

Exploring the English Achievement Assessment of Junior High School Students

Pei-Chuan Cheng¹, Wen-Kuei Hsieh^{2*}

¹Hungkuo Delin University of Technology
Peichuan[at]mail[dot]hdut[dot]edu[dot]tw

²Hungkuo Delin University of Technology
wkhsieh[at]mail[dot]hdut[dot]edu[dot]tw

*corresponding Author

Abstract- In the past years, most traditional item analysis only analyses difficulty and discrimination of each item, and test analysis only analyses overall test reliability and validity. As a result, EFL educators are lack of information on students' response data for both students' learning styles and item types in test questions preparation. Thus, the study presents the various item types of the English achievement assessment of Junior High School Students in Taiwan, and illustrates the various learning styles of the EFL students. The participants were randomly selected from one thousand four hundred and forty two junior high school students, who participated in Taiwan Assessment of Student Achievement in Junior High School English (TASA) held by National Academy for Educational Research (NAER). The data was analyzed based on the dichotomous scoring and the Student-Problem Chart Analysis. The result of Caution Index for Students shows that high achievement students account for one third of the sample students. However, the other students were classified as learning abnormality, inattention learning, and lack of learning adequateness, insufficiency learning, and lack of academic ability. Also, the result of caution index for problems shows that the test items of English Achievement Assessment were capable of measuring the English achievements of junior high school student and it also can differentiate high achievers from the low achiever in Taiwan. To improve the further test question preparation, only partial revisions are suggested for National Academy for Educational Research (NAER).

General Terms- English achievement test; Learning style

Keywords- Student-Problem Chart Analysis ; Dichotomous scorings

1. INTRODUCTION

The educational process consists of instruction objectives, learning process, and assessment [18]. The purposes of assessment include: (a) student problem diagnosis, (b) student performance analysis, (c) providing feedback to students, (d) student placement, and (e) adjustment to teaching materials and methods [1] [5]. The key components of the instructional activities should include determining the teaching objectives, assessing the starting behaviors, implementation of teaching activities, and assessment. In this viewpoint, the assessment is a feedback relay station for student counseling and remedial instruction [2] [6] [7].

In Taiwan, English is regarded as compulsory subject for junior high school students. Students are required to take English classes for three to four hours each week in junior high school. Also, students are required to participant in Taiwan Assessment of Student Achievement in Junior High School English held by National Academy for Educational Research (NAER) each year. In the past years, most traditional item analysis only analyzes the difficulty and discrimination of each item, and test analysis only analyzes the overall test reliability and validity. As a result,

educators are lack of information on students' response data for both students' learning styles and item types in test questions preparation [4] [20].

The purpose of this study aims to explore the various item types of the English Achievement Assessment of Junior High School students in Taiwan, and to describe the various learning styles of the junior high school students. The dichotomous scoring and the Student-Problem Chart Analysis were used to analyze the collected data [10] [15]. The results of this study can provide some suggestions for high school EFL teachers.

2. LITERATURE REVIEW

2.1 Dichotomous Scoring

For many decades, dichotomous scoring has been widely utilized in scoring in order to save time and get consistent results [9]. In addition, among the item response models of dichotomous scoring, one, two, and three parameter logistic models (1PL, 2PL, and 3PL) are the most extensively used item response theory [8] [11] [14] [19].

2.2 Students-Problem Chart Analysis

To analyze students' learning style and to analyze item types, the student and problem chart analysis (Student-

Problem Chart Analysis, SPCA) proposed by Sato(1980) is utilized in this study. Based on SPCA, 'problem' represents test item. In addition, the SPCA (Student-Problem Chart Analysis) include caution index for students (CS) and caution index for problems or item (CP). They are calculated by following equations respectively [17] [16] [20].

