employers from various industries about the general lack of ability to communicate in English. From the perspective of washback effect, the lack of the assessment of listening and speaking skills in the English paper of Taiwan's current high school and university entrance exams on the basis of practicality has received the largest share of the blame for creating this problem. Moreover, as international communication becomes more important in the world today, and more information comes through all kinds of media in the spoken form, it is vital to emphasize the importance of the ability to communicate orally. Therefore, the GEPT aims to assess not only learners' knowledge of English but also their ability to use English in real life situations.

The administration of the GEPT listening and speaking tests was considered a breakthrough in Taiwan's English language testing a decade ago. It was indeed an extremely challenging task for the LTTC to initiate large-scale listening and speaking tests. To achieve the task, the LTTC had to overcome numerous practical constraints. The major problems included securing enough testing rooms with good acoustic condition for listening tests, recruiting sufficient number of qualified assessors for speaking tests, and reliably scoring test-takers' oral performance. Owing to the necessity of assessing listening comprehension and oral proficiency, the GEPT has been fortunate to receive strong support from schools across the country, allowing it to overcome the practical difficulties and accomplish these aims. Schools provide assistance in administering the listening test by renting their facilities and complying with the standardized operation procedures for the GEPT as specified by the LTTC.

Conducting a speaking test on a large scale is by no mean an easy task. A major consideration in developing speaking proficiency component for use within the GEPT program was that it be amendable to large-scale standardized administration at GEPT test centers island-wide. For the first three levels of the GEPT speaking test, each of which was estimated to have over 20,000 test-takers in each administration, it was considered too costly and impractical to use face-to-face interviews, involving direct interaction between the test-takers and an interlocutor who would have had to be a trained native or highly proficient non-native speaker of English. Therefore, a semi-direct speaking test conducted in a digital language laboratory environment was considered more feasible for the first three levels of the GEPT. On the other hand, a direct speaking test was considered more manageable for the two higher levels which

were estimated to have a smaller number of test-takers. A description of task types employed in each speaking test is provided in the appendix.

Teachers of English at the senior high school and university levels are recommended by their schools to apply to work as an examiner or assessor. Before they can become certified GEPT assessors, they are required to go through a rigorous training process and must demonstrate a good degree of rater consistency. The AERA/APA/NCME Standards (1999) make it clear that when the scoring of a test involves judgments by examiners or raters, it is important to consider reliability in terms of the accuracy and consistency of the ratings that are made. Weir and Wu (2006) reported that multi-faceted Measurement (MFRM) has been used as part of the GEPT system which monitors rater variability, parallel-form reliability, and potential effects arising from different raters assessing different performances. These efforts have resulted in a good degree of reliability of the GEPT speaking assessment. While establishing reliability is an essential step for operational high-stakes tests, we must note that, for an operational high-stakes test like the GEPT, providing evidence of reliability may not be sufficient in itself for establishing its validity.

VALIDATION EFFORTS

With the aim of serving as a fair and reliable testing system for use by EFL learners in Taiwan, a major consideration throughout the development of the GEPT was that the test must be amenable to large-scale standardized administrations at GEPT test centers islandwide. The GEPT has demonstrated a good degree of objectivity and consistency in scoring and marking. Measures of the reliability of the GEPT are mostly in the high .80 range, which are comparable to the reliability figures reported by other large-scale tests (Roever & Pan, 2008).

To strike a balance between reliability and validity, the LTTC has sought to provide evidential support for the claims about the validity of the GEPT. Numerous GEPT validation studies have been conducted by LTTC research staff to investigate various aspects of validity, including but not limited to analyses of GEPT test performance data and test-takers' feedback (Wu, 2002a, 2002b), scoring reliability (Wu, 2003), concurrent validity (LTTC, 2003), content validity (Wu, 2003), criterion-related validity (LTTC, 2005), a corpus analysis of written performance (Kuo, 2005), and test impact (Wu & Chin, 2006). All validation reports are available on the GEPT website; however, it should be noted that most of the reports published earlier were written in Chinese and were published as technical monographs rather

than research papers. With an on-going commitment to providing stakeholders with information about the quality of the GEPT, an increasing number of GEPT validation studies have been presented in English and published as refereed articles in books and journals. For example, Wu (2005) and Weir and Wu (2006) investigated empirical construct of three GEPT speaking test forms and established pararellel-form reliability at the task level in the speaking test in terms of code complexity (lexical and syntactical difficulty), cognitive complexity (content familiarity), and communicative demand (time pressure). By means of both qualitative and quantitative analyses, results show that the test forms could be considered parallel.

