Quality Evaluation of The English Proficiency Test for Automotive Specification Certificate Students at Nakhonsawan Technical College

Natanan Tungsujarittam
Nakhonsawan Technical College
yuu.natanan@gmail.com

Jiraporn Intrasai
King Mongkut's Institute of Technology Ladkrabang
kijirapo@kmitl.ac.th

Abstract

The purpose of this study was to evaluate the quality of the English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College. The 200-item test was evaluated by three experts in the field of English teaching with the use of the test quality and the Index of the Item-Objectives Congruence (IOC) forms. It was found that 154 test items with the IOC between 0.5 and 1.00 were congruent with the objectives. The rest with the IOC equal to 0.33, 0.00 and -0.33 was revised in accordance with the experts' comments and suggestions. The validated 200 English proficiency test items divided into two sets, A and B, with 100 test items in each set were used to assess two pilot groups, each of which included 30 first-year Automotive Specification diploma students. The results revealed that the difficulty levels of two test sets were moderate (p = 0.49 for Set A and 0.53 for Set B), but the discrimination levels were low (r = 0.137 for Set A and 0.173 for Set B). The reliability (KR20) was within acceptable range (KR20 = 0.492 for Set A and 0.664 for Set B). The results were discussed in terms of three assessment instruments for each test item: IOC, difficulty levels, and discrimination levels.

Keywords: English proficiency test, the item-objectives congruence, reliability, difficulty levels, discrimination levels.

Introduction

The role of English is important not only in Thailand but also in many other developing countries because English is used to communicate, negotiate and execute transactions by native speakers and non-native speakers of English. It has become a world language as two billion people will be learning English in the next decade (Graddol, 2006). Consequently, educating learners to use English efficiently and competently is crucial in this era.

Recently, Thai government has set a plan to deal with impacts on Thai economy after ASEAN Free Trade Area (AFTA) was effective in 2010. It is going to make a business deal on ASEAN Economic Community (AEC) in the forthcoming year 2015 (Mass Communications Organization of Thailand, 2010). Hence, Thai youths who will be the future professional workforces should be promoted not only in their professional skills but also in language skills, especially English. Furthermore, the government has recently approved the proposal of the Federation of Thai Industries to develop the Vocational Qualification (VQ) based on occupational competency standards. The VQ is a qualification for measuring the vocational students' competency (Office of the Education Council, 2004). Although the

government has established the VQ to support the AFTA and AEC international business agreements, English proficiency of those vocational students has never been assessed. To get ready for the AEC international agreements, Thai vocational students' English language needs to be assessed to assure their international workforce quality.

The 2002 Curriculum for the Certificate of Vocational Education aims to produce and develop skillful workers who have knowledge and vocational skills and let them experience in their areas with morals, ethics, disciplines, personality, intelligence, and ability suitable for careers in accordance with the needs of labor market by using Thai, English and other languages in daily life and careers (Ministry of Education of Thailand, 2002). Thai vocational students should be particularly concerned with not only the occupational skills but also language skills in order to be competent workers. According to Srithongkam (2008), Thai vocational students' English expressions were at low level. Consequently, to prepare the professional workforces in the world's boundless international trades, the Office of Vocational Education (OVEC) needs qualified young learners with good English skills for business communication. English language proficiency is one of the most vital criteria for workforces' employment. Thai entrepreneurs require certificate and diploma graduates with English language proficiency in three top fields, namely, Automotive Mechanic, Mechanic Tools and Maintenance, and Electricity and Electronic (Leekitwattana & Phimdee, 2007). Consequently, the OVEC has to equip vocational students with the standard qualification of English proficiency.

According to the Vocational Education Curriculum, competent graduates with the ability to communicate in English in their workplaces will be produced. As stated by Jeanne (1991), to be valuable vocational education workforce, the vocational students' professional ability has to be assessed as well as the English language proficiency. Therefore, this study aimed to design an English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College which corresponds to their English courses of the certificate curriculum. The test was used to try out the Automotive Specification diploma students who graduated with the Automotive Specification certificate level. With the use of statistical program, the test was analyzed for the difficulty level, the discrimination level and the reliability.

Although much research has been done on the English language tests, such tests were constructed for different levels of students such as secondary and tertiary levels (Chalaysap, 2005; Srithongkam, 2008; Wannarong, 1985; Wongsothorn, 1993; Yanprasart, 1982). To the best of the researchers' knowledge, there have been no tests designed for Automotive Specification Certificate students. Accordingly, this study aimed to design an English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College.

Goal and Objectives

The purpose of the study was to evaluate the quality of the English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College. The objectives of the study were as follows:
1) To design the English proficiency test for Automotive Specification Certificate students.
2) To validate the English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College

3) To analyze the test scores for the test's reliability, difficulty levels, and discrimination levels.

