

A Comparative Study of Construct Validity of Graduation English Proficiency Tests between Universities in Taiwan and Mainland China

Dr. Byron Gong

Abstract

This paper reports findings from an analytical study of how construct validity is reflected in national mandatory English graduation tests that have been widely conducted at universities in Taiwan and Mainland China. The findings relate to key test perceptions for test designers and stakeholders to consider should the mandatory testing systems in use be further endorsed by the educational authorities in the greater China area in the new wave of standardized university graduation English proficiency tests. This study analyzes the impact of backwash effects on normal tertiary ELT (English Language Test) programmes based on certain construct validation. Theoretically and empirically, the study employed documentary analysis, interviews, and investigation of test formats used in most universities in Taiwan and China. On the basis of the analysis, it is argued that positive backwash effects on national ELT programmes can be enhanced only when construct validation can be supported with a needed rationale. In this respect, the state college English test policy in Mainland China exerts a more desirable influence on national college English education while the college English test policy in Taiwan forms construct validity from a different perspective.

I Introduction

The purpose of this paper is to present a comparative study of how tertiary English education is assessed at universities in Mainland China and Taiwan respectively, in the narrow sense of mandatory testing criteria that are closely related to certain kinds of construct validation. To what extent that a large-scale test is designed to measure the intended test contents effectively and satisfactorily can be considered the primary concern for language test designers. In this sense, construct validity of such a test could become one of the most important considerations for test designers to consider. This could be especially critical when a test becomes high stakes because the outcomes of a high stakes test can be closely linked with test candidates' access to upward socio-economic mobility.

This paper provides studies of two different types of benchmark tests for college graduation. College students in Mainland China are under enormous pressure from a mandatory national testing system, i.e. the College English Test (**CET**); while their counterparts in Taiwan have to face the General English Proficiency Test (**GEPT**) or other tests designed by each individual university. To enhance the quality of their English education, universities in both Mainland and Taiwan are trying their best to carry out an unprecedented educational movement in terms of implementing a graduation English threshold test for all university students. In this background, construct validity could become one of the most important considerations for test designers to consider; and it is significant to assess the construct validation of such tests and the implications for test designers and users. However, as can be seen from the large scale and number of test candidates in Mainland China and Taiwan, it is impossible to discuss in depth the issue of how construct validity is reflected in the national tests such as the CET and GEPT in a paper of this scope. Therefore, this paper focuses only on some key issues of construct validation of the English language testing system that is used within the Chinese context. The writer also hopes that this paper will provide a new starting point for a possible exchange of experience in large-scale English language testing in the world. (Note that in this paper, *college* and *university* are used interchangeably, and both of these two words are used to refer to institutions of higher education within the Chinese context.)

II Theoretical Considerations

It is believed that construct validation involves an analysis of the qualities that a test is intended to measure, thus providing a basis for the rationale of a test (Bachman & Palmer, 1996). Generally, construct validity of a language test refers to an indication of how representative it is of an underlying theory of language learning. Construct validation involves an investigation of the qualities that a test measures, thus providing a basis for the rationale of a test (Davies, Brown, Elder, Hill, Lumley, & McNamara, 1999: 33). In other words, construct validity is concerned with the question: Is the study actually investigating what it is supposed to be investigating? (Nunan, 1992) In simple terms, “construct” means the idea used to support a test designer’s decision why s/he should design or construct a test in that certain way.

The writer holds that a good language test should be well supported by the rationale behind the test, and a good language test can bring about positive backwash effects on language teaching. The rationale for a high-stakes test should be effectively reflected in the test purposes, contents, and results with positive backwash effects in terms of social mobility. This study involves a supposition that positive backwash effects on national English test programmes can be enhanced only when construct validation can be supported with a needed rationale, both socially and economically. Construct validity of the above mentioned large-scale and high-stakes tests (the CET and GEPT) in the Chinese context can be viewed at two levels: national and grass-roots levels. At national level, construct validity can be viewed from its symbolic value for the state interest. In other words, the symbolic value of construct validation is related to the strategic needs of national college English education, and to how it can provide positive support for certain English test policies to be continued at macro level. Meanwhile, at grass-roots level, construct validity is more related to its functional value, which is the concern of how universities can have a reliable and valid evaluation system for the purpose of quality English education at college level. Therefore, the fundamental question raised in this paper is *to what extent construct validation helped bring about positive effects on college graduation English tests in Mainland and Taiwan. In other words, the question is about how college English graduation tests in Mainland China and Taiwan are related to their needed testing rationale.*

