Critical Language Testing:
A Case Study of the General English Proficiency Test

Chih-Min Shih
National Institute of Education,
Nanyang Technological University, Singapore
dominiqueshih@hotmail.com

Abstract
Shohamy (2001b) proposed her theory of critical language testing, advocating that test takers should have the right to question the test itself, its value and its methods. She also believes that test takers’ views of the test are sometimes unique and test institutions can learn from test takers. To echo Shohamy’s theory of critical language testing, this study investigated participants’ viewpoints of GEPT-related issues. Topics covered in this study include test takers’ opinions about using the GEPT as one of their degree requirements, and the aspects of the GEPT which they believed can be improved. When it came to adopting the GEPT as a degree requirement, all students’ families and approximately sixty percent of the students I interviewed endorsed this policy. In addition, students suggested that the LTTC administer listening comprehension questions in a different manner, develop a more authentic speaking test, eliminate the interference caused by the babble of surrounding examinees in the speaking test, roll back the registration fees, do a facelift of the answer sheets, provide test centers at well-known locations, endeavor to acquire international recognition for its certificates, and administer the GEPT more frequently.

Key words: critical language testing, General English Proficiency Test, degree requirement
Literature Review

The past decade has witnessed increasing interest in the educational, social, and political dimensions of language testing. Other than the psychometric aspects of language testing, scholars have started paying attention to how tests are employed for educational, social and political purposes and what impact they have induced. Such insight has evolved to form the concept of critical testing.

According to Shohamy (2001b), “viewing tests as powerful tools – embedded in social and political contexts and agendas, related to intentions, effects and consequences and open to interpretations and values – places the field of testing within the broad area of critical pedagogy and is referred to here as critical testing” (p. 131). Shohamy (2001b) said that critical testing promotes critical strategies to investigate the uses and consequences of tests, to monitor the power of the test, and to empower test-takers. It also encourages reducing the deleterious effects testing induces and revealing the misuses of the testing. This notion of critical testing is referred to as critical language testing in the area of language education (Shohamy, 1998). The following are the domains of critical language testing that have been universally discussed in the past decade.

One aspect of critical language testing focuses on the abuse of language testing for political purposes. Shohamy (1998) pointed out that politicians had capitalized on language tests for tackling thorny political issues that they failed to address by other policy-making process. They could set the benchmark for passing a language test for
immigration purposes without any justification, thereby allowing them the flexibility to create immigration quotas. For example, the government of Australia drew on language tests to manipulate the number of immigrants and to determine if refugees could be accepted or rejected. Similarly, Latvia used strict language tests to prevent Russians from obtaining citizenship in the wake of its independence, an initiative that Shohamy (1998) said could result in one type of ethnic cleansing.

Another strain of critical language testing examines the uses of tests for various educational purposes (e.g., for admissions and graduation) as well as their impact on teaching and learning in different contexts (e.g., Qi, 2005; Shih, 2007; Shohamy, Donitsa-Schmidt & Ferman, 1996). For example, Shohamy et al. (1996) critically investigated if the Arabic Second Language (ASL) Test in Israel had induced the impact which the Inspector of Arabic intended. Although results showed that the ASL test had not brought about any favorable impact whatsoever, the Inspector insisted that it continue to be administered. Shohamy (1998) surmised that the Inspector’s insistence was probably because the test rendered him an excuse not to implement effective pedagogical actions.

Still another branch of critical language testing calls for the establishment of ethical standards for language testers. Scholars have discussed the issue of ethics in the past decade or so; the two major journals on language testing, Language Testing and Language Assessment Quarterly, have published its own special issue on ethics. This indicates that ethics has become one of the core considerations in language testing. In addition, the International Language Testing
Association (ILTA) in 2000 has published its Codes of Ethics which enumerates nine principles to professionalize the field of language testing and to serve as a guideline for good testing conduct.

The other domain of critical language testing promotes the rights of test takers. According to Shohamy (2001b), test-takers, in the past and mostly at present, had no rights and were the victims of tests. It was always the testing organizations (e.g., a school, an education system, the government) which determined the contents, methods, scoring, and uses of tests, thereby contributing to their permanent control over knowledge and power. Unfortunately, test takers could only comply without being able to complain or even protest that the test did not fit their knowledge or agenda. In fact, test institutions did not value the voices and concerns of test takers on tests. Despite the apparent imbalanced power relationships between test takers and test institutions, the former seemed to embrace the undemocratic situations. Shohamy (2001b) encouraged test takers to examine tests critically. She asserted that test takers should have been entitled to question the test itself, its value and its methods. She believed that testers could benefit from the valuable insights of test takers (see Shohamy, 2001a, 2001b). She also quoted Punch (1994) as saying that three dimensions of unethical or immoral testing practices were germane to test-takers: consent, deception, and privacy and confidentiality. Consent means that test takers should be allowed to determine if they will be tested or not. Deception denotes the scenario in which test-takers are engaged in a testing situation, but cannot relate it to the abilities which are supposed to be assessed. Privacy and confidentiality refers to if test-takers are denied access to
socio-economic resources due to the results of the test.