Method

Subjects

The subjects of the study were two groups of 30 Automotive Specification certificate graduates. They were studying in their first year diploma in Automotive Specification of the 2012 academic year at Nakhonsawan Technical College.

Research Instruments

Research instruments of this study were the English proficiency test, the evaluation form of the Index of the Item – Objective Congruence (IOC) and the evaluation form of the quality level of the test. They are described as follows:

The English proficiency test

The Automotive Specification Certificate Curriculum was used as a guideline for designing the English proficiency test. Four English-for-daily-life courses focused on listening and language use and usage and two English-for-careers courses focused on reading and writing skills. Hence, the proficiency test was designed to test such skills. In listening and language use and usage, the contents were related to English for daily life. In reading and writing, the contents were about English for careers.

The four-multiple-choice test consisted of four parts, namely listening, language use and usage, reading, and writing. It was constructed in accordance with the objectives of each part. Each part of the test contained 25 items. The total number of test items was 100. Also, the 100 additional test items were constructed to ensure that after the validating process, the total number of the quality test items was 100 and covered the objectives. After the validating process, the 200 test items were revised and divided into two sets: Set A and Set B. Both sets were tried out by two groups of 30 first year Automotive Specification diploma students. Those two test sets were analyzed for the test's reliability, difficulty levels, and discrimination levels by the statistical computer program.

The evaluation form of the Index of the Item – Objective Congruence (IOC)

The evaluation form of the Index of the Item – Objective Congruence (IOC) was employed by three experts in the field of English teaching to evaluate the congruence between the test items and the objectives.

The evaluation form of the quality level of the test

The evaluation form of the quality level of the test was designed as a questionnaire with five-point Likert scale and was used by the same three experts who evaluated the IOC to assess authenticity, interactiveness, and practicality for all four parts of the test.

Statistical Analysis

The test scores from the validated 200 test items were analyzed in terms of the test's reliability, difficulty levels, and discrimination levels.

The test validity

Harrison (1983) claimed that the test's validity is the extent to which the test measures what it is intended to measure. Hence, the Index of the Item – Objective Congruence (IOC) was used to evaluate the congruence between the test items and the objectives. The criteria are as follows:

+ 1   means the question is congruent with the objectives
0   means the question is uncertain to be congruent with the objectives
- 1   means the question is not congruent with the objectives

The questions that obtain the IOC between 0.5 – 1.0 were deemed acceptable.

The criteria of average mean scores

Likert (1970) stated that the evaluation form refer to a measure of the experts' opinions on the quality level of the test. The criteria of the average mean scores are presented as follows:

4.21 - 5.00 means the highest
3.41 - 4.20 means high
2.61 - 3.40 means moderate
1.81 - 2.60 means low
1.00 - 1.80 means the lowest

The test's reliability

To evaluate the test's reliability, the KR – 20 (Kuder Richardson) was used in the study. The test's reliability shows accurate measures of the test takers' proficiency whenever the test is used, with different students and/or different test takers (Genesee and Upshur, 1998). The KR - 20 values between 0-1. The closer the value is to the whole number 1.00, the greater the test's reliability. Conversely, the KR-20 gets close to 0.00, the low consistency of the test is. Hence, the criteria of the test's reliability at 0.60 and over are acceptable.

The difficulty level (p)

The difficulty level is the proportion between numbers of students who got correct answer and those of students who did the test (Bailey, 1998). The test's difficulty values range from 0.20 to 0.80. The criteria of the difficulty level are as follows:

> 0.95   means the test is very easy
0.81 – 0.94   means the test is quite easy
0.20 – 0.80   means the test is good
< 0.20   means the test is difficult or very difficult

The higher level the *p* gets, the easier the test is. While the lower the *p* is, the more difficult the test is. Thus the difficulty level between 0.2–0.8 represents the test with appropriate difficulty level. The test with the *p* below 0.2 and over 0.8 should be deleted.

The discrimination level (r)

The test's discrimination is the test's qualification which can discriminate students in high level group, moderate, or low group (Oller, 1979). In other words, students with high proficiency have higher scores within correct sequence of individual proficiency level. If any test can discriminate those test takers like this, it has test's discrimination. The test's discrimination values range from + 1 to -1, can be interpreted as follows:

0.40	means a perfect discrimination between high scorers and low scorers
0.30 – 0.39	means good discrimination
0.20 – 0.29	means acceptable discrimination
< 0.20	means low discrimination
≤ 0	means no discrimination, or no variance whatsoever

After the statistical analysis, the 200 test items were judged to be of good quality, deleted or revised. Criteria for judging the test items to be of good or low quality are shown below.