III Research Method

Both documentary and empirical data of the CET test in Mainland China and GEPT test (or another representative college test—the SCUEPT test) in Taiwan are used in this paper. As limited by space, only key issues are discussed. Since there is no single best way to study construct validity, the writer specifically looked into the test specifications, its criterion for a standardized test, internal reliability, (Pearson) inter-subtest correlation matrix, the content validity, etc.

As the CET is said to be a kind of criterion-related norm-referenced test (Jin, 2005; Yang and Weir, 1998), the writer first looked into the criterion that the CET is related to and how the criterion is represented in the test specifications. The writer believes that construct validation could be viewed by studying the test specifications of the CET (or GEPT) for a start. As Alderson pointed out (1995), it is generally regarded as the correct method by analyzing test specifications as a starting point in studying the construct validity of a test. Then, the writer analyzed some empirical data of the CET test, the GEPT, and SCUEPT. In particular, much of the first-hand empirical data of SCUEPT test results of college students in Taiwan was collected from a random selection of over 2000 non-English major 2nd-year students (from Soochow University in 2007 and 2008). The empirical results were also discussed. Finally, the writer discussed the effects of the benchmark exams in terms of both symbolic and functional values used in Mainland China and Taiwan. The writer holds that a probe of these aspects is intended for sensible answers to the research question of this paper. The discussion of relevant documentary and empirical data will be described from the next part.

IV Graduation English Tests at Universities in Mainland China

In Mainland China, the central government has been promoting an unprecedented English language testing system at universities with millions of test candidates each year. The national standardized CET test (College English Test), which was introduced by the National College English Testing Committee (NCETC) on behalf of China's Ministry of Education in 1987 and revised in 2005, has been such a high-stakes college English proficiency test that millions of undergraduate university students in China are required to take before their graduation (Han et al., 2004). The number of CET candidates is on the increase every year. In the 1995 academic year, 583,135 students in China took the CET, with a passing rate of 66% (Yang and Weir, 1998); and 9.58 million students took the test in the 2005 academic year (Jin, 2005). Considering the huge number of CET candidates, it is clear that China has not only the largest English test candidates, but also the largest English-learning population in the world. In reality, the CET has become such a high-stakes benchmark test that most universities would demand students to pass the CET so as to obtain their bachelor's degree. Although China's Ministry of Education altered its test policy in 2005 by stating that the CET is not to be directly related to college students' graduation, college students still consider the CET test crucial because the CET test certificate is an important criterion for many employers to consider at a job interview. In other words, practically, it is an irreversible trend for millions of Chinese college students to take the CET test before their graduation. Although there are negative voices against the CET (Han et al., 2004), it is generally believed that such a mandatory standardized English language testing system has brought about a cumulative positive effect on the quality teaching of college English education in Mainland China (Jin, 2005; Yang and Weir, 1998). Therefore, as being linked with both educational and social status, the CET test has become high-stakes for 20 years since its first launch in 1987. (Notice: English-majors have a more advanced national English proficiency test, i.e. TEM-8, the Test for English Majors, to take at the end of the 4th-year in their undergraduate programme.) Nevertheless, the results of in-house analysis and research on the CET will be further discussed in this paper, especially after studying the results of its counterpart testing system in Taiwan.

Construct Validation in the CET Test Specifications (Mainland China)

How much the test specifications of the CET (College English Test) can reflect the intended requirements of China's national teaching syllabus is considerably relevant to the degree of how construct validation can be fully represented in terms of its symbolic value. The CET is a national standardized test designed according to China's National College English Teaching Syllabus for Non-English Majors 1999 (which was revised and called "Requirements" in 2007). This national syllabus stipulates specific quantitative requirements for college students to achieve in terms of their English language proficiency, and skills of reading and listening are of paramount importance (<http://edu.people.com.cn/GB/8216/43375/5995154.html>).