Researchers in Taiwan have paid scant attention to this new theoretical territory, notwithstanding Shohamy’s advocacy of her theory of critical language testing. One of the gaps which emerges from the previous literature is that a paucity of research has been conducted to examine test takers’ views of the test. With this research gap in mind, I used Shohamy’s (2001b) theory of critical language testing as the theoretical framework to investigate test takers’ views of the General English Proficiency Test (GEPT) and its uses for high-stakes purposes. This topic necessitates close scrutiny because a resounding number of students have taken the GEPT and numerous universities have adopted the GEPT for gate-keeping purposes. To guide the present study, I proposed the following research questions:
1. What are test takers’ views of adopting the GEPT as one of their degree requirements? What are their parents’ views of this policy?
2. In test takers’ views, is there any aspect of the GEPT which can be improved?

By investigating the two research questions, the study (1) offers test takers a chance to view the test critically, (2) empowers test takers, so they can voice their concerns and opinions about the GEPT and its use for high-stakes purposes, (3) transmits test takers’ comments on the GEPT to test designers, so the latter can gain insights about it and improve its quality accordingly, and (4) informs school authorities of students’ opinions about adopting the GEPT for high-stakes purposes. Before I elaborate on the research methods, it will be conducive to further contextualize the research by portraying the GEPT.
The GEPT, including the elementary, intermediate, high-intermediate, advanced, and superior levels, is an English proficiency test developed by the Language Training and Testing Center (LTTC) in Taiwan (LTTC, n.d.-a). The superior level assesses listening, speaking, reading, and writing abilities in an integrated test, and the other four levels are comprised of two stages. Reading and listening skills are assessed in the first stage, and listening and speaking components in the second stage. Test-takers are allowed latitude in selecting the level which they deem appropriate for their proficiency, but cannot proceed to its second stage until they have passed its first stage. A certificate of achievement will be awarded if test-takers have passed both stages of a specific level.

Of the five levels of the GEPT, the elementary, intermediate, and high-intermediate levels are the most frequently taken ones, so they are administered twice a year. According to the LTTC (n.d.-b), the proficiency levels of test-takers who pass the elementary level are roughly equivalent to those of junior high school graduates. Those who succeed in the intermediate level have attained the proficiency comparable to senior high school graduates’ (LTTC, n.d.-c). A certificate of achievement at the high-intermediate level guarantees the proficiency similar to that of non-English major university graduates (LTTC, n.d.-d).
METHODS

Interviewing

I adopted interviewing because it can draw out deep information about participants’ opinions, experiences, beliefs, feelings, and thoughts about a topic (Anderson, 1998; Ary, Jacobs & Razavieh, 2002; Best & Kahn, 2003; Holstein & Gubrium, 1995; Rubin & Rubin, 1995). In addition, although interviewing is more time-consuming than survey questionnaires (Byrne, 2004), the former outperforms the latter in several ways. First, interviewing is more natural and spontaneous (Anderson, 1998; Gochros, 2005). Gochros (2005) claimed that participants are more easily involved and can respond to research questions in oral rather than written form. Second, according to Gochros (2005), interviewing can generate a higher response rate than questionnaires. He pointed out that participants, when filling out questionnaires, may leave questions unanswered owing to a variety of reasons. Conversely, interviewers can assure that most research questions are responded to. Third, interviewing permits interviewers to observe and gain insights from non-verbal cues (Anderson, 1998; Gochros, 2005). For example, a pause or a rising intonation may contain implicit meanings that interviewers can detect and ask interviewees to explicate at the scene. Fourth, interviewing allows interviewers to have better control over the environment under which the research questions are posed (Gochros, 2005). For example, researchers do not know whether questionnaires are filled out by intended participants. Fifth, interviewing is flexible in that it allows researchers to orally investigate areas that may be difficult to frame in
questionnaire form (Anderson, 1998; Gochros, 2005). Interviewers can rephrase questions to adapt to each interviewee when necessary or clarify participants’ questions and probe their obscure answers on the spot, so interviewing can generate more comprehensive answers than questionnaires.