Another study mapped the GEPT levels against the CEFR Framework as an external validity criterion, aiming to respond to the MoE's request that all tests should be related to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CoE, 2001). The study followed the 'internal validation' procedure presented by the Manual for Relating Language Examinations (CoE, 2003), including familiarization, specification and standardization (judgment session only). For the purpose of the mapping study, a total of 70 GEPT reading comprehension test questions that exemplify the test constructs of the different levels of the GEPT reading comprehension tests were compiled. The standardization was carried out in April 2007, with the participation of a total of 15 EFL teaching professionals. All participants were first trained to relate their interpretations of the CEFR levels to the calibrated sample items provided by the CoE. During the judgment session, the GEPT tasks and items were presented to the participants who were then asked to determine the minimum CEFR level needed by a candidate to successfully answer the item on a given level of the GEPT. The judgment results show that the first four levels of the GEPT, from Elementary Level to Advanced Level, correspond to the CEFR A2 to C1 levels, with a generally satisfactory rater agreement of 0.91 (Wu & Wu, in press).

Recently, a study was carried out to establish the internal validity of the GEPT. Wu and Liao (2010) calibrated the GEPT listening and reading tests on a common scale by linking different levels of the GEPT vertically onto a common score scale on the basis of Item Response Theory. The results show a pattern of increasing difficulty across the levels, providing evidence to support the internal validity of the GEPT level framework (Wu & Liao, 2010).

Noting that the GEPT level descriptors and can-do statements were not empirically derived, the LTTC began to work on validating the GEPT descriptors through an investigation into the relationship between the GEPT scores and two alternative validity criteria for tests of English language proficiency, namely teacher assessments and student self-assessments (Wu, 2010). The assessments were elicited

by can-do statements adapted from the GEPT can-do descriptors. The results of the study showed that teacher assessment correlated moderately well with students' GEPT scores (.55 for listening and .61 for reading), providing a fair degree of support for the validity of the GEPT scores as an indicator of English language proficiency. Therefore, more validation studies of this kind should be carried out to yield stronger evidence to empirically support the validity of the GEPT descriptors.

Despite the considerable GEPT validation research in the first decade, given the test's wide recognition in Taiwan and the fact that its scores are increasingly used for high-stakes decisions, investigation of construct validity, consequential validity, and social and educational impact remain to be done.

THE INFLUENCES OF THE GEPT

As Bachman (1990) notes, the uses of language tests are heavily influenced by educational and social needs. This has been the case in the growing use of the GEPT in Taiwan in the past years. To strengthen Taiwan's international competitiveness, there has been a strong identification in recent years of a need to acquire competency in English, and this aim is supported by government policies regarding the use of English language assessment. In 2005, the MoE adopted the CEFR as a means to establish a common standard of English proficiency. Following that move, the MoE required students, teachers of English, and civil servants to demonstrate a level of English proficiency by taking an external English language test. With governmental support, a score for an external English test has become influential in qualifying for graduation from college/university or for a job promotion. As a result, the English language testing population has increased enormously in the past years. Among the various English language tests (e.g., Cambridge Main Suite, IELTS, TOEFL, and TOEIC) used in Taiwan, the GEPT is the most widely used, and its impact, both intended and unintended, is evident. The following are some examples illustrating the use of the GEPT and the problems that have arisen.

