The Construction and Development of DPU-Communicative English Proficiency Test (DPU-CEPT)

การสร้างและพัฒนาแบบทดสอบสมิทธิภาพภาษาอังกฤษ เพื่อการสื่อสาร (DPU-CEPT)

*Eakkachai Trikomol[1]

เอกชัย ไตรโกมล

Arnond Sakworawich[2]

อานนท์ ศักดิ์วรวิชญ์

Pornchanok Sukphan[3]

พรชนก สุขพันธ์

Supoj Pongthong[3]

สุพจน์ ปองทอง

Warithorn Samana[3]

วาริธร สมณะ

[1]Lecturer and Testing Centre Manager, Language Institute, Dhurakij Pundit University, Bangkok, Thailand, Correspondence regarding this article should be addressed to prime investigator via eakkacha@dpu.ac.th

[2]Managing Director, Human Capital Consulting and Assessment Co., Ltd., Bangkok, Thailand

[3]Lecturer and the DPU–CEPT Test writer, Language Institute, Dhurakij Pundit University, Bangkok, Thailand

Abstract

The DPU–Communicative English Proficiency Test (DPU–CEPT) has been constructed and developed for teaching and learning English at a university level. The DPU–CEPT was constructed to be analogous to the standardized Test of English for International Communication (TOEIC) developed by the Educational Testing Service (ETS). The DPU–CEPT will be equated with TOEIC for an easier score interpretation. The DPU–CEPT shows an acceptable internal consistency reliability (KR–20=.78). The strong correlations between the DPU–CEPT and TOEIC (r=.79) show evidence of construct validity. Moreover, the DPU–CEPT also correlates with current GPA (r=.71) and GPAX (r=.57), and those scores provide the evidence of concurrent validity. With satisfactorily psychometric properties, the DPU–CEPT can be applied to learning and teaching English as a Second Language.

บทคัดย่อ

  แบบทดสอบ DPU – Communicative English Proficiency Test (DPU–CEPT) ถูกสร้างและพัฒนาขึ้นเพื่อการเรียนการสอนภาษาอังกฤษในระดับมหาวิทยาลัย แบบทดสอบ DPU–CEPT มีลักษณะคล้ายคลึงกับแบบทดสอบมาตรฐาน Test of English for International Communication (TOEIC) ซึ่งพัฒนาโดย Educational Testing Service (ETS) แบบทดสอบ DPU–CEPT ถูกสร้างให้เทียบเคียง (Equating) กับแบบทดสอบ TOEIC เพื่อการตีความคะแนนที่ง่ายขึ้น แบบทดสอบ DPU–CEPT มีค่าความเชื่อมั่น (KR–20 =.78) ในระดับที่ยอมรับได้ สหสัมพันธ์ระหว่าง DPU–CEPT และ TOEIC (r=.79) ยืนยันความตรงตามภาวะสันนิษฐาน (Construct validity) นอกจากนี้แบบทดสอบ DPU–CEPT ยังสัมพันธ์กับเกรดเฉลี่ยในภาคการศึกษาปัจจุบัน (GPA) (r=.71) และเกรดเฉลี่ยสะสม (GPAX) (r=.57) ซึ่งเป็นหลักฐานความตรงร่วมสมัย ด้วยคุณสมบัติเชิงจิตมิติที่น่าพึงพอใจ แบบทดสอบ DPU–CEPT จึงสามารถนำไปประยุกต์ใช้ในการเรียนการสอนภาษาอังกฤษเป็นภาษาที่สองได้

Keyword: Construction and Development, Proficiency Test, Communictive English

คำสำคัญ: การสร้างและพัฒนาแบบทดสอบ, แบบทดสอบสมิทธิภาพ, ภาษาอังกฤษเพื่อการสื่อสาร

The measurement of English language proficiency has contributed many advantages to teaching English as a Second Language. It can be used as a placement test in order to classify and place students into classes suitable for their language proficiency, as a diagnostic test to identify the weak language skills that need improvement, and as an achievement test for measuring learning and teaching effectiveness, for instance (Hughes, 1989).