Research Sites and Participants

I purposely focused on students at private technological institutes because no research, to the best of my knowledge, examined their views of the GEPT and its use for gate-keeping purposes. The research sites were applied foreign language departments of a university of technology (Department A) and an institute of technology (Department B)\(^1\). Among the two departments, only the latter required all students in its day division to pass the first stage of the GEPT’s intermediate level one year before they graduated. If not, they had to take the school-administered make-up examination. If they failed both examinations, they would automatically fail their English Listening and Speaking course which was a required course, and therefore could not graduate from their study. The GEPT requirement constituted the most salient difference between the two applied foreign departments. By recruiting students from departments where the status of the GEPT differed, I could compare if their views of using the GEPT for high-stakes purposes varied.

All of the participants from both departments had had the

---

\(^1\) This research report summarizes the results from a large research project. Another portion of this research project has been published in Canadian Modern Language Review.
experience of taking the GEPT before they participated in the study, so they were cognizant of its content and format and were therefore more likely to view it critically. To maximize the diversity of participants, I considered prospective participants’ division (i.e., day division, night division, weekend division), age, gender, English learning experience (e.g., the experience of studying abroad), the highest level of the GEPT they took, the year they were in the program, and professions if they had a job. Finally, I recruited 15 students in Department A and 14 in Department B. I also solicited help from three students’ parents or family members in each department to investigate if they endorsed the idea of using the GEPT as one of the degree requirements. Table 1 shows the different academic programs in both departments as well as the numbers of participants and total students in each program. Detailed descriptions of all student participants’ demographic information, backgrounds, and GEPT test-taking experiences appear in Appendix A. The information about students’ families who participated in the study is shown in Appendix B.
Table 1
Numbers of Participants and Students in Each Department

<table>
<thead>
<tr>
<th>Department</th>
<th>Day Division</th>
<th>Night Division</th>
<th>Weekend Division</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5YJCP&lt;sup&gt;c&lt;/sup&gt;</td>
<td>4YCP&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2YJCP&lt;sup&gt;e&lt;/sup&gt;</td>
<td>2YCP&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Department A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. Students</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>V. Families</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Classes</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>ANTS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>270</td>
<td>90</td>
<td>270</td>
<td>90</td>
</tr>
<tr>
<td>Department B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V. Students</td>
<td>11</td>
<td>2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>V. Families</td>
<td>2</td>
<td>0</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Classes</td>
<td>8</td>
<td>2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>ANTS</td>
<td>400</td>
<td>90</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Note. Numbers represent the amount of participants, classes, or total students in the program. X means there are no corresponding programs and students in that department. Students in Department A were not required to pass the GEPT, whereas their day-division counterparts in Department B had to pass the first stage of the GEPT’s intermediate level.

<sup>a</sup>V. represents volunteer. <sup>b</sup>ANTS means approximate number of total students. <sup>c</sup>5YJCP, <sup>d</sup>4YCP, <sup>e</sup>2YJCP, and <sup>f</sup>2YCP mean 5-year junior college program, 4-year college program, 2-year junior college program, and 2-year college program, respectively.

Data Collection and Analysis

Although participants could choose to be interviewed in Mandarin Chinese, Taiwanese, or English, they all selected Mandarin Chinese. Interviews with students and their family members lasted
about 20 to 60 minutes and 15-25 minutes, respectively. All interviews were recorded with a digital audio-recorder and were transcribed. Schmidt’s (2004) five-step procedure was utilized as the framework for analyzing the interview data as I described in Shih (2007). In the first step, I determined my analytic categories such as “students’ views of using the GEPT as a degree requirement” after reading all transcripts several times. In the second stage, I further discovered the variants under each analytic category. For example, the analytic category “students’ views of using the GEPT as a degree requirement” had three variants: positive stance, neutral stance, and negative stance. I assembled all of the analytic categories with their variants to form a guide of analysis and coding. In the third stage, each interview transcript was analyzed on the basis of the coding guide. In the fourth step, I calculated the number under each variant if necessary. For example, I tallied the number of students who agreed, disapproved, or took a neutral stance as to whether the GEPT should have been used for gate-keeping purposes. Finally, detailed case interpretations were conducted when necessary. After data analysis was completed, I selected the most representative transcript excerpts to display in this paper. They were translated into English semantically first, and Chinese and English excerpts are juxtaposed in the results section, so readers can interpret the interview excerpts on their own.
RESULTS