Exit Requirement at the Tertiary Level
To enable college and university students to cope with global competitiveness, the MoE has encouraged universities and colleges to adopt English ability as an exit requirement for students. As a result, an increasing number of colleges and universities now require graduates to pass the GEPT at a certain level before they can be awarded a diploma. Pan (2007) found that nearly 30% of technical colleges use the GEPT or other tests as exit tests for non-English majors. However, as Roever and Pan questioned (2008), it is unclear whether the use of the GEPT as an exit requirement

results in pedagogical changes in English curriculum that enable better learning outcomes, or whether it leads to negative washback from test use, e.g., teaching to the test.

As an example to illustrate how the GEPT is used to encourage English learning at the tertiary level, National Taiwan University (NTU) has implemented the use of the GEPT High-Intermediate Level as the target level of English proficiency for students in its curriculum for non-English majors. Unlike other universities that simply use the GEPT as an exit requirement, NTU places sophomores into the Online English Program (OEP) based on their GEPT High-Intermediate test results. Students who attain the passing level of the GEPT High-Intermediate may be exempted from attending the program. However, students who are below the target level of English proficiency are considered to have an immediate need for the OEP. Moreover, the objectives of the curriculum correspond to the level descriptors of the GEPT High-Intermediate. According to NTU, more concrete, task-oriented curriculum planning has encouraged students to study more English. Moreover, it has provided both students and teachers with clearer goals for learning and teaching, respectively (Y. M. Ma, personal communications, July 10, 2009). Although there is a positive trend, whether the students who have completed in the OEP program can reach the proficiency level at the GEPT High-Intermediate as expected is still in question until necessary empirical studies have been conducted. Therefore, the consequences of using the GEPT at the tertiary level are a pressing task for GEPT research.

Young Learners

Taiwan has begun providing English language instruction at elementary schools in recent years. In the meantime, the government has avoided traditional achievement tests (e.g., standardized tests) at the elementary school level; instead alternative assessments (e.g., interviews, classroom observations, student portfolios) are encouraged. However, due to the fact that Taiwan is an exam-driven society, English is treated as a high-stakes subject in Taiwan; the GEPT scores are considered by parents to have a large impact on their children's success in entering a better high school. As a result, parents, especially those with higher socio-economic status, were eager to register their children for the GEPT; approximately 20,000 children at the elementary school level took the GEPT each year before 2006. To meet the parents' expectations, language schools started providing young learners with test preparation programs rather than the English courses regularly offered to children. Having noted the problem, serious concern over test misuse and negative consequences was expressed by the LTTC and a great number of conscientious EFL educators. To

remedy this situation, in 2006 the LTTC proposed that learners at the elementary school level be barred from registering for the GEPT. The proposal was soon accepted by the GEPT Examination Board and the MoE later that year. At present, young learners with a strong interest in taking a standardized test to assess their English ability are directed to one that is specially designed to cater for young learners' needs, e.g., Cambridge Young Learners English Test.

Increasing Teachers' Understanding of Testing and Assessment

As noted earlier, the GEPT was designed from the outset to promote English language teaching and learning in Taiwan. For the GEPT to achieve beneficial washback effects, successful interactions with teachers is essential, and ultimately a feedback loop between teaching and testing should be established (Genesee & Upshur, 1996). Since the GEPT's launch in 2000, positive influences on teachers' understanding of testing and assessment have been observed through efforts in communicating with local teachers by offering teachers' workshops and providing teachers with word lists and teaching activities for classroom materials. For example, in a study (Wu, 2008) investigating teachers' views of the GEPT, some teachers reported that they had learned new ideas about assessment from the GEPT, and they actually used them in the classroom. Others said that the full coverage of macro-skills in the GEPT had encouraged them to include listening and speaking in their teaching and assessment practices, which they had previously neglected in their teaching practices because the entrance examinations do not assess these two skills. Such GEPT washback was also reported by another impact study indicating that the GEPT has enhanced the importance of listening and speaking in senior high school English teaching and learning to some extent (Wu & Chin, 2006).

