A study on CSL proficiency evaluation-Reading and Listening subject

Rih-Chang Chao, Bor-Chen Kuo, Hsuan-Po Wang, Ya-Hsun Tsai

***Abstract*—**Presently there are many Chinese proficiency tests (CPTs) available today measuring participants' proficiency in CSL. Most of them, the AP Chinese language and Cultural examination (AP), Hanyu Shuiping Kaoshi (HSK) and Test of Proficiency-Huayu (TOP), had classified their test results in proficiency levels which corresponded to the levels of CEFR (The Common European Framework of Reference for Languages: Learning, Teaching, and Assessment). However, some of Reading and Listening subjects in Top had not completely conducted their proficiency level in corresponding to CEFR. Therefore, the items implemented in this study were on the basis of CEFR for CSL CPT reading and listening subject construction. This study applied IRT 3PL model to analyze and interpret 751 reading and 762 listening subjects empirical data collected from Grace Christian Collage in Philippine on September 2009 via the computerized based test (CBT). The contribution of this study was not only on the construction of a CSL Proficiency Test on a basis of CEFR but also in comparison with examinees' proficiency scales in referring to their background and explored factors that might affected CSL learning effectiveness.

***Keywords*—**CEFR, Chinese as Second Language, Proficiency Test, CSL proficiency scales

I. Introduction

Under the globalization market, multi-language proficiency becomes very important in the competitive business industry or other sectors of industries today. One of the examples is the recent fever in learning Chinese as second language (CSL). The ability of using Chinese language to communicate with others is another important area which has been neglected in the levels of Chinese Language proficiency test. The Chinese Language prerequisite of entrance to sectors of education or job demands increases which initiates the motivation of participants to take Chinese proficiency test (CPT).

To measure examinee's proficiency and to classify them according to their proficiency levels accurately and effectively has to do with the importance of item implementation incorporate with proficiency index during CPT constructed. Presently there are many CPT in the world today, classified their test results in proficiency levels which corresponded to the levels of CEFR (The Common European Framework of Reference for Languages: Learning, Teaching, and Assessment). The purpose of enable the comparison of proficiency scales between different tests was to distinguish the discrepancy of examinee's proficiency with others further justifies their curriculum engagement. For examples TOEIC, TOEFL, BULATS, TestDaF, and DELF etc. [1, 2]. Regarding as CSL CPT in Taiwan (TOP), TOP Speaking and Writing subjects had also conducted their proficiency scales comparisons corresponding to CEFR. However, some of the TOP Reading and Listening subjects had not completely conducted their proficiency scales comparisons corresponding to CEFR. The proficiency scales comparison between TOP and CEFR had shown below Table 1. Therefore, the item implementation in this study was on the basis of CEFR B1 level for CSL CPT Reading and Listening subject construction.

Table 1. The proficiency level comparison between TOP and CEFR

TOP/ Reading & Listening	CEFR		TOP/ Speaking	CEFR		TOP/ Writing	CEFR
Beginner	A2		Beginner	A2		Beginner	A2
Basic	N/A		Learner	B1		Learner	B1
Intermediate			Superior	B2		Superior	B2
Advanced			Master	C1			

Resource: TOP website

The examinee's proficiency was analyzed from the CSL CPT results based on the Classical Test Theory (CTT). CTT applied observed score (raw score) to classify examinees CSL proficiency level[3, 4]. For example, the proficiency scale of HSK used on the test report was transformed from the raw score [5, 6]. Other example, the test report on SAT was indicated in three different proficiency scales such as raw score, composite total score, and percentile at the same time [7]. The validity of using raw score to represent proficiency scale of a test is not based on the assumptions of meaningful measurement, unidimensionality, linearity, and mutuality of data characters. In addition, the test cut scores were unable to distinguish the proficiency scale of an examinee who participates in different test with more difficult items. Another word that different test could measure examinee's proficiency differently while applying CTT model. On the contrary, IRT model overcomes all the shortcomings of CTT model [8]. This study applied IRT three-parameter logistic (3PL) model to analyze and interpret

751 reading and 762 listening subjects empirical data collected from Grace Christian Collage in Philippine on September 2009 via the computerized based test (CBT). The CSL proficiency scales will be constructed and established to enable the comparison between any two tests in same proficiency level but different item difficulties. The results and conclusions were also indicated at the end of this paper as well as the future study directions.