1. GEPT as a Degree Requirement, Yes or No?

I asked students and their family members if they endorsed the policy of adopting the GEPT as one of the degree requirements. Staggeringly, all of the 6 students’ family members I interviewed did. One of the reasons that underpinned their support was that they believed that the GEPT requirement would guarantee a student’s proficiency to be up to the mark. For example, when asked why he supported the idea of using the GEPT as one of the degree requirements, Eric’s father Tom said:

我們現在的教育已經很嚴重一個問題就是說大學生的程度。大學生的英語程度越來越低落了，所以說你沒有這個門檻的話去鼓勵他的话，那以後大學生出來搞不好他英文能力也差不多國小國中而已。

A serious problem of the current education system is the levels of undergraduate students. The English proficiency of university students is deteriorating. If there is not a benchmark to encourage them, English proficiency of university graduates will probably be similar to that of elementary school students or junior high school students.

One student’s mother had foreseen the potential problems if this policy would be implemented, albeit her support for adopting the GEPT as a degree requirement. For example, YMH, whose daughter Nacci was a night division student in Department B, supported the idea of requiring students to pass the GEPT. However, she was concerned that those students “moonlighting” in the daytime might not be able to squeeze time to prepare for the GEPT. She said:
But students in the night division don’t spend a lot of time on their homework if they work in the daytime. Nacci doesn’t work in the daytime, so she can do it if we require her to spend more time on it. But in general, her classmates need to work in the daytime. I feel that it will be a little bit difficult for them. But in principle I think there should be a benchmark to maintain the quality of the students.

In contrast, student participants were not as supportive as their family members concerning this issue. Eight out of fifteen participants from Department A and nine out of fourteen participants from Department B agreed with the idea of employing the GEPT as one of their degree requirements. Detailed information concerning the students’ opinions is shown in Table 2.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Number of Students’ Opinions of Having the GEPT requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Opinion</td>
<td>Department A</td>
</tr>
<tr>
<td></td>
<td>D</td>
</tr>
<tr>
<td>Agree</td>
<td>1</td>
</tr>
<tr>
<td>Disagree</td>
<td>3</td>
</tr>
<tr>
<td>Neutral</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Only students in the day-division of the Department B were required to pass the first stage of the GEPT’s intermediate level. The rest of the students in both departments were not.

D, N, and W represent day division students, night division students, and weekend division students, respectively.
Results showed that students’ proficiency was not a clear-cut indicator of students’ views on this issue. For example, one student who passed the first stage of the GEPT’s high-intermediate level did not consent to this policy, whereas another student who failed the first stage of the GEPT’s elementary level did. The reasons why students supported this policy were the same as those proposed by their families; they were convinced that the GEPT would encourage them to study and help them achieve higher English proficiency. The other reason was proposed by Elsa who was a student in the night division in Department A. She viewed the GEPT’s certificate of achievement as a must to compete against her day-division counterparts. She said:

Students in the night division don’t study more than those in the day division. If we don’t get one more certificate, we would be at a disadvantage in the future. Because there are more exams in computer related fields as well as other areas in the day division, day-division students have more opportunities than us. We night-division students have a job, and don’t have a lot of time to prepare like students in the day division. So […] it is better for us if the requirement is prescribed.

夜校的學生又不比日校讀書
讀得多。如果不拿一張文
憑的話，以後競爭的話比日
校少，因為日間部他有比較
多的考試，像不論是電腦方
面好了啦，或是別的考試，
一定是機會又比我們多。像
我們夜校大部分都有在工
作，準備時間又沒有像日校
生來的多，所以[…] 2 我覺
得應該要實施，這樣大家才
比較好。
Differing from these positive opinions, objections to the GEPT requirement were also voiced. For example, Nacci and Clarabelle who were Department B students took a staunch position on this issue. They believed that taking the GEPT was a personal affair and should be of their own volition. For example, Nacci disagreed with the GEPT requirement and explained her viewpoint:

就看自己要不要考，學校為什麼規定我們一定要考，有過就自己的事，沒過也是自己的事。

The individual decides if he/she wants to take the test. Why should the school require us to take it? It’s my own business if I pass it. It’s also my own business if I don’t pass it.

Another reason why students were opposed to this policy was because not all of them were interested in English or would pursue their studies or career in English-relevant areas. A Department B student Shalin had passed the intermediate level of the GEPT. She seemed to be perturbed by her classmate’s inability to pass the GEPT. With an outpouring of sympathy, she stated why one of her female classmates ended up being trapped in a program she was not interested in:
Interviewer: Do you agree or disagree with the policy?