Engaging Teachers in Possible GEPT Revisions

The primary concern of any language test revision process is to ensure that the test accurately reflects real-life language use contexts and results in favorable learning outcomes; therefore, local teaching professionals' views on language use contexts should be taken into account when considering possible GEPT revisions. Examples of this are to be found in two recent GEPT revision projects. Based on the findings of a study in which the GEPT was related to the CEFR (Wu & Wu, in press), there was recognition of a need to revise the Elementary Listening and High-Intermediate Reading. More specifically, mini-talks were proposed as a new task type to be included in the Elementary Listening and longer passages were proposed for use in

High-Intermediate Reading. Through teacher forums, views on the proposed changes were exchanged and consensus was reached. Not only were teachers' views collected, the textbooks and learning materials used in the school were also analyzed. To confirm the appropriateness of the changes, the pretest results of the new test tasks were shown to teachers for comments. Owing to teachers' involvement in the revision process, the feasibility of the proposed revisions was able to be confirmed by teachers before they were implemented in 2010. The details of these two revision studies were reported in Chen and Chang (2008) and Ma and Li (2009).

Encouraging Research in Language Testing and Assessment

The introduction of the GEPT has not only influenced teaching and testing, but it has also resulted in an increasing interest in research into language testing and assessment among local EFL professionals in Taiwan. A review of the literature of the past seven years (2002-2009), as published in local major journals and proceedings of English teaching conferences and unpublished theses, reveals that a total of 35 papers and theses, excluding those written by the LTTC research staff, were written about the GEPT from various aspects. The topics of the papers can be broken down to three broad areas: critical review, validity, and impact. Some of these papers are introduced in the following:

The critical reviews of the GEPT (e.g., Pan & Pan, 2007; Shi, 2008) concurred that the GEPT could drive learners to study more English and could increase learners' English ability; however, they suggested a number of aspects for improvement, including acquiring international recognition for GEPT certificates, developing a more authentic speaking test, reducing the registration fee, administering the test more frequently, and providing test centers at more convenient locations. To respond to the request for fee reduction, the LTTC has reduced test fees by an average of 10% since 2006. Also, in order to meet the increasing demand, logistic support of the GEPT has been improved to increase the frequency of test administrations and to open more test centers with convenient locations

Similarly, Vongpumivitch (2010) reviewed research studies that have been conducted by both the LTTC research staff and non-LTTC researchers on the GEPT and evaluated the findings using Bachman and Palmer's (in press) assessment use argument framework. She concluded that the GEPT research studies had yielded strong support for the claim that the GEPT reports are consistent; however, a weaker support for test interpretations was found, given there is a wide range of test uses and different decisions being made based on GEPT scores. She therefore recommends that

more research studies be conducted, either by the LTTC or external researchers, 'to ensure the meaningfulness, impartiality, generalizability, relevance, and sufficiency of the interpretations about the abilities being assessed.' (p.169)

Liao (2006; 2009) investigated the construct validity of the GEPT listening and reading in light of the models of L2 reading and listening abilities and their relations to lexico-grammatical knowledge. In her studies, confirmatory factor analyses were performed to examine the factorial structure of the GEPT Intermediate listening and reading items. Results show that the grammar and vocabulary items in the reading section measured grammatical form and semantic meaning; that the reading comprehension items measured the understanding of literal meaning and pragmatic meaning; and that similarly, the listening comprehension tasks measured the ability to comprehend literal and pragmatic meaning. The structural equation modeling approach and discriminant analysis were then utilized to explore the relationships among lexico-grammatical knowledge, L2 reading ability, and L2 listening ability. It was found that lexico-grammatical knowledge was a strong predictor of reading and listening abilities, but lexico-grammatical knowledge was a stronger predictor of reading than of listening. Such research studies not only have provided support for the validity of the GEPT, but also have helped GEPT stakeholders to better understand the test construct. However, as noted by Vongpumivitch (2010), more qualitative evidence, for example, by means of verbal protocol analyses, should be collected to examine whether the GEPT tasks engage the ability defined in the construct.