1. The test items with discrimination levels (r) below zero were deleted.
2. In the case that *r*-value ranged from 0.20 to 0.29 and difficulty levels (p) ranged between 0.2 and 0.7, the test items were judged to be of good quality.
3. In the case that *r*-value was lower than 0.2 but not lower than or equal to zero and the p-value ranged between 0.2 and 0.8, the test items were revised.
4. In addition to the criteria (3), when *p*-value was higher than 0.8, the test items were deleted.
5. In the case that *r*-value ranged from 0.2 to 0.29, but *p*-value was higher than 0.8 (0.81-0.875), the test items were revised because of quite easy test items but acceptable discrimination.
6. In addition to the criteria (5), when *p*-value was higher than 0.9, the test items were deleted.

Results

The purpose of the study aimed to evaluate the quality of the English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College. The quality was evaluated with the uses of four assessment instruments: IOC, test's reliability, difficulty levels, and discrimination levels. Hence, the results of the study were presented as follows:

1) The Index of the Item-Objective Congruence (IOC)

The Index of the Item-Objective Congruence (IOC) was calculated in order to evaluate the congruence between the questions and the objectives. The test items obtained the IOC

between 0.5 – 1.0 are acceptable, while the test items containing the IOC lower than 0.5 must be revised. The numbers of test item of each IOC levels are shown in Table 1.

Table 1 The numbers of test items of each IOC levels.

The Index of the Item-Objective Congruence (IOC)	The numbers of test item			
	Listening (50)	Language Use and Usage (50)	Reading (50)	Writing (50)
0.33	1	12	13	6
0.00	-	2	10	-
-0.33	-	1	1	-
0.5 -1.0	49	35	26	44

Table 1 illustrates the numbers of test item of each IOC levels. The results show that the IOC of 0.33 is found in 32, 0.00 in 12, and -0.33 in 2 test items. The IOC between 0.5-1.00 which is acceptable is found in 154 test items. The IOC of 0.33 is found in Part 1: Listening for one test item, Part 2: Language Use and Usage for 12 test items, Part 3: Reading for 13 test items, and Part 4: Writing for 6 test items. The IOC of 0.00 is found in Part 2: Language Use and Usage for 2 test items and Part 3: Reading for 10 test items. The IOC of -0.33 was found in Part 2: Language Use and Usage and Part 3: Reading for one test item each. Thus the test items containing the IOC between 0.33 to -0.33 are revised in accordance with the suggestion of the experts and the advisor.

2) **The quality level of the test**

According Bachman and Palmer (1996), a good language test should contain the following measurements qualities of the test: validity, reliability, authenticity, interactiveness, impact, and practicality. The result of the quality level of the test are shown in Table 2

Table 2 Results of the Quality of Test

Quality of test	Part 1		Part 2		Part 3		Part 4		All four parts	
	$\overline{X}$	S.D.	$\overline{X}$	S.D.	$\overline{X}$	S.D.	$\overline{X}$	S.D.	$\overline{X}$	S.D.
Authenticity	3.00	0.94	3.33	0.94	4.33	0.00	4.33	0.00	3.75	0.81
Interactiveness	3.67	0.47	3.67	0.38	3.75	0.32	3.42	0.42	3.63	0.38
Practicality	3.00	0.00	-	-	3.33	1.53	-	-	3.87	0.30
Time allocation of all four parts	-	-	-	-	-	-	-	-	4.00	0.00

Table 2 shows the result of the quality level of the test. The criteria of average mean score from 3.40–4.20 means the high quality level of the test. All three qualities of test which are authenticity, interactiveness, and practicality are at a high level with the average mean ($\overline{X}$) of 3.75 (S.D. = 0.81), 3.63 (S.D. = 0.38), and 3.87 (S.D. = 0.30), respectively. In addition, the authenticity of the test of Part 3 and 4 is at very high level ($\overline{X}$ = 4.33).

3) The statistical results of the overall test items

After the validating process, the 200 test items were revised and divided into two set: Set A and Set B. Both sets were tried out by two groups of 30 first year Automotive Specification diploma students. The statistical results of the overall test items are shown in Table 3 as follows:

Table 3 The statistical results of the overall test items

Set of Test	Numbers of the pilot group	Mean	Median	S.D.	Max	Min	p (Difficulty)	r (Discrimination)	KR-20 (Reliability)
Set A	30	50.267	52.000	5.537	59.000	40.000	0.496	0.137	0.492
Set B	30	53.967	53.000	7.045	67.000	39.000	0.539	0.173	0.664

Table 3 indicates that the difficulty level (p) of those two sets are moderate (p ≈ 0.49 – 0.53) as acceptable p-value is from 0.2 to 0.8. The discrimination levels (r) of the two test sets are lower than the acceptable r-value of 0.2. The discrimination of Set B test (r = 0.173) is slightly higher than Set A (r = 0.137). It can be stated that both tests could not discriminate the students. However, only Set B has acceptable range of the reliability (KR-20) at 0.664.