The CET tests have two basic versions, CET-4, and CET-6. The CET-6 is for students who have passed the CET-4, and have taken elective English course of Band 5-6. The CET Spoken English Test (CET-SET) is administered only to a very small number of students who want to take by themselves on the condition that these students have passed the CET-4 with a score of 80 or above out of a full score of 100, or the CET-6 with a score of 75 or above. However, only the CET-4, which is the focus of this paper, is considered as the benchmark test that virtually all undergraduate students need to pass, and the CET test is administered twice a year, in January and June. According to China's National College English Testing Committee (2006), there are four main components in the CET-4 test: Listening Comprehension (35%—short dialogues and long authentic talks), Reading Comprehension (35%—careful reading and fast reading), Cloze (15%—one cloze and sentence translation), and Writing (15%—one short essay).

As for test specifications of the CET, the in-house report states that the guiding

principle is to reflect the requirements of the national syllabus. According to the research report by Yang and Weir (1998), the CET test specifications have generally met the requirements of the national syllabus. As the test is to help to implement this national teaching syllabus, the CET test designers had paid attention to the following aspects so as to construct a theoretical framework for the CET test:

- 1) The relationship between knowledge and ability: This means, conceptually, language is a tool for communication. The ultimate aim of EFT is to ensure that students can use English to communicate. Therefore, the CET should test more language skills rather than language knowledge.
- 2) The relationship between fluency and accuracy: The designers of the CET have set specific speed requirements of reading, listening and writing (i.e. 50 wpm and 129 wpm are set for reading and listening in the CET-4 test).
- 3) The relationship between sentence understanding and discourse comprehension: As communication is based on discourse comprehension, the CET should not only take into consideration of sentence structures, but also the ability to understand discourse.
- 4) The relationship between receptive ability and productive ability: This means the CET specifications require that both passive and active skills are to be examined.

According to the research report by Yang and Weir (1998), the CET test is designed according to the above four major considerations which constitute the basis for its construct validation at a macro level. The symbolic value of the construct validation of the CET test is therefore can be indicated by the degree of how the CET test can be accepted by both the educational authorities and university English teachers. According to an official survey by China's National College English Testing Committee (2006), the CET has successfully achieved the aims of its test specifications; and the construct validation based on the theoretical framework can be well represented in each delivery of the CET test. Specifically, the statistics also provide the following implications:

- The internal reliability of objective items in the CET test reaches 0.9 or above every time when the CET test is conducted, indicating that the reliability of the CET is high.
- A series of studies of questionnaires on the CET has indicated that 92% of college teachers in China agree that the CET test can effectively reflect students' actual English proficiency level, indicating a high validity in terms of expert judgment.
- As the CET is a criterion-related norm-referenced test, the passing score set in the CET correlates with the teachers' assessment of the test candidates' passing score with a correlation coefficient of 0.82. In addition, the CET test scores correlate with the order of class assessment results given by the teachers with a correlation coefficient of 0.7, which is very good because it is difficult to achieve such a high coefficient in large-scale standardized tests.
- Over 86% of college teachers agree that the contents of the CET are appropriately designed and each part has a proper weighting.
- The CET has a complete testing system, including item bank management, test formation and organization, administration, statistical analysis of test results, test fairness, and practicality.

Therefore, upon conclusion of this part, the writer maintains that the construct validity of the CET is mainly associated with the state interest at national level. In other words, the test specifications of the CET reflect governmental initiatives for centralization and standardization of language testing at a national level, with a centralized definition of ability construct. Furthermore, empirically, China's national educational authorities have gained solid statistical support for its CET policy to be continued nationwide. Next, the graduation benchmark testing system in Taiwan will be discussed.

V Graduation English Tests at Universities in Taiwan

In Taiwan, there is not a mandatory island-wide English proficiency test set by Taiwan's Ministry of Education for undergraduate students to take for graduation. Unlike its counterpart in the Mainland, the Chinese educational authority in Taiwan has been carrying out an American style of autonomous and decentralized language assessment within colleges and universities. The educational authority in Taiwan has transferred power to lower levels at each individual university, which gives more freedom to universities to decide what kind of English proficiency is needed for their undergraduate students according to each university's own principles; and many universities in Taiwan have recently announced that they will carry out their own benchmark English testing system for college students to take. In other words, a kind of threshold English test is about to be carried out in the near future across the university campuses in Taiwan. In addition, undergraduate students can also take other English proficiency tests as a proof of their English proficiency before they finish the 4-year university education, such as the GEPT (General English Proficiency Test, a criterion-referenced test set by a non-governmental organization in Taiwan), or TOEIC, IELTS, and TOEFL. Nevertheless, the local GEPT test is the most popular English proficiency test for college students to take although the passing rate for college students is around 32% (LTTC, 2007).