Strong evidence was reported to support the criterion-related validity in various studies. For example, Gong (2002) compared the GEPT intermediate level with the Public English Test System (PETS) in China. Similarly, Sim (2006) compared the GEPT High-Intermediate level with the English language proficiency test developed by Tunghai University. In addition, Tung et al. (2005) found the match between the GEPT test content and junior and senior high school textbooks in terms of number of words used and topics of passages.

Numerous scholars (e.g. Shih, 2006; Vongpumivitch, 2006) have made significant efforts to collect evidence on the impact of the GEPT on instructional settings. The evidence supports claim that the test helps promote good instructional practice and effective learning. Taiwanese EFL classrooms generally emphasize grammar and vocabulary because the English test papers in the entrance examinations cover only reading and writing. As a result, due to the inclusion of oral language assessment in the GEPT system, listening and speaking now receive greater attention in the classroom. Nevertheless, it was observed that the impact on teaching methods and teachers' teaching philosophies was actually quite limited. Shih (2006) reported an observable washback effect of the GEPT at one university which requires students

to pass the GEPT intermediate level, while only little washback effect was observed on teachers' practices at a different university which did not require students to take the GEPT.

Vongpumivitch (2010) indicated that it is still unclear whether the consequences of using the GEPT and of the decisions that are made are beneficial to all stakeholders, because GEPT scores are used for various purposes. She therefore recommends that input from stakeholders be collected to support the warrant that all stakeholders benefit from the decisions they made based on GEPT scores.

While the GEPT was not always viewed positively in terms of its quality and delivery in these studies, they are valuable, constructive comments and will contribute to the improvement of the GEPT. In short, these studies not only benefit the GEPT, but they also certainly enrich the field of language testing and assessment in Taiwan. Hopefully, the interest in and enthusiasm for conducting studies related to language testing and assessment will continue to grow.

CONTINUITY AND INNOVATION

Like any large-scale language test, the GEPT faces numerous challenges in the future. For the continuity and innovation of the GEPT, agendas for future research and development are recommended as follows:

Ongoing Improvement of Test Quality

Ongoing efforts should be made to further improve the quality of the GEPT and to defend its claims with sufficient evidence and convincing argumentation (Bachman, 2005). As suggested in Kunnan (2000; 2004; 2008) and Kunnan and Wu (2010), validation, reliability, absence of bias, access and accommodations, administration and security, and social consequences should be examined continuously. Empirical validation of the GEPT level descriptors and can-do statements are regarded by the LTTC as its most urgent task. Additional evidence in support of the claims of the GEPT can be made available to test score users and other stakeholders in the near future.

Utilization of new technologies is another aspect for the improvement of the GEPT. With technological advances in computing and the Internet, the LTTC has begun to explore the possibility of online automated scoring. Moreover, in the interests of promoting test validity and efficient administration, the LTTC is currently developing the computer-based GEPT Advanced Writing Test. In the future, test takers will be free to choose this mode as an alternative to the current

paper-and-pencil test.

Addressing the Need to Facilitate Better Teaching and Learning

Taiwan is an examination-oriented society where examinations have long been used as tools to facilitate better teaching and learning. Moreover, because the development of the GEPT was supported by the government, with the aim of introducing beneficial washback effects to EFL education in Taiwan, the GEPT is thus seen to occupy an important position in Taiwan's language education. However, the relationship between teaching and testing is not simple, given that the needs of learners, the educational system, and society at large are changing (Bachman, 1990). Therefore, as an important language test in Taiwan, the GEPT has to respond as appropriately as possible to meet the needs of learners and society. To achieve this task, coordinated efforts were made to enable testing and teaching to function collaboratively and complementarily. The following are two examples.