1. Caution Index for Students

$$CS_i = 1 - \frac{\sum_{j=1}^n (y_{ij}) (y_j) - (y_i)(\mu')}{\sum_{j=1}^n y_j - (y_i)(\mu')} \quad \text{-----2-1}$$

2. Caution Index for Problems

$$CP_j = 1 - \frac{\sum_{i=1}^n (y_{ij}) (y_i) - (y_j)(\mu)}{\sum_{i=1}^n y_i - (y_j)(\mu)} \quad \text{-----2-2}$$

In which, y_i is total score of student i ; μ is average score; y_j denotes number of correct answers of item j ; μ' is average number of correct answers.

3. METHODOLOGY

3.1 Participants

The total participants for this study were one thousand four hundred and forty two junior high school students and they were randomly selected from eight thousand five hundred and eighty seven junior high school students, who participated in 2013 Taiwan Assessment of Student Achievement in Junior High School (TASA, 2013) held by National Academy for Educational Research (NAER, 2013).

3.2 Procedures for implementation

The procedures for this study are described as following. First, the ability index and two-way schedules of junior high school English were collected. Second, prepare test questions and complete formal test by Testing and Assessment Center of National Academy for Educational Research (NAER). Third, analyze the data based on dichotomous scoring and the Student-Problem Chart. Then, calculate caution index for students (CS) and caution index for problems (CP), and perform student diagnosis and problem diagnosis.

3.3 Instrument

The instrument for this data analysis is the educational testing and assessment analysis software Tester for Windows version 3.0. It was developed by professor Min-Ning Yu of National Cheng-Chi University [20]. The Tester for Windows version 3.0 was utilized in this study because it can provides overall analysis including item analysis, test analysis, S-P Chart analysis etc. Item analysis provides information of each item, including difficulty, discrimination, disparity index, and item option distraction. Test analysis comprises information of overall items, which includes reliability, validity, and disparity index. Also, S-P Chart analysis offers information of various students' learning styles and item types. The arithmetic logics of Tester for Windows version 3.0 are described in the following. First, convert the raw data matrix into dichotomously scoring matrix data. Then, sort students'

scores and the number of correct answers. Next, calculate the correct response rates for both higher and lower groups, and item characteristics for each item. Finally, calculate the caution index for students and caution index for problems.

4. RESULTS

4.1 Student Diagnosis

According to caution index for students, students' learning styles were classified into six categories, including learning stability (A), inattention learning (A'), insufficiency learning (B), lack of learning adequateness (B'), lack of academic ability (C), and learning abnormal ability (C'). The results of caution index for students were described in the following.

1. Learning stability A: It means high stability for high academic achievement. A total of 468 students were calculated and it accounted for 32.45 percent of the sample students.
2. Inattention learning A': It means learning with moderate stability and good ability. A total of 286 students were calculated and it accounted for 19.83 percent of the sample students.
3. Insufficiency learning B: It shows students require further efforts. A total of 117 students were calculated and it accounted for 8.11 percent of the sample students.
4. Lack of learning adequateness B': It means learning is occasionally careless and has inadequate preparation. A total of 150 students were calculated and it accounted for 10.40 percent of the sample students.
5. Lack of academic ability C: It means lack of basic skills and low academic achievement. A total of 78 students and it accounted for 5.40 percent of the sample students.
6. Learning abnormality C': It means that learning extremely is unstable for the students and they are with casual reading habits. A total of 343 students were calculated and it accounted for 23.78 percent of the sample students.

Table 1. Student Diagnosis

Category	Number of participants	%age
A	468	32.45
A'	286	19.83
B	117	8.11
B'	150	10.40
C	78	5.40
C'	343	23.78

The Caution index for students is summarized in Table 1. The result shows those high achievement students

accounted for 32.45 percent of the sample students. However, students whose learning is extremely abnormal accounted for 23.78 percent of the sample students. Also, the lower achievers accounted for 19.83 percent of the sample students who might need more attention in learning and remedial teaching from English teachers.

4.2 Item Diagnosis

Item types were classified into four categories, according to caution index for problems and percentage of correct answers of each item, namely, appropriate item (A), heterogeneous item (A'), difficult item (B), and awkward item (B'). The result of Caution index for problems is described in the following.