4) The comparison of the IOC, difficulty levels, and the discrimination levels

According to the criteria for judging the test items mentioned above, the comparison of three assessment instruments of IOC, the difficulty levels and the discrimination levels is shown in Table 4.

Table 4 The comparison of the IOC, the difficulty levels and the discrimination levels

Part	Item	Set A						Set B					
		IOC	p	r	Good	Revised	Deleted	IOC	p	r	Good	Revised	Deleted
Part 1: Listening	1	1.00	0.250	-0.500			✓	1.00	0.813	0.375	✓		
	2	0.67	0.500	0.500	✓			1.00	0.313	-0.625			✓
	3	1.00	0.000	0.000			✓	0.67	0.125	-0.250			✓
	4	0.67	0.688	0.625	✓			0.67	0.125	-0.250			✓
	5	0.67	0.500	-0.250			✓	1.00	0.875	0.250		✓	
	6	0.67	0.875	0.250		✓		1.00	0.125	-0.250			✓
	7	1.00	0.750	0.250	✓			1.00	0.125	-0.250			✓
	8	1.00	0.750	0.500	✓			1.00	0.250	-0.250			✓
	9	0.67	0.750	0.500	✓			0.67	0.750	0.250	✓		
	10	0.67	0.750	0.500	✓			0.33	0.250	0.000			✓
	11	1.00	0.938	0.125			✓	1.00	0.813	0.375		✓	
	12	1.00	0.375	-0.250			✓	1.00	0.250	-0.250			✓
	13	1.00	0.000	0.000			✓	0.67	0.688	0.375	✓		
	14	1.00	0.750	0.250	✓			1.00	0.688	0.125		✓	
	15	1.00	0.500	0.500	✓			0.67	0.750	0.500	✓		
	16	1.00	0.563	0.125		✓		0.67	0.813	0.375		✓	
	17	1.00	1.000	0.000			✓	0.67	0.813	0.375		✓	
	18	1.00	0.438	-0.375			✓	1.00	0.750	0.000			✓
	19	1.00	0.563	0.375	✓			1.00	0.313	-0.375			✓

Table 4 (Continued)

Part	Item	Set A						Set B					
		IOC	p	r	Good	Revised	Deleted	IOC	p	r	Good	Revised	Deleted
Part 1: Listening	20	1.00	0.625	0.500	✓			1.00	0.500	0.250	✓		
	21	0.67	0.938	-0.125			✓	1.00	0.188	0.125			✓
	22	1.00	0.188	0.125			✓	1.00	0.750	0.500	✓		
	23	1.00	0.250	0.000			✓	1.00	0.875	0.250		✓	
	24	1.00	0.000	0.000			✓	1.00	0.813	0.125			✓
	25	0.67	0.250	0.000			✓	0.67	0.250	0.000			✓
P Part 2: Language use and usage	26	0.67	0.063	0.125			✓	1.00	0.188	-0.125			✓
	27	1.00	0.125	0.000			✓	0.67	0.438	0.625	✓		
	28	1.00	0.875	0.000			✓	0.33	0.188	-0.375			✓
	29	0.33	0.000	0.000			✓	0.33	0.750	0.500	✓		
	30	1.00	0.250	-0.250			✓	1.00	0.845	0.250		✓	
	31	1.00	0.813	0.125			✓	1.00	0.750	0.250	✓		
	32	1.00	0.875	0.000			✓	0.67	0.938	0.125			✓
	33	1.00	0.125	0.000			✓	1.00	0.063	-0.125			✓
	34	1.00	0.875	0.000			✓	0.33	0.313	-0.375			✓
	35	1.00	0.813	0.375		✓		0.67	0.750	0.500	✓		
	36	0.33	0.000	0.000			✓	1.00	0.875	0.250		✓	
	37	1.00	0.875	0.250		✓		1.00	0.188	0.125			✓
	38	1.00	0.688	-0.125			✓	1.00	0.125	-0.250			✓
	39	1.00	0.875	0.000			✓	0.33	0.125	0.000			✓
	40	1.00	0.750	0.250	✓			1.00	0.938	0.125			✓
	41	1.00	0.063	0.125			✓	0.33	0.813	0.375		✓	
	42	1.00	0.063	-0.125			✓	0.33	0.125	-0.250			✓
	43	0.33	0.188	0.125			✓	0.33	0.438	0.625	✓		
	44	0.33	0.313	0.125		✓		0.00	0.813	0.125			✓

Table 4 (Continued)