II. Methodology

A. Item Implementation

CEFR is intended to overcome the barriers to communication among professionals working in the field of modern languages arising from the different educational systems in Europe. It provides the means for educational administrators, course designers, and teachers, teacher trainers, examining bodies, etc., to reflect on their current practice, with a view to situating and coordinating their efforts and to ensuring that they meet the real needs of the learners for whom they are responsible [9]. There are A1, A2, B1, B2, C1, and C2 six different levels in CEFR represented the magnitude of communication proficiency from basic to proficient user respectively.

CEFR emphasized the importance of communication proficiency but also classified and defined communication activities and strategies into four different categories. The content of those four categories refer to as the proficiency indexes. Those four categories are productive activities and strategies (Productive), receptive activities and strategies (Receptive), interactive activities and strategies (Interactive), and mediating activities and strategies (Mediating). CEFR incorporate those four categories individually included reading, listening, speaking, and writing. This is the major difference between CEFR and other proficiency indexes, others which simply classified proficiency description in reading, listening, speaking, and writing only.

The proficiency index of learning English as Second Language (ESL) in Taiwan was adopted CEFR as the reference of the proficiency index. Tsai [10] compared the language differences between English and Mandarin, after the comparison the differences by means of CEFR to modified these four different activities and strategies to meet the needs for CSL learners. Therefore, the study applied CEFR (hereinafter referred to the one modified by Tsai) as the reference on the basis for items implementation during this CSL Proficiency Test construction.

B. Application of IRT

In reference to the test theory, it is divided into CTT and IRT [11]. CTT proposed a mathematical model (CTT model) refer to (1) below and used the observed score to describe the characteristics of examinees' performance on the test.

X (observed Score) = T (true score) + E (error) (1)

According to the definition of CCT model, its item difficulty is defined as the proportion of examinees correctly responds an item [12]. On the other words, the higher proficiency the examinee had, the higher probability of correctly answer a difficult item (correct percentage); if the proficiency of the examinee is comparatively low, then the probability of correctly answer a difficult items (correct percentage) would become low [13]. This imply that an item either easy or difficult is totally depends on the sample of examinee's ability. Therefore, CTT model is a type of answer behavior model which has nothing to do with a mathematical model. CTT is a theory of sample dependent. On the contrary, the scale of IRT model is separately constructed from the examinee's attribute distribution [13]. The major reason is due to IRT model has incorporated the item parameters and examinee's proficiency into the model [8]. Therefore, IRT model is a type of answer behavior model which consisted with a mathematical model. IRT is a theory of sample free. Due to the limitations and constrains of CTT model, this study had adopted IRT model as the application model for CSL CPT.

One of the major objectives in this study was the selected-response items implemented while constructing CSL CPT on reading and listening subjects. Those dichotomous data had most often been analyzed by the following three types of IRT models: one-parameter logistic (1PL) model proposed by Rasch [14], two-parameter logistic (2PL) model and 3PL model proposed by Lord [15]. These three different IRT models were described as below [16-18].

In IRT 3PL model, the possibility of kth examinee correctly response jth item was refer to (2) as below：

$$P(x_j = 1 \mid \theta_k, a_j, b_j, c_j) = c_j + \frac{(1-c_j)}{1+\exp^{-D^* a_j(\theta_k - b_j)}} \equiv P_{j1}(\theta_k) \quad (2)$$

where x_j represent the jth item response (1 means correctly answer this item, otherwise 0).

a_j represent the jth item discrimination parameter, and $a_j > 0$;

b_j represent the jth item difficulty parameter, and $-\infty < b_j < \infty$;

c_j represent the jth item guessing parameter, and $0 \le c_j < 1$;

D represent the scaling factor, and assume $D = 1.702$

3PL model was assumed that there was guessing phenomenon occurred during the test administration [8, 19]. On the contrary, if there was no guessing phenomenon appeared during the test administration ($c_j = 0$), the 3PL model in equation (2) will be transformed to 2PL model equation refer to (3) as below：

$$P(x_j = 1 \mid \theta_k, a_j, b_j) = \frac{1}{1+\exp^{-D^* a_j(\theta_k - b_j)}} \equiv P_{j1}(\theta_k) \quad (3)$$