Shalin: But I think this is unfair for people whose studies were not English. What can they do if they keep failing the test? Because one of my classmates in the applied English department doesn’t like to study English, she kept failing the elementary level. Then I am worried what she can do if she has to pass the intermediate level.

[*****]

Interviewer: If she’s not interested in English, why does she study in the applied foreign language department?

Shalin: Her mother, parents, pressure from parents! They think her future will be rosier if she studies applied foreign languages. But that’s not her interest. So I don’t think this is applicable to everybody.
Shalin identified a phenomenon that is common in Taiwan; parents determining their children’s path of study, and disregarding their offspring’s actual interests. In Shalin’s testimony, her classmate’s parents foresaw a promising future if their daughter was enrolled in an applied foreign language department. However, students who were not interested in English might be gifted in other areas of study. They could pursue other tracks of academic study their school offered. For example, a Department A student Super Killer reported that some of his classmates were not skilled in English, but were conversant with other academic areas. He used one of his classmates as an example:

Although he studies in the same department, he is innately worse in this area {English}. He wants to transfer to the field of business. If you ask him to develop in languages, he will have a headache. […] Although these people are not good at English, they are competent at business. I do not think that this requirement can be imposed on everybody.

In addition, Debbie who was a night-division student in Department A did not uphold the idea of requiring students to pass the GEPT as a degree requirement due to the potential resistance of her classmates to the inflicted pressure. She also believed that the GEPT should be confined only to the workplace, and should not have been introduced into the education system:
Interviewer: Would you agree if your school required you to pass the GEPT as one of your degree requirements?

Debbie: There will be some difficulties for classmates if the GEPT is used as a benchmark for graduation. [...] when I sometimes proposed some ideas, everybody felt tons of pressure, let alone using the GEPT as a gate keeper. This will be even heavier pressure for them.

Interviewer: If you don’t consider your classmates, would you agree?

Debbie: I don’t think the GEPT should be introduced to schools. The GEPT is only needed in the workplace, or is for personal improvement. But the school’s teaching is systematic. If passing the GEPT is not a must, our learning will probably be more enjoyable.

Another reason why students were opposed to the GEPT requirement was the considerable proficiency gap among students, especially in Department A’s weekend division. Some students believed that this GEPT policy would not fully demonstrate students’ learning progress, especially for those with lower English proficiency. For example, Nancy who was a student in the weekend division in Department A said:
Students’ proficiency in our department differs widely. Some students can’t make significant progress after studying for two years not because they don’t study hard, right? This may be painful for students with extremely low English proficiency.

To sum up, all of the 6 students’ family members and a marginal majority of students seemed to endorse the policy or proposal of employing the GEPT as a degree requirement. Without a second thought, they believed that the GEPT requirement would prompt students to study and therefore enhance their English abilities. In addition, Elsa maintained that students in the night division would have remained competitive with their counterparts in the day division if the GEPT requirement had been established. A dozen of the students that I interviewed, however, verbalized their opposition to this policy. Two students asserted that they should be given freedom of choice about whether they would take the test. In addition, some students were resistant to this policy due to its imposed pressure. Moreover, students also asserted that the educational authorities had to take students’ diverse interests and proficiency gaps into account.

2. Issues and Suggestions

I asked students what aspects of the GEPT needed amelioration, and was bombarded with testing issues and suggestions, which ranged from the test itself to other logistical problems. Here I will enumerate them from issues related to validity.

**Validity.** One validity issue which bewildered my participants
was the criteria for selecting certain contents or vocabulary words covered by a specific proficiency level. For example, Liromensky questioned who decided what vocabulary words were tested at a specific level. He said:

Vocabulary words at the elementary level are too basic, those at the intermediate level do not seem to belong to that level, and those at the high-intermediate level are too difficult. Probably those who design test items are influential, so they determine which word belongs to what level. Everybody has a different viewpoint.

Like Liromensky, another Department A student Little P was not convinced that certain sections of the GEPT reflected the level claimed by the LTTC. She said:

Interviewer: Do you think any aspect of the GEPT can be improved?

Little P: Don’t deceive the society. Actually that’s the senior high school level, not junior high school level.

Interviewer: What do you mean?

Little P: The elementary level is somewhat like the senior high school level.

Interviewer: What do you mean?
Little P: I think its listening section is at the junior high school level. But some vocabulary and grammar questions in the reading comprehension section belong to the senior high school level. And those questions that require us to re-write the sentences are all at the senior high school level.