Meanwhile, among the English proficiency tests designed by individual universities at grass-roots level in Taiwan, different universities have their own testing systems and criteria. In contrast to the common practice of using achievement test of the last term of the university programme, the SCUEPT test (Soochow University English Proficiency Test) appears to be at the forefront of the campaign for a standardized English proficiency threshold test for undergraduate students to pass for graduation. Many other universities are also trying to design their own proficiency tests now. By and large, it is clear that a variety of benchmark English tests will soon become high-stakes tests for thousands of college students in Taiwan to take; as such a test certificate would help college graduates to have better opportunities in the job market, too.

Construct Validation in the GEPT/SCUEPT Test Specifications (Taiwan)

Now let us take a look at the benchmark English test for college students before graduation in Taiwan. There are more than 150 officially accredited universities in Taiwan. However, there lacks a cohesive paradigm of college English assessment at the tertiary level. There exists no requirement from the educational authorities in Taiwan to demand all college graduates to take an English proficiency test before they graduate. As mentioned earlier, the educational authority in Taiwan has transferred power to lower levels at each individual university, which gives more freedom to individual universities to decide what graduation threshold test should be. According to the writer's investigation, few universities actively demand their students to take GEPT (General English Proficiency Test) or any other public tests as part of the requirements for graduation. According to a 2007-year report on GEPT, only 22% of GEPT testees took the GEPT in order to give their test scores to their universities for reference. As for the specifications, the GEPT is not designed to test just college students' English proficiency, but to test the English proficiency of the general public, which is very different from that of the CET. In addition, there exists little research of the possibility of a large-scale mandatory testing system designed especially for college English education in Taiwan.

Notwithstanding this, the freedom from government control regarding using a standardized and centralized assessment within colleges and universities in Taiwan reflects autonomy in defining the construct of ability, whose rationale could have a different framework for different socioeconomic purposes. However, the negative side of such autonomy in defining construct ability can also cause various problems for local universities to solve. Practically, the general scenario of the evaluation of English

programmes among universities in Taiwan is that test scores may be inconsistent and incompatible. In other words, the macro-relationship between college students' English competence and the applied evaluation methods in Taiwan is not clear, which means different universities adopt their own methods in evaluating students' English proficiency, and such methods may be inconsistent each year and differ in different department, even differ from one individual teacher to another. Thus, it is far from the desirable situation that testing results can be considered mutually compatible among colleges in Taiwan in both theory and practice. This is because current evaluation criteria for college graduates' English proficiency level in Taiwan are not based on an island-wide or nationally agreed standard, such as that of the CET-4 used in Mainland China (Yang and Weir, 1998). Therefore, the evaluation results conducted by different colleges and universities are difficult to interpret in terms of statistical analysis at national level.

In addition, at present, graduation examinations across colleges and universities are mostly of progress tests or achievement tests, and the contents of such achievement tests could be widely different from one university to another on the basis that different teaching materials are used. That is to say, such assessment practice of English language programs can hardly provide reliable and valid evaluation of college graduates' English language proficiency (Gong 2004). For example, let us look at the achievement test scores of the same English course at two campuses of one university.