1PL model was a special case of 2PL model by defined the

item discrimination parameter equal to 1($a_j = 1$). Therefore, the 2PL model equation (3) will be transformed to 1PL model equation refer to (4) as below：

$$P(x_j = 1 | \theta_k, b_j) = \frac{1}{1 + \exp^{-D(\theta_k - b_j)}} \equiv P_{j1}(\theta_k) \qquad (4)$$

IRT 1PL model only considered the proficiency estimation variation caused by item difficulty parameter (b_j) proposed by Rasch in 1960 [14]. Therefore, IRT 1PL model also named as Rasch model.

There were 32 reading items and 32 listening items implemented in this study. All of them were B1 level with four option multiple choice items. According to the item fit analysis between IRT 1PL, 2PL, and 3PL models, this study applied IRT 3PL model to analyze and interpret these empirical data.

III. Model Selection

There were two subjects of CSL CPT constructed in this study. One is reading subject and the other is listening subject. This study collected 751 reading and 762 listening subject empirical data for analysis and interpretation. Those data were collected from grade 7 to 10 students of Grace Christian Collage in Philippine on September 2009 via CBT.

Both of the reading and listening subject of CSL CPT conducted in this study were approximately 30 minutes each in length. The test structure of these two subjects was constructed as below:

A. Reading Subject

The reading subject had included three sections and consisted of 32 multiple-choice items. The first section was consisted of 14 items of reading selection of syntax. The second section was consisted of 15 items of reading selection of single communicating sentence. The third section was consisted of 3 items of reading selection related to the materials used in daily life. In addition, the items implementation of reading subject incorporated with CEFR four categories had shown as below table 2.

Table 2. Items implementation of reading subject incorporated with CEFR

Category	Item Sequence No.	Item No.	(%)
Productive	3, 6, 7, 12, 17, 18, 22, 23, 26, 28	10	31.25
Receptive	4, 5, 10, 11, 15, 20, 21, 31, 32	9	28.13
Interactive	8, 9, 13, 14, 16, 24, 25, 27	8	25
Mediating	1, 2, 19, 29, 30	5	15.63

B. Listening Subject

The listening subject had also included three sections and consisted of 32 multiple-choice items. The first section was consisted of 15 items of listening selection of single conversation sentence. The second section was consisted of 15 rejoinder items. The third section was consisted of 2 items of listening selection of a passage. In addition, the item implementation incorporated with CEFR four categories had shown as below table 3.

Table 3. Items implementation of listening subject incorporated with CEFR

Category	Item Sequence No.	Item No.	(%)
Productive	14, 29	2	6.25
Receptive	3, 5, 6, 7, 13, 15, 17, 31	8	25
Interactive	1, 2, 4, 8-12, 16, 18-28, 30	21	65.63
Mediating	32	1	3.13

C. Item Fit Analysis

The model fits the data more, the more accuracy of the data analysis resulted[20]. Therefore, the objective of the item fit analysis was to select the most fitness model for data analysis in this study. This study utilized BILOG-MG [21] and EAP method for item parameter estimation and conducted the item fit comparison between IRT 1PL, 2PL and 3PL model. For reading subject, there were 18 not fit items with IRT 1PL model application compared with 3 and 1 not fit items with IRT 2PL and 3PL model respectively. This comparison, shown as below table 4, indicated that the IRT 3PL model fit the data more compared with 1PL and 2PL model for data analysis on reading subject.

Table 4. IRT Model Comparison for CSL CPT Reading Subject

Reading Subject of CSL CPT	
IRT Model	Item not fit (P<.05)
1PL	1, 3 , 5, 9, 10, 12, 13, 16, 17, 18, 19, 20, 23, 24, 25, 26, 29, 32
2PL	9, 16, 18
3PL	12

For CSL CPT listening, there were 16 not fit items with IRT 1PL model application compared with 3 and 3 not fit items with IRT 2PL and 3PL model respectively. This comparison, shown in below table 5, indicated that the IRT 3PL model fit the data more compared with 1PL and 2PL model for data analysis on listening subject.