1. Appropriate item (A): Items are properly prepared, and are capable of differentiating low achievers from other students. A total of 31 questions out of 32 were calculated and it accounted for 96.87 percent.
2. Heterogeneous item (A'): It requires partial amendments or items with poor options. A total of 0 questions out of 32 were calculated and it accounted for 0 percent.
3. Difficult item (B): It was with high difficulty, and was capable of differentiating high achievers from other students. A total of 1 question out of 32 was calculated and it accounted for 3.13 percent.
4. Awkward item (B'): It means unclear meaning and requires item deletion or amendments. A total of 0 questions out of 32 were calculated and it accounted for 0 percent.

Table 2 presents the result of Caution Index for problems. As seen in Table 2, the appropriate item (A) and hetero generous item (A') accounts for 96.87 percent of the test items. That is, it is capable of both measuring the English Achievement of the junior high school students. In addition, the difficult item (B) and awkward items (B') accounted for 3.13 percent of the test items.

Table 2 Item Diagnosis

Category	Number of items	% age
A	31	96.87
A'	0	0
B	1	3.13
B'	0	0

4.3 Item Analysis

Table 3 presents test items, difficult item (B), under correct response rate less than 0.5 according to correct response rate items. The table includes item number, number of correct answers, correct response rate, difficulty, discrimination, caution index, and item types.

Table 3. Difficult items

Item No/	No. of correct	Correct response	Difficulty	Discrimination	Caution index	Item types
----------	----------------	------------------	------------	----------------	---------------	------------

option	answer	rate				
19	600	0.42*	0.54	0.64	0.48	B

Note: The correct response rate is less than 0.5.

Item 19 was classified as difficult item (B) with correct response rate of 0.42 and it required partial amendments. Also, under the condition of discrimination greater than 0.50, item 19 was capable of differentiating high achievers from the low achievers with discrimination of 0.64. Table 4 describes item response cases analysis.

Table 4. Item Response Cases Analysis

Item no/ Options	1	2	3	4	Other
19	0.14	0.27	0.17	0.42*	0

Note: The caution indices is greater than 0.5)

Table 5 describes distraction analysis to difficult items (B). With respect to item 19, the correct answer was option four; number of correct answers was 600 with option rate of 0.42; the numbers of response for each option were 204, 396, 241, and 600 with option rates of 0.14, 0.27, 0.17, and 0.42 respectively. Moreover, option two and four were distractive, but option one and three were lack of distractive and required partial revise because only 9 responses case each in high achievers.

Table 5 Distraction Analysis to Item 19

Options	1	2	3	4*	Other
Response/ Option rate	204/ 0.14	396/ 0.27	241/ 0.17	600/ 0.42	0
High achievers/ Option rate	6/ 0.02*	38/ 0.11	6/ 0.02*	310/ 0.86	0
Low achievers/ Option rate	80/ 0.22	111/ 0.31	90/ 0.25	79/ 0.22	0

5. DISCUSSIONS

5.1 Students' learning styles

According to the results of this study, around one third of the participants (32.45 percent) show high learning stability, while two third of the participants might need more help or remedial instruction from English teachers in high school. As proposed by Norris (2000), "the ultimate goal of language assessment is to use tests to better inform us on the decisions we make and the actions we take in language education" (p. 18). Thus, to improve junior high school students' English ability in Taiwan, high school English teachers should take students' learning styles into consideration when they design instructional activities. For low achievement students who were with learning abnormality, English teachers could cultivate their reading habits in English classes. In addition, to improve students' learning, high school teachers or junior college teachers also need to give more attention to low achievement students and provide more remedial instruction for the

students who have insufficiency learning and are lack of academic ability.

5.2 Item Preparation

5.2.1 Item Diagnosis

Based on the results of this study, firstly, appropriate items (A) and heterogeneous items (A') accounted for 96.87 percent of the test items. It reveals that they were capable of both measuring the English achievements of junior high school student. Also, it can differentiate high achievers from the low achievers. Secondly, difficult items (B) and awkward items (B') accounted for 3.13 percent of the test items which required of item analysis.