Table 1 Pearson Correlation of Freshmen English Scores between Taipei and Kaohsiung Campuses (N=900; 1/2005)

	Taipei Campus	Kaohsiung Campus
Taipei Campus	1.000	.034
Kaohsiung Campus	.034	1.000

Table 1 reflects the fact that test scores are not comparable even within the same university. No significant correlation can be found between the test scores at its two campuses (Kaohsiung and Taipei) of the same university, with a Pearson correlation of 0.034 ($p < 0.01$). This may suggest that as there are no established English test syllabus and test specifications for universities in Taiwan to follow, different tests are used for evaluation. As a result, there exist different kinds of English tests at each university. According to the writer's survey of the assessment of college students' English proficiency at different colleges and universities in Taiwan, there is not a unified standardized criterion for these educational institutions to follow. In fact, each university has to take its own approach to the assessment of its English programmes at different levels. Hence, in the light of different teaching materials, the tests used for graduation examinations, if required, are unsurprisingly related to different teaching materials. The test results are accordingly not comparable due to the fact that there are different test contents; and there are hardly any test specifications written for such wide-ranging tests.

Hence, as far as construct validity is concerned, it shows that the understanding and interpretation of language ability could vary at different universities. With an autonomous and decentralized language testing system, each individual college or university may decide their own testing criteria according to their own needed rationale. Therefore, at national level, the symbolic value of construct validation of both GEPT and other individual college tests in Taiwan appears to be limited when compared with that in the Mainland where the construct of the CET test is closely related to the national English teaching syllabus. But for universities at grass-roots level, their test results appear to be inconsistent and incompatible, which may considerably affect the construct validation of the test designed by each individual university. Thus, the English language testing is said satisfactory and successful only in terms of the interpretation of the needed test construct, or "ability" by each university itself.

VI Discussion of Empirical Data of the CET, GEPT, and SCUEPT

Test designers in both Mainland China and Taiwan paid much attention to the issue of validity in their tests at different levels. Now, let us look at some empirical data so as to have a better view of the CET-4 test and GEPT (or SCUEPT). First, reliability results of the above three tests are reported as follows: the reliability (Cronbach's alpha) of the CET is 0.9, and 0.86 for the SCUEPT, while the GEPT is said to be 0.85 (LTTC 2003:23). But the CET has kept its reliability as 0.9 for nearly 20 years, which is remarkably good. As for GEPT and SCUEPT, the reliability is not stable, and it changes over time.

The internal correlation coefficients of these tests could be also very useful for us to evaluate construct validation of these tests. The following Tables 2—5 show the internal correlation coefficients.

Table 2 Internal Correlation Coefficients of the CET-4 (Yang and Weir, 1998:60)

	LC	RC	VS	CL	WR	Total
LC	1.000					
RC	.5630**	1.000				
VS	.5390	.6147	1.000			
CL	.4671	.5310	.6264	1.000		
WR	.3879	.3594	.4698	.4042	1.000	
Total	.7922	.8915	.8022	.7076	.5813	1.000

** Correlation is significant at the 0.01 level (2-tailed).

(LC=listening comprehension; RC=reading comprehension; VS=vocabulary & structure; WR=Writing)

Table 3 Internal Correlation Coefficients of the GEPT (High-Intermediate Report, 2007)

	Listening	Reading	Writing	(Speaking)
Listening	1			
Reading	0.37**	1		
Writing	0.19	0.32	1	
Speaking	0.39	0.23	0.38	1

** Correlation is significant at the 0.01 level (2-tailed).

The above Table 2 shows that the internal correlation coefficients of each part in the CET-4 are between 0.3—0.7, which is fairly good. Compared with Table 2, Table 3 shows that the internal correlation coefficients in GEPT in Taiwan is less desirable, especially the coefficient between Listening and Writing in GEPT (2007) is on the small side in comparison with that of the CET shown in Table 2. But as we mentioned earlier that the statistics of the GEPT may not be constant, its earlier 2000 report provided a different analysis than the 2007 report, which can be seen in the following Table 4.

Table 4 Internal Correlation Coefficients of the GEPT (2000 LTTC Report)

Sub-test	Reading Part A	Reading Part B	Reading Part C
Reading Part A	1.000		
Reading Part B	0.681	1.000	
Reading Part C	0.686	0.722	1.000
Listening Part A	0.591	0.590	0.598
Listening Part B	0.629	0.624	0.648
Listening Part C	0.620	0.605	0.680

N=375 p≤0.01

The internal correlation coefficients of the GEPT (2000 test) appear to be more convergent than that of 2007 test. It is revealed in the 2000 GEPT Report that high

coefficients between Listening and Reading in GEPT may be caused by the fact that the test format and content are similar, i.e. both are of paragraphs compression (LTTC, 2000).