Table 5. IRT Model Comparison for CSL CPT Listening Subject

Listening Subject of CSL CPT	
IRT Model	Item not fit (P<.05)
1PL	2, 3, 8, 10, 11, 15, 16, 20, 21, 22, 23, 24, 25, 26, 27, 30
2PL	11, 26, 27
3PL	15, 26, 30

In addition, according to the internal consistency reliability analysis for item evaluation applied IRT 3PL model, the reliability values were up to 0.848 and 0.844 (Cronbach's α) for reading and listening subject respectively. Therefore, this study applied IRT 3PL model for reading and listening subject data analysis.

IV. Data Analysis

A. Item Analysis

The average of item discrimination parameter on reading subject was 1.254 and average item difficulty and guessing parameters were 0.438 and 0.249 respectively. Other than this, the average of item discrimination parameter on listening subject was 1.116 and average item difficulty and guessing parameters were 0.351 and 0.245 respectively. These results indicated that the B1 level items implemented in this study could effectively discriminate examinees' proficiency.

B. Examinee Analysis

The average proportion of all examinees correctly responded the items incorporated with CEFR four categories in both reading and writing subject had shown in below table 6. The results of the examinees' CSL proficiency from high to low in reading subject were 56.40%, 55.15%, 54.36%, 40.66% for Receptive, Productive, Interactive, and Mediating respectively. In addition, there were 61.68%, 59.51%, 51.30%, 41.62% for Productive, Receptive, Interactive, and Mediating respectively on listening subject.

Table 6. Average proportion regarding incorporate with four categories

Category	Reading Subject		Listening Subject	
	Item No. (%)	Propotion	Item No. (%)	Propotion
Productive	10 (31.25%)	55.15%	2 (6.25%)	61.68%
Receptive	9 (28.13%)	56.40%	8 (25%)	59.51%
Interactive	8 (25%)	54.36%	21 (65.63%)	51.30%
Mediating	5 (15.63%)	38.68%	1 (3.12%)	41.62%
Total	32	52.73%	32	53.70%

Most of the average proportion of all examinees correctly responded the items incorporated on both reading and listening subject were less than 60%. Overall speaking, all the examinees' CSL proficiencies were below B1 level except Productive in Listening subject. One of the major reasons of examinees more capable utilizing productive activities and strategies during their listening communication might due to their individual background.

V. Examinees Proficiency Differences Analysis

The analysis of examinees' proficiency differences was based on proficiency estimated by IRT 3PL model. The analysis compared the examinees' proficiency difference on the basis of their individual background and explored the factors that might affect CSL learning effectiveness. These factors including gender, ethnicity, grade, parents' education, regular home language speaking or hearing, and traveling experience were analyzed and explored as following:

A. Gender

The examinees' proficiency comparison on the basis of their gender differences that might affect CSL learning effectiveness was shown in below table 7. The average CSL proficiency was -0.177 and 0.132 and the standard deviation was 1 and 0.978 for male and female respectively on listening subject. The proficiency of female examinees was significantly different and relatively higher than male examinees had on listening subject. On the contrary, there were no significantly different between male and female examinees' proficiency on reading subject.

Table 7. Examinees' Ability Comparison on Gender Difference (*** $p < .001$)

Variables		Number	Mean	S.D	t value
Listening	Male	326	-0.177	1	-4.272***
	Female	436	0.132	0.978	
Reading	Male	310	0.028	1.025	0.634
	Female	441	-0.019	0.984	

B. Regularly Language Speaking or Hearing at Home

The examinees' proficiency comparison on the difference of their regular home language speaking or hearing was shown in below table 8. The average CSL proficiency was 0.233 and -0.136 and the standard deviation was 1.048 and 0.947 for examinees who regularly speak or hear Chinese at home and examinees do not regularly speak or hear Chinese at home on listening subject respectively. The proficiency of examinees that regularly speak or hear Chinese at home were significantly different and much higher than who didn't on listening subject. On the contrary, there were no significantly different between them in CSL CPT reading subject.