Part	Item	Set A						Set B					
		IOC	p	r	Good	Revised	Deleted	IOC	p	r	Good	Revised	Deleted
Part 2: Language use and usage	45	-0.33	0.188	-0.375			✓	0.67	0.313	0.375	✓		
	46	1.00	0.500	0.000			✓	0.33	0.438	0.375	✓		
	47	1.00	0.125	0.000			✓	1.00	0.563	0.375	✓		
	48	1.00	0.688	0.125		✓		1.00	0.188	0.375		✓	
	49	1.00	0.125	0.000			✓	1.00	0.313	-0.125			✓
	50	1.00	0.313	0.375	✓			0.00	0.188	-0.125			✓
Part 3: Reading	51	1.00	0.188	-0.125			✓	1.00	0.813	0.125			✓
	52	0.33	0.750	0.500	✓			1.00	0.563	0.625	✓		
	53	0.33	0.250	0.250	✓			0.00	0.563	0.875	✓		
	54	0.00	0.875	0.250		✓		0.33	0.688	0.375	✓		
	55	1.00	0.625	0.500	✓			1.00	0.625	0.500	✓		
	56	1.00	0.063	-0.125			✓	1.00	0.313	-0.125			✓
	57	1.00	0.750	0.500	✓			0.00	0.813	0.125			✓
	58	1.00	0.688	0.375	✓			0.33	0.750	0.500	✓		
	59	-0.33	0.875	0.250		✓		0.00	0.000	0.000			✓
	60	0.00	0.875	0.250		✓		0.00	0.688	0.625	✓		
	61	0.33	0.250	0.000			✓	0.33	0.438	0.375	✓		
	62	0.33	0.625	0.750	✓			0.33	0.688	0.375	✓		
	63	0.33	0.688	0.625	✓			0.33	0.625	0.250	✓		
	64	0.33	0.500	0.500	✓			0.33	0.625	0.750	✓		
	65	1.00	0.063	0.125			✓	1.00	0.500	0.500	✓		

Table 4 (Continued)

Part	Item	Set A						Set B					
		IOC	**p**	**r**	**Good**	**Revised**	**Deleted**	**IOC**	**p**	**r**	**Good**	**Revised**	**Deleted**
Part 3: Reading (Continued)	66	1.00	0.625	0.750	✓			1.00	0.625	0.500	✓		
	67	1.00	0.750	0.500	✓			1.00	0.750	0.500	✓		
	68	1.00	0.000	0.000			✓	0.67	0.625	0.750	✓		
	69	0.33	0.750	0.250	✓			0.67	0.375	-0.500			✓
	70	1.00	0.813	0.125			✓	1.00	0.500	0.750	✓		
	71	1.00	0.688	0.625	✓			1.00	0.500	0.250	✓		
	72	1.00	0.750	0.500	✓			1.00	0.688	0.125		✓	
	73	1.00	0.438	0.375	✓			1.00	0.625	0.250	✓		
	74	0.00	0.375	0.000			✓	0.00	0.688	0.375	✓		
	75	0.00	0.750	0.250	✓			0.00	0.313	0.375	✓		
Part 4: Writing	76	0.33	0.750	0.250	✓			0.67	0.875	0.000			✓
	77	0.67	0.313	-0.375			✓	0.67	0.875	0.250		✓	
	78	0.33	0.813	0.125			✓	0.67	0.813	0.375		✓	
	79	0.67	0.000	0.000			✓	0.67	0.063	0.125			✓
	80	0.67	0.125	0.000			✓	0.67	0.313	-0.375			✓
	81	1.00	0.125	0.000			✓	0.33	0.063	-0.125			✓
	82	1.00	0.063	0.125			✓	0.33	0.750	0.500	✓		
	83	0.67	0.875	0.000			✓	0.67	0.625	0.750	✓		
	84	0.67	0.063	0.125			✓	0.33	0.250	0.000			✓

Table 4 (Continued)

Part	Item	Set A						Set B					
		IOC	p	r	Good	Revised	Deleted	IOC	p	r	Good	Revised	Deleted
Part 4: Writing (Continued)	85	0.67	0.188	-0.125			✓	0.33	0.875	0.250		✓	
	86	1.00	0.938	0.125			✓	1.00	1.000	0.000			✓
	87	1.00	0.188	-0.125			✓	1.00	0.938	0.125			✓
	88	1.00	0.938	0.125			✓	1.00	0.813	0.375		✓	
	89	1.00	1.000	0.000			✓	1.00	0.938	0.125			✓
	90	0.67	0.875	0.250		✓		1.00	1.000	0.000			✓
	91	1.00	0.938	0.125			✓	1.00	0.188	-0.125			✓
	92	1.00	0.125	0.000			✓	1.00	0.063	-0.125			✓
	93	1.00	0.063	-0.125			✓	1.00	0.875	0.000			✓
	94	1.00	0.938	0.125			✓	1.00	0.875	0.250		✓	
	95	1.00	0.125	-0.250			✓	1.00	0.750	0.250	✓		
	96	1.00	0.688	0.125		✓		1.00	0.313	0.125		✓	
	97	1.00	0.313	-0.125			✓	0.67	0.500	-0.250			✓
	98	1.00	0.250	0.000			✓	1.00	0.313	0.375	✓		
	99	1.00	0.625	0.500	✓			1.00	0.250	-0.250			✓
	100	1.00	0.563	0.375	✓			1.00	0.563	0.375	✓		