Test administration reliability. Brown (2004) pointed out that the conditions in which the test is administered might contribute to unreliability of test results. The most oft-criticized problem raised by students was the broadcast quality of the listening comprehension questions. Students called for a thorough revamp of the broadcast. For example, a Department A student Liromensky suggested using public announcement systems in lieu of portable cassette players.

再來是環境很差，在聽那個...<listening comprehension>的...時候，它沒有達到說很清楚，因為我考過三次嘛，每一次考試的時候我就知道一定會有人在後面舉手說他聽不到，前面卻說老師太大聲了[…]它其實應該用集體廣播的。

Then the environment is abysmal. When we were tested <listening comprehension>, it was not clear. I have taken it three times. Every time when I was taking the exam, I knew for sure that somebody sitting in the back would raise his/her hand and say he/she couldn’t hear. But people sitting in the front would say to the teachers that it was too loud. […] Public address systems should have been used.

5 < >: code switching to English in the interview
The other issue that students referred to was the babel uttered by surrounding test takers that interfered with their performance in the speaking tests. For example, a Department A student Tracy complained:

我們那時候是剛好來這裡考,它們是這樣一排間隔一個
一個去考，有的後面講話好
大聲，我剛好坐在中間，都
被後面干擾，好吵喔！都聽
到他們在講話，有時候會把
你的注意力分散掉，到時候
它在問什麼你沒注意聽都
變成聽到他們在講話，很
吵！

We happened to take the test here {Tracy’s university}. Test-takers were arranged to sit in every other row to take the test. Voices from some people sitting in the back were really loud. I happened to sit in the middle, so I was bothered by those sitting in the back. That was very noisy! I heard them talking. I was sometimes distracted and couldn’t hear the questions I was asked, but I heard what others were saying. It was very noisy!

Authenticity. Another complaint concerning the speaking test was its authenticity which Bachman and Palmer (1996) defined as “the degree of correspondence of the characteristics of a given language test task to the features of a target language task” (p. 23). According to the LTTC (2002), a portion of the questions in the speaking test section required test takers to listen to each pre-recorded question twice. Then they would have fifteen or thirty seconds to orally answer each question to the extent that they could. Their answers were audio-recorded and later evaluated. Here is an example from the LTTC (2002):

Question no. 1. Did you exercise this morning? Why or why not?

A Department B student Shalin was not poised to answer questions of this item type and called for a face-to-face speaking test.
When asked whether she considered any aspect of the GEPT could be improved, she replied:

訪談者：你認為全民英檢在哪一方面可以再做改進？
Joice： [...] 它可以不用費用那麼高，因為它一級一級費用不一樣。

Interviewer: Do you think any aspect of the GEPT can be improved?
Joice: [...] Can the registration fee be lower because different levels will charge different registration fees?

訪談者：你覺得以初級來說費用在多少比較合理？

Interviewer: Take the elementary level for example, how much is more reasonable?

Access. Apart from discontent with the administration of the listening and speaking tests, another problem that had plagued some interviewed students from all divisions of both departments was the financial access of the GEPT. They wondered if the exorbitant registration fees could be rolled back. Some of them were also bewildered why a higher level of the GEPT came with a heftier bill. For example, Joice said:

訪談者：你認為全民英檢的話，如果它們改成是對話，就是<one by one>對話，這樣可能會比較好。
訪談者：你覺得以初級來說費用在多少比較合理？

Interviewer: I think it will be better if the GEPT changes to conversation, <one by one> conversation.

6 Kunnan (2004) proposed his Test Fairness Framework which included five main qualities of a test: validity, absence of bias, access, administration, and social consequences. One dimension of access is financial access.
Joice: Money might not be the LTTC’s primary concern. I forget the first stage of the elementary level costs 500 or 600. I think around 300! [...] Or the LTTC can just charge once. Don’t charge for the second time. [...] Otherwise, I think it’s better to have test centers in certain universities in Taichung city, or some other well-known locations. But I was assigned to go to Da-Li to take the test. Where is it? The map was enclosed, but the problem was that I easily got disoriented.

Other suggestions. In addition to this financial concern, a Department B student Christine implied that the computerized answer
sheet should have been spruced up to make it aesthetically pleasing. She complained:

我每次看到那個格子就會很想睡覺，就覺得很小，就讓人覺得很緊湊。
Whenever I saw those oval bubbles, I felt sleepy. I felt they were very small and tightly packed.