Table 8. Regularly Speaking or Hearing Chinese at Home (*** $p < .001$)

Variables		Number	Mean	S.D	t value
Listening	Yes	281	0.233	1.0479	4.859***
	No	481	-0.1361	0.9469	
Reading	Yes	275	-0.0022	0.9911	-0.046
	No	476	0.0013	1.0072	

C. Grade

The examinees' proficiency comparison on the basis of their grade differences that might affect CSL learning effectiveness was shown in below table 9. The average CSL proficiency was -0.1724, -0.3547, 0.2006, and 0.3580 and the standard deviation was 0.8028, 0.9226, 1.0125, and 1.0321 for grade 7, 8, 9, and 10 respectively on listening subject. In addition, according one-way ANOVA analysis as shown in table 10 below, the F value was 23.322. This indicated that there was significant different in CSL proficiency between grades on listening subject. Furthermore, according to the Scheffe's Method, the CSL CPT Proficiency of Grade 10 and 9 were both significantly higher than Grade 8 and 7. According to the comparison and one-way ANOVA analysis as shown in table 9, and 10 below on reading subject, the F value was 13.451. This indicated that there was significant different in CSL proficiency between grades on reading subject. The Scheffe's Method analysis also indicated that the CSL CPT Proficiency of Grade 10, 9, and 7 were all significantly higher than Grade 8.

Table 9. Examinees' Ability Comparison on Grade Difference

Variables		Number	Mean	S.D
Listening	Grade 7	116	-0.1724	0.8028
	Grade 8	243	-0.3547	0.9226
	Grade 9	242	0.2006	1.0125
	Grade 10	161	0.3580	1.0321
	Total	762	0.0000	1.0007
Reading	Grade 7	77	-0.0075	0.8735
	Grade 8	179	-0.3866	0.9041
	Grade 9	251	0.0827	1.0357
	Grade 10	244	0.2023	0.9934
	Total	751	0.0005	1.0004

Table 10. One-Way ANOVA and Scheffe's Analysis on Grade (*** $p < .001$)

Variables		SSE	d. f	MSE	F test	Scheffe's Method
Listening	between	64.391	3	21.464	23.322***	Grade 10 > 7, 8
	within	697.609	758	0.92		Grade 9 > 7, 8
	total	762	761			
Reading	between	38.461	3	12.82	13.451***	Grade 10 > 8
	within	709.123	744	0.953		Grade 9 > 8
	total	747.584	747			Grade 7> 8

D. Traveling Experience

The examinees' proficiency comparison on the basis of their traveling experiences over one month in either to Taiwan, Mainland China, or Hong Kong was shown in below table 11. The average CSL proficiency was -0.1815, 0.7458, 0.2592, and 0.5242 and the standard deviation was 0.9133, 0.9472, 1.2521, and 1.0616 for no traveling experiences to Mainland China, Hong Kong or Taiwan, and with travel experiences to Mainland China, Hong Kong or Taiwan on listening subject respectively. In addition, according to one-way ANOVA analysis as shown in table 12 below, the F value was 31.748. This indicated that there was significant different in CSL proficiency on listening subject if the examinees had traveling experiences oversea for more than one month. Furthermore, according to the Scheffe's Method, the CSL CPT Proficiency of examinees who had traveling experience to Taiwan or Hong Kong for more than one month were both significantly higher than examinees didn't. On the contrary, there were no significant different on reading subject.

Table 11. Examinees' Ability Comparison on Traveling Experience

Variables		Number	Mean	S.D
Listening	No (Experience)	581	-0.1815	0.9133
	Mainland (China)	74	0.7458	0.9472
	Hong Kong	22	0.2592	1.2521
	Taiwan	85	0.5242	1.0616
	Total	762	0.0000	1.0007
Reading	No (Experience)	543	0.0110	1.0009
	Mainland (China)	86	0.0745	1.0498
	Hong Kong	27	-0.4368	0.9153
	Taiwan	95	-0.0062	0.9614
	Total	751	0.0000	1.0007

Table 12. One-Way ANOVA and Scheffe's Analysis on Traveling Experience (*** $p < .001$)