There are many commercial standardized tests available for assessing communicative English proficiency; however, the Test of English for International Communication (TOEIC) developed by the Educational Testing Service (ETS) is one of the most widely acknowledged among business organizations worldwide, especially as a tool for selection, training, and placement (Educational Testing Service, 2002). The TOEIC test comprises listening and reading comprehension tasks in a business context, and is taken by people whose native language is not English. In Thailand, many leading corporations now require an official TOEIC score from applicants prior to their applying for a job in those companies (Centre for Professional Assessment [Thailand],

2006), for example, Siam Cement Group, PTT Group, Thai Airways International, etc. For the above reasons, most university graduates seem to be obligated to take the TOEIC test as soon as possible after their graduation to use it as part of their job application and career advancement.

The TOEIC test has been well researched and documented (Educational Testing Service, 1998, 2002, 2007). Abundant evidence (e.g., Wilson, 1993a; Wilson, 1993b, 1999, 2000; Woodford, 1982) shows that TOEIC is a valid, reliable, and practical test of English language proficiency, particularly in the Thai academic and employment context (Wilson, Komarakul Na Nagara & Woodhead, 2004).

However, teachers of English as a Second Language and university students seem to be reluctant either to take or use TOEIC for both teaching and learning because of its rather high testing fee in Thailand. Hence, the testing centre at Dhurakij Pundit University, as an authorized English language–testing agency, would like to construct and develop an English standardized test, titled the DPU–Communicative English Proficiency Test (DPU–CEPT) for assessing English language proficiency

in a business context similar to TOEIC's, to be utilized as part of English learning and teaching. The DPU–CEPT will be constructed and developed as a standardized test with satisfactorily psychometric properties and practicality, and the DPU–CEPT score will be equated with ETS's TOEIC score for easier interpretation. It is expected that the DPU–CEPT will benefit both learning and teaching English at Dhurakij Pundit University in various ways, e.g., for placement and diagnosis, as well as personnel recruitment tools.

TOEIC: An Overview

Because the DPU–CEPT aims to be equated with TOEIC, a review of the TOEIC format, validity, and reliability is a crucial step for the DPU–CEPT test blueprint specifications and item writing.

TOEIC Test Format

According to the TOEIC examinee handbook (Centre for Professional Assessment [Thailand], 2006), the TOEIC is a 2–hour multiple–choice test containing 200 questions. The test aims to assess the test candidates in two main areas: Listening Comprehension (LC), and Reading Comprehension (RC). The test contents were developed to meet the language requirements of the business workplace where English is used for communication. Test questions cover everyday English use. The TOEIC test does not require specialized knowledge or vocabulary beyond that of a person who uses English in everyday work activities. Each part consists of 100 test items. The two sections of the test are subdivided as shown in Table 1, and the DPU–CEPT will use these TOEIC table of content specifications as a test blueprint for item writing and selection.

Table 1: The TOEIC Table of Contents Specifications

Part	Listening Comprehension	No. of items	No. of choices
Part I	One Picture	20 items	4
Part II	Questions-Response	30 items	3
Part III	Short conversation	30 items	4
Part IV	Short talks	20 items	4
Part	**Reading Comprehension**	No. of items	No. of choices
Part V	Incomplete sentences	40 items	4
Part VI	Error recognition	20 items	4
Part VII	Reading comprehension	40 items	4

TOEIC Validity

The ETS developed the TOEIC test by conducting needs analyses to ask companies from many nations to delineate the language functions necessary for English business communication (Educational Testing Service, 1998). The purpose of conducting needs analyses was to examine the content–related validity (Woodford, 1982). Afterwards, the ETS continued the construct validity study.

Regarding the construct–related validity of the TOEIC test, the ETS collected the massive data from an ample sample of test takers who took the TOEIC test by comparing candidates' performance with other established methods that aim to measure the same construct. The TOEIC test measures two main constructs: listening comprehension and reading comprehension. Theoretically, the ETS investigated the correlations between TOEIC and direct measures of different language skills. This was to ensure that its test could measure the identical constructs (Cronbach, 1989; Cronbach & Meehl, 1955), in other words, that it could also implicitly measure similar English language skills.