In addition, Carrie suggested that the LTTC strive to gain international recognition. Although none of the other students I interviewed mentioned this as a problem requiring improvement, participants all acknowledged that the GEPT’s certificate of achievement would be ineffective overseas, a fact that undermined participants’ perceptions of the GEPT. If the GEPT intends to gain ground, worldwide recognition is an imperative. Last but not the least, a Department A student Joice suggested that the LTTC administer the GEPT more frequently. She said:

我覺得它考的次數可不可以多一點，不要今年考完我還要等明年之類的。
I wonder whether the GEPT can be administered more frequently. So after I take the test this year, I don’t have to wait until the next year.

DISCUSSION

Using the GEPT or other English tests as one of the degree requirements has become more and more universal among universities in Taiwan. However, little research has investigated students’ views of this extra requirement before the policy came into
effect at universities. According to the theory of critical language testing, this testing policy is undemocratic because it does not take students’ opinions into account. In this study, I offer students and their parents an opportunity to voice their concerns and considerations regarding this policy, so university authorities and policy makers can know what responses students and their parents might have. Surprisingly, all students’ family members and 17 out of 29 student participants acceded to this requirement. They supported this policy because they believed that it could drive students to study English and that it could keep students’ English proficiency up to the mark. Conversely, other participants opposed this policy because they believed students’ manifold academic interests, various English proficiencies, and personal affairs (e.g., part-time jobs) should be taken into account. They also objected to this policy because they believed that taking the GEPT is a personal choice and that an external test should not be introduced to the formal education system.

One major concern about the GEPT which emerges from interviews with my participants is the logistical support of the administration of the GEPT, so the test administration reliability and geographic access of the test are seriously affected. In my interviews with students, complaints concerning the listening and speaking tests recurred. Participants suggested improving the quality of the listening comprehension component by distributing headphones to each test taker or by adopting a public announcement system. I realize that the feasibility of these suggested solutions depends on the availability of the equipment. Whether there is copious equipment to accommodate the skyrocketing number of test takers is an open question, especially
when all test centers are inundated with test takers. If there is a lack of sufficient equipment to offer test takers a comfortable and reliable test administration environment, should the GEPT be administered to so many test-takers concurrently, an impractical measure which results in the invalidity of the test?

The study suggests that examinees’ views of the test are valuable for test developers as Shohamy (2001b) claimed. With their firsthand experience, they reflected on the pitfalls of the tests in the stage of administration. Their testimonies are conducive to developing a better quality of the test because this aspect of information is always overlooked by researchers who primarily focus on validity, reliability, and other psychometric components of the test. It is therefore fair to conclude that the quality of a given test can be examined and improved by incorporating psychometric research on it and test takers’ views of it. In this study, students suggested that the LTTC administer listening comprehension questions in a different manner, develop a more authentic speaking test, eliminate the babble of surrounding examinees in the speaking test, roll back the registration fees, do a facelift of the answer sheets, provide test centers at well-known locations, endeavor to acquire international recognition for its certificates of achievement, and administer the GEPT more frequently. I am convinced that the quality of the GEPT will be improved if these issues are addressed.

**Limitation of the Study**

The advantage of qualitative interviewing, as I have elaborated, is to elicit deep information about a person’s opinions, experiences,
beliefs, feelings, and thoughts about a topic. Therefore, it is the most appropriate research method for the present study. However, there is one limitation in the study; the results cannot be generalized to all Taiwanese test takers due to the limited number of participants. All of my participants were students at technological institutes. Their views cannot represent opinions of the rest of the test takers such as high school students or those who are no longer students. I suggest investigating participants from a wide spectrum of background to learn the full gamut of the views of test takers on the GEPT.

ACKNOWLEDGEMENTS

I would like to thank three anonymous reviewers from the journal for their valuable comments on the draft of this paper.

REFERENCES


http://www.lttc.ntu.edu.tw/E_LTTC/gept_eng_i.htm


ABOUT THE AUTHOR

Chih-Min Shih received his Ph.D. in second language education from Ontario Institute for Studies in Education (OISE) of the University of Toronto in 2006. He is currently an assistant professor in English Language and Literature Academic Group of Nanyang Technological University in Singapore. His research interests include language testing and teacher education. His recent paper has been published or has been accepted for publication in *Canadian Modern Language Review, Language Assessment Quarterly*, and *English Teaching: Practice and Critique.*
## APPENDIX A