Variables		SSE	d.f	MSE	F test	Scheffe's Method
Listening	between	85.144	3	28.381	31.784***	Taiwan > No
	within	676.856	758	0.893		Hong Kong > No
	total	762	761			
Reading	between	5.699	3	1.9	1.904	
	within	745.301	747	0.998		
	total	751	750			

E. Ethnicity

The examinees' proficiency comparison on the basis of their ethnicity different that might affect CSL learning effectiveness was shown in below table 13. The average CSL proficiency was -0.0611, 0.2106, and -0.1221 and the standard deviation was 0.9655, 1.0598, and 0.9790 for Filipino, Chinese, and others respectively on listening subject. In addition, according to one-way ANOVA analysis as shown in table 14 below, the F value was 6.365. This indicated that there was significant different in CSL proficiency between ethnicity on listening subject. Furthermore, according to the Scheffe's Method, Chinese examinees were significantly higher than Filipino and others on CSL proficiency. On the contrary, there were no significant different on reading subject.

Table 13. Examinees' Ability Comparison on Ethnicity

Variables		Number	Mean	S.D
Listening	Filipino	424	-0.0611	0.9655
	Chinese	202	0.2106	1.0598
	Others	136	-0.1221	0.9790
	Total	762	0.0000	1.0007
Reading	Filipino	422	-0.0537	0.9887
	Chinese	193	0.0896	1.0124
	Others	136	0.0395	1.0172
	Total	751	0.0000	1.0007

Table 14. One-Way ANOVA and Scheffe's Analysis on Ethnicity (** $p < .01$)

Variables		SSE	d.f	MSE	F test	Scheffe's Method
Listening	between	12.569	2	6.284	6.365**	Chinese > Filipino Chinese > others
	within	749.431	759	0.987		
	total	762	761			
Reading	between	2.978	2	1.489	1.489	
	within	748.022	748	1		
	total	751	750			

F. Parents' Education

The examinees' proficiency comparison based on the level different of their parents' education was shown in below table 15 to 18. For example, according to table 15 and 16, the average CSL proficiency was 0.4152, 0.0639, 0.0535, and -0.2777 and the standard deviation was 1.1062, 1.0916, 0.9775, and 0.9229 for father's education level below high school, high school, college, and above college respectively on listening subject. In addition, according to one-way ANOVA analysis as shown in table 16 below, the F value was 7.748. This indicated that there was significant different in CSL proficiency between father's education level on listening subject. Furthermore, according to the Scheffe's Method, the CSL CPT Proficiency of examinees whose father's education level were below high school or at college both significantly higher than examinees whose father's education were above college. The same phenomenon appeared on the different of mother's education level shown as table 17 and 18. However, there were no significant different on reading subject. On the other hand, the parents' education level different had nothing to do with the examinees' CSL proficiency on reading subject.

Table 15. Examinees' Ability Comparison on Father's Education Level

	Variables	Number	Mean	S.D
Listening	Below (High)	44	0.4152	1.10616
	High	92	0.0639	1.09162
	College	452	0.0535	0.97749
	Above (College)	174	-0.2777	0.9229
	Total	762	0.0000	1.00066
Reading	Below (High)	39	-0.1585	1.10115
	High	100	0.0854	1.11182
	College	475	0.0084	0.98451
	Above (College)	137	-0.0463	0.94355
	Total	751	0.0000	1.00067

Table 16. One-Way ANOVA and Scheffe's Analysis on Father's Education Level (*** $p < .001$)

Variables		SSE	d.f	MSE	F test	Scheffe's Method
Listening	between	22.672	3	7.557	7.748***	Below > Above College > Above
	within	739.328	758	0.975		
	total	762	761			
Reading	between	2.036	3	0.679	0.677	
	within	748.964	747	1.003		
	total	751	750			

Table 17. Examinees' Ability Comparison on Mother's education level

	Variables	Number	Mean	S.D
Listening	Below (High)	43	0.2714	1.1353
	High	47	0.1541	1.2314
	College	494	0.0582	0.9803
	Above (College)	178	-0.2679	0.907
	Total	762	0.0000	1.0007
Reading	Below (High)	35	-0.0460	1.2038
	High	53	-0.1386	1.0452
	College	521	0.0176	0.9872
	Above (College)	142	-0.0014	0.9854
	Total	751	0.0000	1.0007