For example, the relationship between the TOEIC test and direct speaking measures (LPI: Language Proficiency Interview; ASLPR: Australian Second Language Proficiency Rating) were .66 and .70 respectively. For direct listening measures, the findings revealed that the TOEIC test was strongly related to other measures of listening comprehension tests. Concerning

reading, the TOEIC test illustrated an accurate measure of examinees' comprehension of written English. In line with other language skills, the relationships between the TOEIC test and a direct writing measure were .83; this means that the TOEIC reading comprehension score strongly correlated (Wilson, Komarakul Nagara & Woodhead, 2004).

Apart from the relationship between the TOEIC test and other direct measures, the ETS also linked the TOEIC score with classroom performance. It was found that as classroom grades increased, TOEIC Listening Comprehension, Reading Comprehension and total scores also tended to increase. Moreover, the TOEIC Listening Comprehension scores were most strongly correlated with classroom grades for listening, and also presented a moderate correlation with speaking grades (Wilson, 1999). In addition, according to Woodford (1982), it could be concluded that TOEIC and TOEFL (Test of English as a Foreign Language) were strongly related, and could be comparable.

From the aforementioned research, it is hypothesized that DPU-CEPT will be correlated with the TOEIC score, students' cumulative grade point average, and students' current grade point average.

TOEIC Reliability

With regard to its reliability, the TOEIC test applies an internal consistency measure using the KR-20 reliability index. The reliability of the TOEIC Listening and Reading section scores across all form of their samples has been approximately .90 (Educational Testing Service, 2007), which is rather high according to Nunnally's criterion (Nunnally, 1978; Nunnally & Bernstein, 1994).

Methodology

Item Selection

In the development of the DPU-CEPT, test items were drawn and selected from a test bank previously created by the DPU Language Institute. The DPU testing center administers the DPU-CEPT for DPU students annually at the beginning of the year, and selected test items are drawn from the test item bank for a particular year's test. Following the students' completion of that year's test, test items are submitted for item analysis. Four criteria were used for item selection: 1) appropriate item difficulty

index (.2 to .8); 2) high item discrimination power (.2 or more); 3) high corrected item–total correlation statistics (.2 or more); and 4) item formats consistent with ETS' TOEIC format. As a result, DPU–CEPT consists of 200 items measuring listening and reading comprehension (100 items each). The time allowed for the test is approximately 120 minutes (45 minutes for the listening task, and 75 minutes for the reading task).

Data Collection

Fifty-one participants were selected from fourth-year English students at DPU; 82.35 % were female, with cumulative grade point averages (GPAX) ranging from 1.69 to 3.69 (M = 2.78, SD = 0.47). They were invited and encouraged to take the DPU–CEPT on January 11th, 2008, and one week later, they were also invited to take the real TOEIC test without any cost. Three participants were absent from the TOEIC test; thus, the remaining 48 participants' scores were submitted for test equating and validation study. Both the DPU–CEPT and the TOEIC tests were administered as normal paper-pencil tests in order to avoid computer literacy problems. No extra credits and/ or any

penalties were applied to their classroom performance. Students were informed of the test results within a week. For this reason, the two tests were voluntarily taken for the test takers' own benefit.

Results

Evidence of Reliability of the DPU–CEPT

The DPU–CEPT showed evidence of sound psychometric properties as intended.

The KR–20 (Kuder & Richardson, 1937) showed evidence for acceptable internal consistency reliability (Nunnally & Bernstein, 1994) (.79, .68, and .78 for listening, reading, and whole test, respectively).

Evidence of Validity of the DPU–CEPT

Two types of validity, construct and concurrent, were investigated in this study.

Construct validity

Construct validity concerns the hypothetical construct that tests purport to measure. Nomological network or intercorrelations with intended constructs are the most frequently used evidence for construct validity (Cronbach & Meehl, 1955).