i) Responding to Educational Needs

One example of a project geared to the integration of testing and teaching/learning is the creation of Taiwan's EFL Learner Corpora (TELC). The TELC is a joint research project undertaken by the LTTC and National Taiwan University's Graduate School of Linguistics. The TELC contains GEPT test-takers' written and oral performances, aiming to describe the features of test-takers' language production in terms of the lexico-grammatical components. The information will be very valuable to those who are involved in English language teaching in Taiwan, including teachers, curriculum developers, and writers of learning materials. Two studies utilizing the TELC data have so far been conducted: one is on topic familiarity in writing performance (Chung & Wu, 2009) and the other is on formulaic language and pauses in speaking performance (Skoufaki, 2009), and the findings have been presented at academic conferences. Moreover, the TELC is being used in the development of an error identification system and an automatic essay scoring system for the GEPT. It is hoped that, through collaborative efforts like the TELC project, language teaching/learning and testing in Taiwan will work together in a more complementary way, so that the ultimate goal of upgrading the language proficiency of Taiwan's EFL learners can be achieved.

ii) Responding to Social and Economic Needs

While GEPT scores have been used to gain better access to employment, test-takers' data (LTTC, 2009) indicate that the use of the GEPT for employment purposes is

obviously less common than for academic purposes. One speculation about the GEPT's lower popularity in the workplace is that the GEPT, in terms of content and format, may not meet the needs of professionals in Taiwan. To better address social and economic needs, the LTTC is exploring the possibility of developing an ESP module on the basis of the GEPT level framework. The GEPT-ESP project started with the banking and finance sectors, which have a great need to acquire competency in English to adapt to the continuing globalization of markets. The project is currently investigating the use of English as well as the need for English tests in the context of banking and finance. To achieve the task, Wu and Chin (2010) conducted a needs analysis study in which the informants included executive-, managerial-, and staff-level employees in Taiwan's banking and financial sectors. The data were collected through semi-structured interviews and a questionnaire survey. Preliminary findings show that the informants considered reading and listening skills to be more important to their work. Moreover, managerial employers in the financial industry suggested that the future GEPT-ESP test for finance professionals should focus on reading and listening skills and the tasks in the test should simulate the work tasks they perform in the work contexts. Such information is valuable input for developing a user-oriented assessment system as it can better inform the GEPT-ESP project of the needs of its users in workplaces.

CONCLUSION

In this article, I have presented a review of the GEPT from various perspectives, including test development, test validation, test use, and impact. I have also discussed the GEPT's directions for ongoing improvement and innovation to meet the emerging needs of EFL education and society in Taiwan. Like any large-scale language test, the GEPT faces some uncertainties in the 21st century. However, with a decade's efforts as the foundation and a strong commitment to continuing research and development efforts, it is hoped that the GEPT can meet the present and future challenges and exert a positive influence on English language education in Taiwan.

REFERENCES

American Psychological Association, American Educational Research Association & National Council for Measurement in Education.(1999). *Standards for Educational and Psychological Testing.* Washington, DC: Author.

Bachman, L.F. (1990). *Fundamental Considerations in Language Testing.* Oxford: Oxford University Press.

Bachman, L.F (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1-34.

Bachman, L.F. & Palmer, A.S. (in press). *Language Assessment in the Real World.* Oxford University Press.

Chen, C., & Chang, M. (2008). Assessing listening at elementary level: A GEPT case study. Paper presented at the 17th International Symposium on English Teaching, Taipei.

Chung, S.F. & Wu, C.Y. (2009). Effects of topic familiarity on writing performance: A study based on GEPT Intermediate test materials. Paper presented at the 2009 LTTC International Conference on English Language Teaching and Testing. Taipei, Taiwan.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Council of Europe (2003) Relating language examinations to the common European framework of reference for languages: Learning, teaching, assessment. Manual. Preliminary pilot version, Strasbourg, France.

Genesee, F. & Upshur, G. (1996). *Classroom-based Evaluation in Second Language Education.* Cambridge: Cambridge University Press.

Gong, B. (2002). Comparative studies of GEPT and PETS in Taiwan and China. *Selected Papers from the Special International Symposium on English Teaching/ Fourth Pan Asian Conference.*

Kunnan, A. J. (Ed.). (2000). *Fairness and validation in language assessment.* Cambridge, UK: Cambridge University Press.

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27-48). Cambridge, UK: Cambridge University Press.