Table 4 shows that p and/or r values are not consistent with the IOC values in 109 test items (59 test items of Set A and 50 items of Set B). That is, while the IOC values are higher than 0.5, which are judged to be of good validity, p and/or r values are not in the acceptable levels. Moreover, there are 27 test items (12 test items of Set A and 15 items of Set B) containing the IOC values lower than 0.5, but p and /or r values are at acceptable levels. There are 64 test items the IOC values of which are consistent with the p and r values. Forty-five good test items contain the IOC values higher than 0.5 and the acceptable p and r values. Furthermore, there are 19 poor test items (8 test items of Set A and 11 items of Set B) with the IOC values lower than 0.5 and unacceptable p and r values.

Discussion

The purpose of the study was to evaluate the quality of the English proficiency test for Automotive Specification Certificate students at Nakhonsawan Technical College. The test scores of the validated 200 English proficiency test items were analyzed for the test's reliability, difficulty levels, and discrimination levels.

Regarding the test items with the acceptable IOC, p, and r values, the test item 100 of Set A, for example, is shown below:

Choose the one word or phrase that best completes the resume.
 	
RESUME
Krisada Srichan
18 Jesadabordin Rd.
Muang, Uttaradit 53000
Tel: (055) 441-682 Fax: (055) 441-683
email: krisri@gmail.com
Personal Detail
Age: 23
_____96___: Single Nationality: Thai Date of Birth: February 1, 1989
_____97___: Uttaradit, Thailand
_____98___
2009 – 2011 Uttaradit Technical College
          (Diploma in Automotive Specification with
          GPA 3.50)
	2005 – 2008 Uttaradit Technical College
	          (Certifcate in Automotive Specification with
	          GPA 3.50)
	Working Experience
	March – April 2010 Maintenance trainee at Uttaradit TOYOTA, Uttaradit
	May – August 2008 Maintenance trainee at _____99_____, Uttaradit
	Interests / Activities
	• Committee of Thailand Future Technician Organization
	• Football
	• Crosswords
	_____100_____
	• Proficient with Microsoft Word, Excel, and PowerPoint, and Internet

100. (A) **Computer Skills** (**H:L = 6:3, p = 0.563, r = 0.325**)
    (B) Public Relation (H:L = 0:4, p = 0.250, r = 0.500)
    (C) Interpersonal Skills (H:L = 2:0, p = 0.125, r = -0.250)
    (D) Correspondence (H:L = 0:1, p = 0.063, r = 0.125)

According to the appropriate difficulty level ($p = 0.563$) and the acceptable discrimination level ($r = 0.325$), it can be seen that more high scorers ($H = 6$) than low scorers ($L = 3$) chose the correct answer A. The distractors B and D could well trick the low scorers because many low scorers but no high scorers chose them. As a result, the factors that might influence the well-designed test with the acceptable values of IOC might be derived from test takers. The result is consistent with Sukamonsan's study (2006). He stated that test's quality is very high since test designers are experts in language testing and test takers are intellectual. That is, knowledgeable test takers choose the correct answer, while the good distractors are chosen by examinees who do not know the content.

However, as regards a number of the students who selected the wrong answer (distractor C), more high scorers than low scorers chose the distractor C. It might be because of the students' response patterns of plodding and guessing. According to Haladyna (1999), students may not have enough time to answer all items due to their plodding nature. Plodding test takers are very careful and meticulous in approaching each item but lack a test-taking skill to encourage time management strategies. As the test item was the last item, in this study, the test takers might presume that the test time was over. Hence, they might guess for the correct answer.

For some items, the IOC values were lower than acceptable level while the p and r values were at acceptable levels. For instance, the test item 52 of Set A is shown below:

52.   What tool is used to measure the dimension of a component?

(A) Shafts   (H:L = 0:2, p = 0.125, r = 0.250)
(B) Dimensions   (H:L = 0:1, p = 0.063, r = 0.125)
(C) Micrometers   **(H:L = 8:4, p = 0.750, r = 0.500)**
(D) Components   (H:L = 0:1, p = 0.063, r = 0.125)

It is plausible that the test item was revised well before it was piloted. The low scorers were distracted by the good distractors. Moreover, the high scorers chose the correct answer. As claimed by the study of Wannarong (1985), the tests were analyzed, revised and selected appropriately so that they were at acceptable levels of p and r values.