5.2.2 Item Analysis

1. Difficult Items

Item 19 was classified as difficult item (B) with correct response rate of 0.42 and required partial amendments; under the condition of discrimination greater than 0.50, item 19 was capable of differentiating high achievers from the low achievers with discrimination of 0.64.

2. Distraction Analysis to Difficult Items

With respect to item 19, the correct answer was option four; number of correct answers was 600 with option rate of 0.42; the numbers of response for each option were 204, 396, 241, and 600 with option rates of 0.14, 0.27, 0.17, and 0.42 respectively. Moreover, option two and four were distractive, but option one and three were lack of distractive and required partial revise because only 9 responses case each in high achievers.

6. CONCLUSION

The results of this study demonstrate that around one third of the participants (32.45 percent) show high learning stability, while two third of the participants might need more help or remedial instruction from English teachers in high school. Also, students' learning styles should be taken into consideration when high school English teachers design instructional activities in class. Furthermore, for students who are with learning abnormality, English teachers could cultivate their reading habits in English classes. Furthermore, to improve low achievement students' insufficient learning and academic ability, follow-up remedial instructions are needed in senior high school or junior college in Taiwan.

As for the result of item type's analysis, almost 96.87 percent of the test items were capable of measuring the English achievements of junior high school student. Also, it can differentiate high achievers from the low achiever. However, for further test question preparation only partial revision will be needed for National Academy for Educational Research (NAER).

REFERENCES

- [1] Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment: Concepts and applications* (6th ed.). New York: McGraw-Hill.

- [2] Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519-521.
- [3] Guttman L. (1944). A basic for scaling qualitative. *American sociological review*, 9, 139-150.
- [4] Haladyna, T. M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- [5] Hart, D. (1994). *Authentic assessment: A handbook for educators*. New York: Addison-Wesley.
- [6] Huang, K. S. (1988). *Teaching principles*. Taipei: ShTa Bookstore.
- [7] Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Upper Saddle River, NJ: Prentice-Hall.
- [8] Liu, C. C. & Gen, M. F. (1993). Some fast estimation formulas of reliability for binary items test. *Psychological Testing*, 40, 169-188.
- [9] Lord, F. M. (1952). *A theory of test scores*. Psychometric Monograph, 7.
- [10] Lord, F. M. (1980). *Application of item response theory to practical testing problems*. Hill-Isdale, NJ: Lawrence Erlbaum Associates.
- [11] Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- [12] National Academy for Educational Research (NAER) (2011). *Taiwan Assessment of Student Achievement in Vocational High School English* (2011 TASA). Taipei: National Academy for Educational Research of Taiwan, R.O.C.
- [13] Norris, J. M. (2000). Purposeful language assessment: Selecting the right alternative test. *Forum*, 38(1), 18-23.
- [14] Ou, C. H. (2002). *Educational Testing and Assessment*. Taipei: Psychological Publishing.
- [15] Sato, T. (1980). The S-P chart and the caution index. *NEC Educational Information Bulletin*, 80(1), C&C Systems Research Laboratories, Nippon Electric Co. Ltd., Tokyo, Japan (In Japanese).
- [16] Sato, T. (1985). *Introduction to student-problem curve theory analysis and evaluation*. Tokyo: Meiji Tosho (In Japanese).
- [17] Sato, T., & Kurata, M. (1977). Basic S-P score table characteristics. *NEC Research and Development*, 47, 64-71.
- [18] Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation* (7th ed.). Englewood Cliffs, NJ: Prentice Hall.
- [19] Yu, M. N. (2009). *Item Response Theory and Its Application*. Taipei: Psychological Publishing.
- [20] Yu, M. N. (2012). *Educational Testing and Assessment: achievement testing and instruction*

assessment (3rd ed.). Taipei: Psychological Publishing.

Author's Biography



Language Assessment.

Pei-Chuan Cheng is an Assistant professor in Department of Applied English in Hungkuo Delin University of Technology. Her research interests are writing Instruction, Computer-Assisted Language Learning and



Wen-Kuei Hsieh is an Associate Professor in Applied Foreign Languages Center in Hungkuo Delin University of Technology. His research interests are Language Teaching and Testing, and Language Assessment.