Kunnan, A. J. (2008). Towards a model of test evaluation: Using the test fairness and wider context frameworks. In L. Taylor & C. Weir (Eds.), *Multilingualism and Assessment: Achieving transparency, assuring quality, sustaining diversity.* Papers from the ALTE Conference, Berlin, Germany (pp. 229-251). Cambridge University Press.

Kunnan, A.J. & Wu, J. (2010). The General English Proficiency Test. In L. Cheng & A. Curtis (Eds.), *English Language Assessment and the Chinese Learner* (pp.77-92). New York: Routledge.

Kuo, G. (2005). A preliminary corpus study on EFL test takers' writing proficiency. Proceedings of the Eighth Academic Forum on English Language Testing in Asia, Hong Kong, SAR, China, 27-35.

Language Training and Testing Center. (1999). GEPT Intermediate Level Research

Report. Taipei, Taiwan: Author.

Language Training and Testing Center. (2003). Concurrent validity studies of the GEPT Intermediate Level, GEPT High-Intermediate Level, CBT TOEFL, CET-6, and the English Test of the R.O.C. College Entrance Examination. Taipei, Taiwan: Author.

Language Training and Testing Center. (2005). Mapping the GEPT to the Common English Yardstick for English Education in Taiwan (CEY). Taipei, Taiwan: Author.

Language Training and Testing Center. (2009). GEPT Score Data Summary. Taipei, Taiwan: Author.

Liao, Y. (2006). Validating the GEPT: Lexico-grammatical ability as a predictor of L2 listening ability. Paper presented at annual Language Testing Research Colloquium, Melbourne, Australia.

Liao, Y. (2009). A construct validation study of the GEPT reading and listening sections: Re-examining the models of L2 reading and listening abilities and their relations to lexico-grammatical knowledge. Unpublished doctoral dissertation, Teachers College, Columbia University.

Ma, M., & Li, S. F. (2009). Bridging test construct and beneficial washback effects: Revising the GEPT High-Intermediate reading test. Paper presented at the 26th International Conference of English Teaching and Learning in National Tsing Hua University, Hsinchu, Taiwan.

Pan, Y.C. (2007). Consequences and evaluation of test use: The washback of exit requirements on tertiary English education in Taiwan. Manuscript in preparation.

Pan, Y.C. & Pan, Y.C. (2007). A review of the General English Proficiency Test (GEPT) in Taiwan. Working Papers of National Pingtung Institute of Commerce, 9, 55-78.

Roever, C. & Pan, Y.C. (2008). Test Review: GEPT; General English Proficiency Test. *Language Testing*, 25, 3, 403-418.

Shih, C.M. (2006). Perceptions of the General English Proficiency Test and its washback: A case study at two technological institutes. Unpublished doctoral dissertation, University of Toronto, Canada.

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

Sims, J. (2006). A model for creating a reliable and valid university proficiency exam. *Proceedings of the Twenty-third Conference on English Teaching and Learning in the Republic of China* (pp. 978-988). Taipei, Taiwan: Kuan Tung International Publications.

Skoufaki, S. (2009). Formulaic language and pauses in the speech of Taiwanese

learners of English. Proceedings of the 2009 International Conference on Applied Linguistics & Language Teaching, National Taiwan University of Science and Technology. Crane Publishing Company.

Teng, S.C., Kuo, C.Y., Wang, C.H. & Chang, M.C. (2006). Teachers' perceptions of the GEPT and its impact. *Proceedings of 2006 International Conference and Workshop on TEFL & Applied Linguistics,* (pp. 339-346). Department of Applied English, Ming Chuan University.

Vongpumivitch, V. (2006). An impact study of Taiwan's General Proficiency English Test (GEPT). Paper presented at annual Language Testing Research Colloquium, Melbourne, Australia.

Vongpumivitch, V. (2010).The General English Proficiency Test. In L. Cheng & A. Curtis (Eds.), *English Language Assessment and the Chinese Learner.* New York: Routledge.

Weir, C.J., & Wu, J. (2006). Establishing test form and individual task comparability: a case study of semi-direct speaking test. *Language Testing*, 23(2), 167-197.