However, there were test items with unacceptable p and r values. Those test items were divided into three groups as follows:

Firstly, more high scorers chose the distractor, as shown in the test item 30 of Set A:

30.   (man):   _____, Marie?
      (woman):   Me? I'm a salesperson.
      (A) How are you   (H:L = 7:4, p = 0.688, r = -0.375)
      (B) How do you do   (H:L = 0:1, p = 0.063, r = 0.125)
      (C) What do you do   **(H:L = 1:3, p = 0.250, r = -0.250)**
      (D) What are you doing   (H:L = 0:0, p = 0.000. r = 0.000)

It can be seen that the correct answer is C, the question about a job. However, most of the high scorers identified the distractor A (H = 7) as the correct answer probably by noticing the use of the verb to be "am" in the sentence "I'm a salesperson." Also, they might be familiar with the basic greeting expression "How are you?". On the other hand, the distractor D made with the use of the verb to be "are" followed by the present participle form of the verb "do" (doing) was not selected (H:L = 0:0). It might be the case that the students relied on the surface match between the stem "I'm a salesperson." and the distractor D without understanding the item. That is, the verb "do" did not appear in the stem. As claimed by Linn and Gronlund (1995), the students who depend on the rote memory and verbal associations will then be led away from, the correct answer.

Secondly, more low scorers than high scorers chose the correct answer, as shown in the test item 38 of Set A:

38. (woman):   What does your sister look like?
     (man):   _____.
     (A) She's a super star.   (H:L = 0:1, p = 0.063, r = 0.125)
     (B) She loves going out.   (H:L = 1:1, p = 0.125, r = 0.000)
     (C) She likes tennis.   (H:L = 2:0, p = 0.125, r = -0.250)
     (D) She's got long curly hair.   **(H:L = 5:6, p = 0.688, r = -0.125)**

For this test item, the question "What does your sister look like?" was related to personal description. Almost the same number of high and low scorers chose the correct answer (H:L = 5:6). It is possible that the item was so easy (p = 0.688) that the low scorers guessed the correct answer. According to Matlock–Hetzel (1997), a poor student may make a guess, select that response, and come up with the correct answer. Fewer high scorers than low scorers, moreover, chose the correct answer D, and a small number of high scorers chose the distractors B and C. As claimed by Matlock–Hetzel (1997), good students may be suspicious of a question that looks too easy. They may take the harder path to solving the problem, and

read too much into question. At that point, they may end up being less successful than those who guess.

Lastly, no low scorers chose the distractor, as shown in the test item 32 of Set A:

32. (man):   _____.
     (woman):   Let me see. It's ten fifteen.
     (man):   Thanks.
     (A)   Is time up?   (H:L = 0:1, p = 0.063, r = 0.125)
     (B)   **Do you have the time?**   (**H:L = 7:7, p = 0.875, r = 0.000**)
     (C)   Can I borrow your watch?   (H:L = 1:0, p = 0.063, r = -0.125)
     (D)   Are you a time keeper?   (H:L = 0:0, p = 0.000, r = 0.000)

The item was very easy as the p–value was very high at 0.875 and could not discriminate between high and low scorers. The result supports Phinit-akson's study (1976). He stated that a difficult exam always discriminates between good and poor students better than an easy exam. An easy exam may not discriminate between good or poor students at all. According to the sentence "It's ten fifteen.", it is interpreted as time telling. Though all options identify the pattern of question, only the distractor A and the keyed response B ask about the time while the distractors C and D represent requesting and asking for information, respectively. It can be seen that the distractors A and D show Yes/No Question which was not related to the item. The distractor A were chosen by only one low scorer (L = 1), while the distractor D drew no responses. There was no discrimination between high and low scorers. On the contrary, the correct answer B was chosen by the same number of high and low scorers (H:L = 7:7). It was very obvious. However, one of the high scorers chose the distractor C. The possible explanation is that the item was about time telling. Since the watch could tell the time, the students might guess the answer from the word "watch" to complete the time telling. Thus, the distractors A, C and D must be rewritten as they could not trick test takers; they were poor distractors. As claimed by Professional Testing Inc. (2006), the distractors must be clearly incorrect but plausible. That is, the distractors should seem reasonable to an examinee who is not sufficiently knowledgeable in the content area to choose them. If a distractor appears so unreasonable that almost no examinee will select it, it is not contributing to the performance of the item. In fact, implausible distractors in a multiple choice item can make the item artificially far easier than it ought to be.

Suggestions for further studies

Another area of concerns for teachers, in addition to teaching methodologies and materials, would be the test quality, including validity, reliability, discrimination, and difficulty. In the present study, some test items cannot perform well in separating better students from weaker ones. It can be seen from the evidence that some distractors drew no responses and that the same number of high and low scorers chose the keyed response or correct answer. Meanwhile, many high scorers chose some distractors. Such evidence could guide the improvement of the test items. For example, it may be necessary to alter the syntax or vocabulary of a distractor, or perhaps its semantic thrust. Then, the test must be piloted again.