The following Table 5 is about the internal correlation coefficients of the SCUEPT test at Soochow University in Taiwan conducted in May 2008.

Table 5 Internal Correlation Coefficients of SCUEPT (Pearson / N=2004, 2008)

	Sentence	Fast reading	Careful reading	Cloze	Short conversation
Fast reading	0.256**				
Careful reading	0.320**	0.327**			
Cloze	0.305**	0.249**	0.362**		
Short conversation	0.384**	0.322**	0.436**	0.360**	
Long talks	0.312**	0.291**	0.426**	0.283**	0.611**

** Correlation is significant at the 0.01 level (2-tailed).

Looking at Tables 2 and 5, we can see the inter-subtest correlations of the GEPT and SCUEPT in Taiwan generally appear to be on the small side when compared with those of the CET. This might indicate that each inter-subtest in GEPT or SCUEPT is too disintegrative regarding language communication skills. Probably, further efforts are needed so as to achieve less divergent construct validity in the GEPT, which needs to keep a fair balance between convergent and divergent validity by using test items of better discrimination and difficulty index. Next, let us look at content validity.

VII Content Validity

Construct validation can be viewed from the perspective of content validity. One good way to study content validity is to gather the judgment of experts. Alderson et al points out: "Typically, content validation involves 'experts' making judgements in some systematic way. A common way is for them to analyze the content of a test and to compare it with a statement of what the content ought to be" (1995:173).

Tables 6 and 7 show the relevant results of the CET-4 in Mainland China and SCUEPT in Taiwan.

Table 6: Content Validity of the CET-4 Test Based on Teachers' Evaluation

(Yang and Weir, 1998:175)

General Comments	English Teachers	Students
1) useful/reflecting students' ability	68%	41.6%
2) Useful/good for jobs	24%	26.5%
3) Useless/not reflecting ability	4%	14.6%
4) Useless/students unwilling to take	4%	11.4%

Table 6 shows only 8% of teachers have negative views on the CET-4, but the majority believes the CET-4 is creditable.

On the other hand, no judgment of experts has been reported regarding the GEPT in Taiwan. As for the SCUEPT, the writer interviewed and collected questionnaires from 20 college English teachers to make their judgment of the content validity of the SCUEPT. The question to these 20 teachers is: *To what extent does the SCUEPT test not suffer from construct under-representation or construct irrelevant variance?* Out of a scale from 1 to 10, the bigger the number is, the less the test suffers from construct under-representation. The feedback is shown in Table 7.

Table 7 20 Teachers' Feedback of Construct Under-representation of SCUEPT (2007)

T	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
+	8	7	8	9	9	8	6	9	7	9	8	8	7	9	9	8	9	7	9	8
-																				

This “expert” feedback shows that 81% (162/200) hold that the test content does not suffer from construct under-representation. Therefore, statistically, the content validity for the SCUEPT can be rated as significantly strong.

VIII Discussion

So far we have briefly studied the construct validation of the College English Test (CET) in Mainland China and the General English Proficiency Test (GEPT), and SCUEPT (a representative English proficiency test designed by a local university) in Taiwan from different aspects. We can see there are huge differences in the understanding and interpretation of construct validation in theory and practice.

At macro level, for the Chinese educational authorities in the Mainland, the CET test serves the state interest and the needs of the national higher education. The CET reflects governmental initiatives for centralization and standardization of language testing at a national level, with a centralized definition of ability construct. The CET appears to be such a mandatory standardized English language testing system that all college students in the Mainland need to pass this high-stakes benchmark test for graduation. Therefore, the CET has virtually mobilized all Chinese college students to study English hard. Although there are criticisms of the CET (Han et al., 2004), the major positive backwash effects of the CET test is that all universities in China have realized the importance of college English education, and have taken various actions to promote college English actively, which has brought about a cumulative positive effect on the quality teaching of college English education nationwide (Jin, 2005). In this sense, the construct validation of this high-stakes national test is linked with the needs at national level. The symbolic value of construct validation of the CET in China is related to the needed rationale not only educationally but also socially and economically.