Wu, J. (2002a). Assessing English proficiency at advanced level: The case of the GEPT. *Proceedings of the International Conference on Language Testing and Language Teaching* (pp. 93-100). Shanghai, China.

Wu, J. (2002b). Investigation of test-takers' views on difficulty at task level: A case study of GEPT-Intermediate spoken performance. *English Teaching & Learning,* 26(4), 107-119.

Wu, J. (2003). Task difficulty in semi-direct speaking tests – code complexity. *English Teaching & Learning, 27*(4), 79-98.

Wu, J. (2005). Task difficulty in semi-direct speaking tests. Unpublished doctoral dissertation, University of Roehampton, UK.

Wu, J. (2008). Views of Taiwanese students and teachers on English language testing. *Research Notes*, 34, 6-9.

Wu, J. (2010). Validating GEPT scores against teacher and student assessments. Paper presented at 2010 International Conference on Applied Linguistics & Language Teaching, Taipei, Taiwan.

Wu, J & Wu, R.Y.F. (in press). Relating the GEPT Reading Comprehension Tests to the CEFR In Martyniuk, W. (Ed.), *Aligning Tests with the CEFR: Reflections on Using the Council of Europe's Draft Manual.* Cambridge University Press.

Wu, R.Y.F. (2003). Assessing advanced-level English writing: A report on the case of the GEPT. *Proceedings of the Sixth International Conference on English Language Testing in Asia* (pp.75-101), Seoul.

Wu, R.Y.F., & Chin, J. (2006). An impact study of the Intermediate Level GEPT. *Proceedings of the Ninth International Conference on English Testing in Asia,* (pp. 41-65), *Taipei.*

Wu, R.Y.F., & Chin, J. (2010). An Investigation into the English Language Needs of Banking and Finance Professionals in Taiwan. *Proceedings of the 12th Academic Forum on English Language Testing in Asia* (pp. 73-87)*, Taipei*.

Wu, R.Y.F., & Liao, C.H.Y. (2010). Establishing a common score scale for the GEPT Elementary, Intermediate and High-Intermediate Listening and Reading Tests. In Kao, T.E & Lin, Y.F. (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: The Language Training and Testing Center.

Appendix: A description of GEPT levels and test tasks

Level Description	Skills	Task Types
Test-takers who pass the **Elementary Level** have basic ability in English and can understand and use rudimentary language needed in daily life.	Listening	Picture description; answering questions; short conversations and stories
	Reading	Sentence completion; cloze; simple short passage reading comprehension
	Writing	Paragraph writing (50 words)
	Speaking	Simulated format including reading aloud; answering questions
Test-takers who pass the **Intermediate Level** can use basic English to communicate about topics in daily life.	Listening	Picture description; answering questions; conversations
	Reading	Sentence completion; cloze; comprehension of narrative, descriptive, and expository texts
	Writing	Guided writing (120 words)
	Speaking	Simulated format including reading aloud; answering questions; picture description
Test-takers who pass the **High-Intermediate Level** have a generally effective command of English and can handle a broader range of topics.	Listening	Answering questions; conversations; talks
	Reading	Sentence completion; cloze; comprehension of different types of texts
	Writing	Guided writing (150 words)
	Speaking	Simulated format including answering

		questions; picture description; discussion
Test-takers who pass the **Advanced Level** are able to communicate fluently in English with only occasional errors related to language accuracy and appropriateness, and to handle academic or professional requirements and situations.	Listening	Long conversations and talks
	Reading	Careful reading; skimming and scanning
	Writing	Two 250-word essays based on two articles and two charts
	Speaking	Face-to-face interview (warm-up questions; discussion; presentation)
Test-takers who pass the **Superior Level** have English language abilities almost equivalent to the linguistic competence of a native speaker who has received higher education. They can use English effectively and precisely under all kinds of circumstances.	Integrated writing	A 750-word essay based on a 10-15 minutes' video/radio program and a 3,000-word article
	Integrated speaking	Face-to-face presentation; follow-up questions and answers