As the DPU–CEPT intends to measure English communicative proficiency within the business context, the correlations with the TOEIC test were investigated. The DPU–CEPT score highly correlated with the TOEIC score (r = .79, p < .01). Moreover, the DPU– CEPT listening comprehension score also highly correlated with the TOEIC listening comprehension score (r = .70, p < .01), and DPU–CEPT reading comprehension score still strongly correlated with the TOEIC reading score (*r* = .71, *p* < .01) as shown in Table 2.

Table 2: Descriptive Statistics, KR–20, and Intercorrelations Among Variables

Variables	Min	Max	M	SD	1	2	3	4	5	6	7
1.TOEIC Total Score	335	835	481.67	105.04	-						
2.TOEIC: LC	170	455	277.81	64.68	.93	-					
3.TOEIC: RC	130	380	203.85	51.23	.88	.64	-				
4.DPU Total Score	68	123	90.90	13.75	.79	.76	.67	(.78)			
5.DPU: LC	31	74	47.67	9.85	.63	.70	.40	.85	(.68)		
6.DPU: RC	27	59	43.23	7.44	.64	.47	.71	.72	.25	(.79)	
7.Current GPA	1.33	3.87	2.81	0.54	.71	.66	.62	.70	.47	.66	
8.GPAX	1.67	3.69	2.78	0.47	.57	.48	.57	.60	.37	.62	.82

Notes: LC refers to Listening Comprehension, RC stands for Reading Comprehension.
All correlations are statistically significant at .01 level.
Numbers in the parentheses on diagonal are KR–20.

Concurrent validity

Concurrent validity refers to the relationship between variables and their related criterion being measured at the same time (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

In educational settings, academic achievement, using grade point average, and cumulative grade point average (GPAX) as proxy, are among the most widely used criteria. Previous research by the ETS also found a strong correlation between classroom performance and the TOEIC score, especially among the intermediate class students (Wilson, 1999).

In this study, the DPU–CEPT also still exhibited a high level of concurrent validity. However, it should be noted

that the correlation between the DPU–CEPT score and current GPA (r = .47 to .70) is significantly higher than the correlation between cumulative grade point average (r = .37 to .62).

Equating DPU–CEPT with TOEIC Scaled Score

As the DPU–CEPT score strongly correlates with the real TOEIC score, test score equating can be done with the DPU–CEPT via regression analysis for easier score interpretation in line with the TOEIC Can–Do Guide (Chauncey Group International, 2000). Ordinary Least Squared (OLS) Regression was used as the method of equating, while its assumptions were checked thoroughly with scatter plots, residual plots, and histogram (Cohen & Cohen, 1983; Pedhazur, 1982). DPU–CEPT scores and equated TOEIC scores are presented in Table 3 below, while regression equation is displayed in Table 4.

Standard Error of Measurement (SEM)

The Standard Error of Measurement (SEM) estimates how repeated measures of a person on the same instrument tend to be distributed around his or her true score according to a classical test theory (Anastasi & Urbina, 1997; Gulliksen, 1950). It is used to calculate the confidence interval of the true score for an individual; as a result, it is very helpful for test users to interpret test scores.

To illustrate, if Mr. A earns a 90 score from test X which has a mean of 85 and a SEM of 10, a 95% confidence interval for the true score can be computed as 1.96 (z–score corresponding to 95% of the area under normal curve) multiplied by SEM. Mr. A's true score will certainly be located in the range of 90 ? 19.6 with a 95% confidence or 5% of error.