Table 18. One-Way ANOVA and Scheffe's Analysis on Mother's Education Level (*** $p < .001$)

Variables		SSE	d.f	MSE	F test	Scheffe's Method
Listening	between	18.733	3	6.244	6.368***	Below > Above College > Above
	within	743.267	758	0.981		
	total	762	761			
Reading	between	1.254	3	0.418	0.416	
	within	749.746	747	1.004		
	total	751	750			

VI. Conclusions and Future Study Directions

A. Conclusions

The benefit of this study has enabled the constructed and established CSL proficiency scales comparable between any two different tests in the same proficiency level but different item difficulties. The contribution of this study was not only on the construction of a CSL Proficiency Test on a basis of CEFR but also in comparison with examinees' proficiency scales in referring to their background and explored factors that might affected CSL learning effectiveness.

This study was applied IRT 3PL model on B1 level reading and listening subject data analysis. The items were fit to the examinees' proficiency and could effectively discriminate examinees' proficiency. There were four categories of communication activities and strategies, Productive, Receptive, Interactive, and Mediating subscale scores, had estimated in this study. However, the examinees' CSL proficiency was below B1 level except Productive in Listening subject. This is probably the major reason caused no significant different on reading subject during further analysis on those six different background factors.

B. Future Study Directions

Language contains various concepts, customs of the history and culture extra during their communicating activities with others. Because of these cultural factors, languages can not be directly interpreted or transformed from words themselves. Therefore, the items of CSL CPT should integrate several dimensions. Unfortunately, this will definitely contaminate the IRT basic assumption of unidimensionality. Therefore, the future study might include the following directions:

1) Conduct this study and analyze samples collected from higher grades such as grade 11 or above.

2) Appling multidimensional IRT model for this study and analysis the culture factors that affected CSL proficiency.
3) Conduct CSL CPT on cultural subject.

REFERENCES

[1] G. Kecker and T. Eckes, "*Linking the TestDaF to the CEFR: The case of writing proficiency*," in *Fourth Annual Conference of EALTA*, 2007.
[2] R. J. Tannenbaum and E. C. Wylie, "*Mapping English proficiency test scores onto the Common European Framework*," TOFEL Research Rep. NO. RR-80, 2005.
[3] B. D. Wright and M. H. Stone, "Best Test Design," in *MESA Press* Chicago, 1979.
[4] B. D. Wright and G. N. Masters, "Rating scale analysis," in *MESA Press, Chicago, IL* Chicago, 1982.
[5] HSK, "HSK." vol. 2009, 2009.
[6] C. Test, "C. Test." vol. 2009, 2009.
[7] NCACLS, "*National Preparation Test for SAT Subject Test in Chinese with Listening*," 2009.
[8] F. M. Lord, "Applications of item response theory to practical testing problems," New Jersey: Lawrence Erlbaum Associates, 1980, p. 274.
[9] C. Council, "The Common European Framework in its political and educational context," CEFR Council, p. 260.
[10] Y.-H. Tsai, *CSL Teaching Materials Construction*. Taipei: Zhong Zheng, 2009.
[11] S. E. Embretson and S. P. Reise, *Item response theory for psychologists*. Mahwah, N.J.: L. Erlbaum Associates, 2000.
[12] W. H. Angoff, *Scales, Norms, and Equivalent Scores*. Princeton, New Jersey: Educational Testing Service, 1984.
[13] T. G. Bond and C. M. Fox, *Applying the Rasch Model: Fundamental measurement in the human sciences*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum, 2007.
[14] G. Rasch, *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmarks Paedagogiske Institut, 1960.
[15] F. M. Lord, "A theory of test scores," *Psychometric Monograph*, vol. 7, 1952.
[16] R. K. Hambleton and H. Swaminathan, *Item response theory : Principles and applications*. Hingham, MA: Kluwer-Nijhoff, 1985.
[17] F. B. Baker, *Item Response Theory: Parameter Estimation Techniques*. New Yook: Marcel Dekker., 1992.
[18] R. J. Mislevy and R. D. Bock, "*PC-BILOG-Item analysis and test scoring with binary logistic models*," Mooresville, IN: Scientific Software., 1990.
[19] A. Birnbaum, "*Some latent trait models and their use in inferring an examinee's ability*," in *Statistical theories of mental test scores*, F. M. L. M. R. Novick, Ed. Reading, MA: Addison-Wesley, 1968.
[20] R. M. Smith, "Comparision of the power of Rasch total-and between-item fit statistics to detect measurement disturbances," *Educational and Psychological Measurement*, vol. 54, pp. 42-55, 1994.
[21] M. F. Zimowski, E. Muraki, R. J. Mislevy, and R. D. Bock, "*BILOG-MG*," *Scientific Software lnternational*, 2003.