### Information on Student Participants

#### GEPT Test-Taking Experience

<table>
<thead>
<tr>
<th>Pseudonym</th>
<th>Age</th>
<th>Sex</th>
<th>Division</th>
<th>Program</th>
<th>1st Stage</th>
<th>2nd Stage</th>
<th>1st Stage</th>
<th>2nd Stage</th>
<th>1st Stage</th>
<th>2nd Stage</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allen</td>
<td>25</td>
<td>Male</td>
<td>Night</td>
<td>2YJCP</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Automobile technician</td>
</tr>
<tr>
<td>Carrie</td>
<td>33</td>
<td>Female</td>
<td>Night</td>
<td>2YJCP</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td>President of a company</td>
</tr>
<tr>
<td>Clarabelle</td>
<td>18</td>
<td>Female</td>
<td>Day</td>
<td>5YCP</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Debbie</td>
<td>27</td>
<td>Female</td>
<td>Night</td>
<td>2YCP</td>
<td>O</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kindergarten English teacher</td>
</tr>
<tr>
<td>Elba</td>
<td>20</td>
<td>Female</td>
<td>Night</td>
<td>2YCP</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Accountant</td>
</tr>
<tr>
<td>Joice</td>
<td>27</td>
<td>Female</td>
<td>Weekend</td>
<td>2YCP</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kindergarten English teacher</td>
</tr>
<tr>
<td>Linda</td>
<td>46</td>
<td>Female</td>
<td>Weekend</td>
<td>2YCP</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Housewife</td>
</tr>
<tr>
<td>Lironwency</td>
<td>20</td>
<td>Male</td>
<td>Day</td>
<td>5YCP</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td></td>
<td></td>
<td>Three years in the US</td>
</tr>
<tr>
<td>Little P</td>
<td>18</td>
<td>Female</td>
<td>Day</td>
<td>5YCP</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Kindergarten English teacher</td>
</tr>
<tr>
<td>Nancy</td>
<td>31</td>
<td>Female</td>
<td>Day</td>
<td>2YCP</td>
<td>O</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randy</td>
<td>19</td>
<td>Male</td>
<td>Day</td>
<td>5YCP</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Super Killer</td>
<td>18</td>
<td>Male</td>
<td>Day</td>
<td>5YCP</td>
<td>XXX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tim</td>
<td>48</td>
<td>Male</td>
<td>Weekend</td>
<td>2YCP</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>President of a company</td>
</tr>
<tr>
<td>Tracy</td>
<td>43</td>
<td>Female</td>
<td>Weekend</td>
<td>2YCP</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Employee of a company</td>
</tr>
<tr>
<td>Vincent</td>
<td>25</td>
<td>Male</td>
<td>Night</td>
<td>2YCP</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>Part-time worker at school A</td>
</tr>
</tbody>
</table>

#### Notes

- Student participants are arranged alphabetically. O = passing that level of the GEPT, X = failing that level of the GEPT. Each O or X indicates one-time test-taking and its results. Two or more symbols in the same cell show the consecutive test-taking at the same stage of the GEPT.
- 2YJCP, 2YCP, 5YJCP, and 4YCP denote 2-year junior college program, 2-year college program, 5-year junior college program, and 4-year college program, respectively.
- ESL indicates that the participant had studied short-term in an English language program in an English-speaking country.
APPENDIX B

Information of Students’ Families

Department A
1. Julia, 47, is Tim’s wife. She is a public servant.
2. Mimi, 50, is Super Killer’s mother. She is a temporary employee of a municipality.
3. Tony, 52, is Randy’s father. He is a chief executive officer of a company.

Department B
1. LCC, 52, is Summer’s mother. She is a housewife.
2. Tom, 48, is Eric’s father. He is a land agent.
3. YMH, 46, is Nacci’s mother. She is an elementary school teacher.
評判性語言評量：全民英檢的個案研究

摘要
Shohamy (2001b) 提出了評判性語言評量的理論，提倡考生應該擁有質疑考試本身、考試的價值、以及考試方法的權利。她也相信參與測驗者有時對考試看法獨特，而測驗機構能夠向考生學習。為了回應 Shohamy 的評判性語言評量的理論，本研究調查受訪者對於全民英檢相關議題的看法。本研究涵蓋的主題包括考生對於將全民英檢納入畢業門檻的看法，以及考生認為全民英檢有哪些方面可以改善。全部受訪的學生家人以及大約 60% 的受訪學生支持將全民英檢納入畢業門檻這個政策。除此之外，學生們並建議語測中心以不同的方式施行聽力測驗、發展出真實的口說測驗、消除口說測驗中周圍考生發出的吵雜聲造成的干擾、降低報名費、美化答案卡、於知名的地點設立考場、致力於其證書獲得國際間的承認，以及更頻繁的施行全民英檢的測驗。

關鍵詞：評判性語言測驗  全民英檢  學位要求