The SEM for the DPU–CEPT is also calculated via equation (1) (Gulliksen, 1950) and presented in Table 3 for both the DPU–CEPT raw score and the TOEIC scaled score.

$$SEM = SD * sqrt\ (1-r_{tt}) \text{ ----------------------------------- (1)}$$
Whereas,   SEM = Standard Error of Measurement

     SD = Standard Deviation

   r_{tt} = Reliability Coefficient

Table 3: DPU–CEPT Score vs. TOEIC Scaled Score

	DPU	TOEIC	DPU: Listening	TOEIC: Listening	DPU: Reading	TOEIC: Reading
M	90.65	480.16	47.61	277.54	43.04	202.93
SD	13.76	83.48	9.66	44.56	7.63	37.21
SEM	6.45	39.15	5.47	25.21	3.50	17.05
	123	677.45	74	399.26	59	280.74
	120	658.28	65	357.75	58	275.87
	111	603.66	64	353.14	57	270.99
	109	591.53	63	348.53	56	266.12
	104	561.19	59	330.08	54	256.37
	103	555.12	56	316.25	52	246.61
	101	542.98	55	311.63	51	241.74
	100	536.91	54	307.02	50	236.86
	99	530.84	53	302.41	49	231.99
	98	524.78	52	297.80	48	227.11
	97	518.71	51	293.19	47	222.24
	96	512.64	50	288.57	46	217.36
	95	506.57	49	283.96	45	212.49
	94	500.50	48	279.35	44	207.61
	93	494.44	47	274.74	43	202.74
	90	476.23	46	270.13	42	197.86
	89	470.16	45	265.51	41	192.99
	88	464.09	44	260.90	40	188.11
	87	458.03	43	256.29	39	183.24
	85	445.89	42	251.68	38	178.36
	83	433.75	41	247.07	37	173.49
	81	421.62	40	242.45	36	168.61
	80	415.55	39	237.84	35	163.73
	79	409.48	38	233.23	34	158.86
	78	403.41	37	228.62	33	153.98
	77	397.34	36	224.01	32	149.11
	76	391.28	35	219.39	30	139.36
	74	379.14	34	214.78	27	124.73
	73	373.07	31	200.95		
	72	367.00				
	68	342.73				

Table 4: DPU–CEPT Score vs. TOEIC Scaled Score

Dependent Variables	Independent Variables	Unstandardized Coefficient	t-test	p-value
TOEIC	constant	-69.91	-1.11	.272
	DPU	6.07	8.86	.000
TOEIC: LISTENING	constant	57.98	1.73	.091
	DPU: Listening	4.61	6.69	.000
TOEIC: READING	constant	-6.90	-0.22	.827
	DPU: Reading	4.88	6.81	.000

Discussion

The DPU–CEPT shows appropriate psychometric properties. Reliability analysis revealed that internal consistency reliability was significantly high (r_{tt} = .68 to .79); however, it is lower than TOEIC's average reliability (average r_{tt} = .90). Item unreliability can cause large SEM and wide confidence interval of the true score. Higher reliability can be attained by selecting homogeneous items meticulously, and it should be implemented in future test development. For future research, item response theory (IRT), in which item parameter estimations are sample–independent, will be more suitable for item analysis and calibration (Lord & Novick, 1968). Moreover, the IRT should be employed to investigate various item parameters, e.g., guessing parameter, difficulty index, and discrimination power for the better refinement of the test items in the DPU–CEPT.

However, the DPU–CEPT validity is quite satisfactory for both construct and concurrent validity. For construct validity, the high correlation between TOEIC and DPU–CEPT ensures that the two tests have measured the same construct, and, based on this correlation, the test score may be similarly equated. If the sample size allows doing so, future research should draw attention to the test dimensionality of the DPU–CEPT to investigate its factorial validity, similar to how the TOEIC has explored this measurement issue (Wilson, 2000). The other direct measures of English

language skills should also be investigated together with the DPU–CEPT to ascertain its construct validity. For concurrent validity, the DPU–CEPT correlates with GPA more than GPAX. Unsurprisingly, correlations between DPU–CEPT and current GPA were higher than with GPAX because the latter may be affected by many kinds of extraneous variables, e.g., inability to adjust to a new university environment, learning curiosity change, test takers' motivation, etc. However, the DPU–CEPT predictive validity should be studied through predictive validation design, especially in the case of using it as a personnel selection tool.