However, as for the GEPT test, the Chinese educational authorities in Taiwan have adopted a completely different approach to the assessment of college English education. Universities in Taiwan have much more freedom to decide what kind of English proficiency is needed for their undergraduate students according to each university's own understanding and interpretation of test construct abilities. The autonomous and decentralized language assessment within colleges and universities in Taiwan provides autonomy in defining the test construct of ability according to their own local needs. Therefore, the GEPT is not high-stakes, and the symbolic value of construct validation of the GEPT appears to be limited when compared with that in the Mainland where the construct of the CET test is closely related to the national English teaching syllabus.

In view of symbolic value of construct validity, the writer believes that the CET has comparatively got more credit because a set of dedicated test specifications has been designed, which is directly in line with the purpose of the CET test in the Mainland. By contrast, the GEPT's test specifications are not specially designed to evaluate college graduates' English competency, not to mention graduation threshold. Therefore, strictly speaking, the purpose of the GEPT and its backwash effects on college English education do not accord with the needs of college English programs in Taiwan. There is no comparison between CET and GEPT in this sense. On the credit side, SCUEPT is designed for graduation threshold, but the symbolic value of its construct validity is limited as it is just for one individual university in Taiwan.

At grass-roots level, the functional value of construct validation of both the CET and

GEPT (or SCUEPTS) can be viewed with their specific needed rationale. Generally, the CET, GEPT, or SCUEPT all have solid empirical data to support its own rationale regarding internal reliability, correlation coefficients, content validity, etc., which are all satisfactorily acceptable at micro level. However, in the light of different tests used for graduation examinations in Taiwan, if required, the test results are unsurprisingly not comparable due to the fact that there are different test contents; and there are hardly any test specifications written for such wide-ranging tests. So, the functional value of construct validity is limited to each individual university in Taiwan.

Conclusion

The findings of this comparative study have revealed that positive backwash effects on national ELT programmes can be enhanced effectively when construct validation can be linked with not only a needed rationale but also strong government support. As for the research question raised in this paper, i.e. *to what extent construct validation helped bring about positive effects on college graduation English tests in Mainland and Taiwan*, the writer would hold that the state English language testing policy in Mainland China, generally, exerts a more desirable influence on her national college English programmes. Its CET testing system has produced far-reaching positive effects on college ELT in Mainland China. As a final point, a standardized benchmark college English test might provide a long-term solution to the malaise of the English language tests at universities in Taiwan, which could be especially meaningful when construct validation is considered not only at grass-roots college level, but also at national level.

References

- Alderson, J. C., Clapham, C., and Wall, D.** (1995) *Language Test Construction and Evaluation*. Cambridge: CUP.
- Bachman, L.F. and Palmer, A.S.** (1996) *Language testing in practice*. Oxford: OUP.
- Cheng Liying** (2008) The key to success: English language testing in China. *Language Testing* 25(1): 15-37.
- Davis, A., Brown, A., Elder, C., Hill, K., Lumley, T., and McNamara, T.** (1999) *Studies in Language Testing 7: Dictionary of language testing*. Cambridge: CUP.
- Gong, Byron** (2004) "A Need for a Unified Assessment of College English Language Programs—Some Theoretical and Practical Considerations for Quality ELT in Taiwan" *Shih Chien Management Commentary*, Issue 1.
- Jin, Y.** (2005) *The National College English Test of China*. In Hamp-Lyons, L. (Chair), *the big tests: Intentions and evidence*. Symposium presented at International Association of Applied Linguistics (AILA) 2005 Conference in Madison, WI.
- LTTC** (2007, 2003, 2000) *A Statistical Report on the Scores of a GEPT® Test*. Taipei, LTTC.
- National College English Testing Committee, PRC** (2006) *College English Test Sample Papers*. Shanghai, China: Shanghai Foreign Language Education Press.
- Nunan, D.** (1992) *Research Methods in Language Learning*. Cambridge: CUP.
- Yang, H. and Jin, Y.** (2000) Score interpretation of CET. *Proceedings at the Third International Conference on English Language Testing in Asia*. Hong Kong: Hong Kong Examinations Authority, 32–40.
- Yang, H. and Weir, C.** (1998) *Validation study of the National College English Test*, third edition. Shanghai, China: Shanghai Foreign Language Education Press.