However, one of the drawbacks to the multiple-choice format is guessing, which can lead to low reliability. Guessing is another strategy that the test takers may use. They may

have partial knowledge that allows for the elimination of implausible distractors. They may also make a guess among the remaining options or simply guess in the absence of any knowledge (Haladyna, 1999). Moreover, test takers without necessary knowledge may select the correct answer (Suen, 1990). Accordingly, it is possible that the quality evaluation of the test can include guessing – an assumption of how test takers without necessary knowledge can select the correct answer (Suen, 1990).

In addition, language test administration is of major concern. It involves a variety of procedures for actually giving a test and also for collecting empirical information in order to evaluate the qualities of test and make inferences about test takers' abilities (Bachman and Palmer, 1996).

Futhermore, the tests in the present study were constructed in accordance with the English courses offered in the Vocational Education Curriculum. The test scores are, however, bound to raise questions about the knowledge of the test takers. Such knowledge would bring about the reliable results of the test quality. As a result, material designs or teaching methodology may be investigated in order to find out whether the test takers are knowledgeable and good enough to be qualified subjects of the language testing research.

References

Bachman, L. F & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language test.* New York: Oxford.

Bailey, K. M. (1998). *Definitions from learning about language assessment: Dilemmas, decisions, and directions.* California: Heinle & Heinle Publishers.

Chalaysap, N. (2005). English Proficiency of master's degree students to be graduating from National Institute of Development Administration. *Journal of Development Administration, 45*(1), 1-26.

Genesee, F. & Upshur, A. J. (1998). *Classroom-based evaluation in second language education.* UK: Cambridge University Press.

Graddol, D. (2006). *English next.* London: British Council.

Haladyna, T. M. (1999). *Developing and validating multiple-choice test items.* (2nd ed). New Jersey: Lawrence Erlbaum Associates.

Harrison, A. (1983). *A language testing handbook.* Hong Kong: Macmillan.

Jeanne, L.V. (1991). *Succeeding with the LEP in vocational education: Common concerns and solution.* Illinois: Vocational and Technical Education.

Leekitwattana, P. & Phimdee, P. (2007). *The demand of technical workforce with industrial major by the entrepreneurs in industrial estate.* Bangkok: King Mongkut's Institute of Technology Ladkrabang.

Likert, W. J. (1970). *Research in education.* (2nd ed). Englewood Cliffs, New Jersey: Prentice-Hall.

Linn, R. L. & *Gronlund*, N. E. (1995). *Measurement and assessment in teaching*. (7th ed**.**) New Jersey: Merrill.

Mass Communications Organization of Thailand. (2010). *PM Assures Thailand's Readiness to Cope with AFTA Impacts*. Retrieved January 12, 2011 from http://enews.mcot.net/view.php?id=13417&t=1

Matlock-Hetzel, S. (1997). *Basic concepts in item and test analysis*. Retrieved January 20, 2014 from http://ericae.net/ft/tamu/Espy.htm

Ministry of Education, Thailand. (2002). *Curriculum for certificate of vocational education*. Retrieved December 28, 2010 from http://bsq.vec.go.th/course/Eng/certificate_eng.pdf

Office of Education Council. (2004). *Education in Thailand 2004.* Bangkok: Amarin Printing and Publishing.

Oller, W. Jr. (1979). *Language tests at school.* London: Longman.

Phinit-akson, V. (1976). *Modern language testing (Testing English as a second language).* Bangkok: Thammasat University Press.

Professional Testing Inc. (2006). *Step 9 conduct the item analysis.* Retrieved December 3, 2013 from http://www.proftesting.com/test_topics/pdfs/steps_9.pdf

Srithongkam, C. (2008). *A Study of the English Conversational Competency of Vocational Students*. Master of Art thesis, Ramkhamhaeng University.

Suen, K. H. (1990). *Principle of test theory.* New Jersey: Lawrence Erlbaum Associates.

Sukamonsan, S. (2006). *A standardization of CULI test of English proficiency for occupational and international communication purposes and their testing programs.* Bangkok: Chulalongkorn University

Wannarong, D. (1985). *The design of diagnostic tests of reading comprehension for Mathayomsuksa 3*. M.Ed. Thesis, Srinakharinwirot University.

Wongsothorn, A. (1993). *Levels of English skills of Thai students.* Bangkok: Chulalongkorn University Language Institute.

Yanprasart, Y. (1982). *The design of the diagnostic reading test of word recognition skills for Mathayomsuksa 1 in Chiangmai*. M.Ed. Thesis, Srinakharinwirot University.