Another suggestion for future research arises from the practical side. Alternate or various forms of the DPU–CEPT should be further developed and revised in the near future in order to prevent students cheating on the test, testing effects, memory effects (for the case of pre–test vs. post–test research), and other practical testing issues.

Regarding the future use of the DPU–CEPT, this standardized English test will be beneficial for learning and teaching English language as it can be used as a placement test for classifying students' English language proficiency, for assigning them to suitable classes, and as a diagnostic test to indicate students' language weaknesses that need support. With its high correlation to the TOEIC test, the DPU–CEPT can also be taken as a helpful preparatory test for students who wish to see how proficient their English language is before taking the actual TOEIC test. Apart from its academic use, the DPU–CEPT can also be employed as an effective personnel recruitment tool for the Human Resource Department to screen and select staff possessing adequate English proficiency. In addition, in terms of personnel development, this standardized test can be applied as an English diagnostic test for staff to indicate which English language skills are weak and need amelioration. Clearly, the DPU–CEPT has versatile purposes for both the academic arena and the university's personnel recruitment and development units.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). **Standards for educational and psychological testing.** Washington, D.C.: American Educational Research Association.

Anastasi, A., & Urbina, S. (1997). **Psychological testing.** New Jersey: Prentice-Hall.

Center for Professional Assessment (Thailand). (2006). **TOEIC examinee handbook.** Bangkok, Thailand: Author.

Chauncey Group International. (2000). **TOEIC Can-Do Guide: Linking TOEIC scores to activities performed using English.** Princeton, NJ: Educational Testing Service.

Cohen, J., & Cohen, P. (1983). **Applied multiple regression/correlation analysis for the behavioral sciences.** NJ: Lawrence Erlbaum Associates.

Cronbach, L. J. (1989). Construct validation after thirty years. In R. L. Linn (Ed.), **Intelligence, measurement, theory, and public policy: Proceedings of a symposium in honor of Lloyd G.** Humphreys. Chicago: University of Illinois Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. **Psychological Bulletin,** 52, 281-302.

Educational Testing Service. (1998). **TOEIC technical manual.** Princeton, NJ: The Chauncey Group International.

Educational Testing Service. (2002). **TOEIC examinee handbook.** Princeton, NJ: Author.

Educational Testing Service. (2007). **TOEIC user guide.** Princeton, NJ: Author.

Gulliksen, H. (1950). **Theory of mental tests.** New York: Wiley.

Hughes, A. (1989). **Testing for language teachers.** Cambridge: Cambridge University Press.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of the test reliability. **Psychometrika,** 2, 151-160.

Lord, F. M., & Novick, R. (1968). **Statistical theories of mental test scores.** MA: Addison-Wesley.

Nunnally, J. C. (Ed.). (1978). **Psychometric theory** (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (Eds.). (1994). **Psychometric theory** (2nd ed.). New York: McGraw-Hill.

Pedhazur, E. J. (1982). **Multiple regression in behavioral research: Explanation and prediction.** Holt, Rinehart and Winston.

References

Wilson, K. M. (1993a). **Enhancing the interpretation of a norm-referenced second-language test through criterion referencing: A research assessment of experience in the TOEIC testing context. Princeton,** NJ: Educational Testing Service.

Wilson, K. M. (1993b). **Validating a test designed to assess ESL proficiency at lower developmental levels. Princeton,** NJ: Educational Testing Service.

Wilson, K. M. (1999). **Validity of global self-ratings of ESL speaking proficiency based on an FSI/ILR-referenced scale.** Princeton, NJ: Educational Testing Service.

Wilson, K. M. (2000). **An exploratory dimensionality assessment of the TOEIC test.** Princeton, NJ: Educational Testing Service.

Wilson, K. M., Komarakul Na Nagara, S., & Woodhead, R. (2004). TOEIC/LPI **relationships in academic and employment contexts in Thailand.** Princeton, NJ: Educational Testing Service.

Woodford, P. (1982). **An introduction to TOEIC: The initial validity study.** Princeton, NJ: Educational Testing Service.