Author

Rih-Chang Chao was born in Kaohsiung, Taiwan. Day of birth is on April 7th, 1963. After graduated from the Fu-Jin Catholic University, Taipei, Taiwan acquired Mathematics major of Bachelor Degree. He was accepted to enroll Northern Illinois University to take the course in the Department of Applied Probability and Statistics with full Scholarship. He was also the teaching assistant in the department for 3 years. In 1992 he earned his master degree from Department of Applied Probability and Statistics at Northern Illinois University. He is currently a PH. D candidate of Graduate Institute of Educational Measurement and Statistics (GIEMS) at National Taichung University, Taichung, Taiwan.

He was working for GFK Asia based in Taipei, Taiwan as a FIELD MANAGER, responsible for market research, data collection and data analysis for three years. He was in charge of managing a group of staffs to complete several internal system projects. With the competitive market environment he had an opportunity to join AC Neilson Consultant Company base in Taipei, Taiwan, as a SENIOR MANAGER in the Department of Client Solutions. He was responsible of managing large account to handle and resolve any market information on particular products and services. He also handles retail and customized marketing research project. He has to manage group of 7 people in taking research project requests from all over the Asia. AC Neilson Consultant Company is currently ranked as the World NO. 1 and GFK ranked NO. 5. In 2008, he has started teaching in National Taiwan Normal University, Taipei, Taiwan and become a lecturer of Mathematics and Science department. He has other teaching experience before becoming permanent lecturer; He was teaching in Hsin Sheng College of Medical Care and Management, and Overseas Chinese University as a Lecturer. Teaching course involves of Calculus and CSL Assessment. He has published a book called CBEST Mathematics for student preparing to take the CBEST Examination in United State of America. This book was published by Wen-He Taipei Taiwan 2007. Major focus on his research is on the language testing. His research interests are specialized in the areas of proficiency measurement and assessment for learning Chinese as Second Language (CSL).

Mr. Rih-Chang Chao, he is currently a member of Chinese Association of Psychological Testing and Chinese Grey System Association.

Co-authors

Bor-Chen Kuo received the Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, in 2001. He is currently a Professor with the Graduate Institute of Educational Measurement and Statistics (GIEMS), National Taichung University, Taichung. His research interests are pattern recognition, remote sensing, image processing, and nonparametric functional estimation.

Hsuan-Po Wang was born in Nantou, Taiwan. His day of birth is on July 1st, 1982. He earned his master degree from Graduate Institute of Educational Measurement and Statistics (GIEMS) at National Taichung University, Taichung, Taiwan.. He is currently a PH. D candidate of GIEMS at National Taichung University, Taichung, Taiwan. His research interests are specialized in the areas of proficiency measurement and assessment for learning Chinese as Second Language (CSL).

Ya-Hsun Tsai is a Professor and Department Head of Applied Chinese Language and Literature at National Taiwan Normal University, Taiwan. She has been in this position since 2008. Before joined NTNU, she was the Department Head of Applied Linguistics and Language Studies at Chung Yuan Christian University.

Professor Tsai has published many articles, journal and books in the areas of Teaching Chinese as Second Language (CSL). She is the author of 『Teaching Material Construction for Learning Chinese as Second Language』 and co-authored a book on the 『Introduction of Teaching Chinese as Second Language』. She also had been officially nominated as one of the committees by both Ministry of Education and Overseas Compatriot Affairs Commission and participated with several CSL related projects. Her current research interests are also including CSL Testing and Assessment, CSL Teaching Syllabus Design, and Standard Setting for CSL vocabulary.