Curriculum and Teaching Methodology (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/curtm.2023.061102
ISSN 2616-2261 Vol. 6 Num. 11

Study on Students' Learning Ability from the Design and Analysis of Primary English Test Paper

Qishan Liao[1,a,*]

[1]*Faculty of Education and Human Development, The Education University of Hong Kong, Hong Kong SAR, 999077, China*
[a]*liaoqishan2023@163.com*
**Corresponding author*

Keywords: Test design and analysis, bloom taxonomy, rasch analysis, learning ability

Abstract: This paper mainly focuses on the design and analysis of English test papers which can better understand primary students' level of learning ability and test quality to promote students' English learning and teachers' teaching quality. TOS (Table of Specifications) and the theory of Bloom's Taxonomy of Educational Objectives are the guidance to the design of the test paper. For the test analysis part, Winsteps Rasch Model is applied to analyze the reliability of items and students, difficulty factor, and level of discrimination. As a result, it can be employed to precisely identify the questions and students based on their level of achievement, revealing the actual level of the learners' ability regardless of the small number of samples.

1. Introduction

Assessment should be considered a coherent and important part of teaching and learning. The goal of evaluation design is to provide teachers and students with advice on how to improve learning by explaining a student's learning process, assessing each student's progress in their work, clarifying certain problems the student may encounter in the learning process, and more [3]. At the same time, test making and test result analysis are also important ways to analyze students' learning and teachers' teaching quality. In this report, taking "In China", unit 1, Volume 1, Grade 6, people's Education Edition as an example, this report first designs the table of specifications (TOS) according to the textbook content combined with Bloom's Taxonomy educational objectives. The test paper is then created to evaluate students' ability to remember, understand, and a little bit of application of the knowledge they have learned. I gathered the students' answers, and then I ran the data through SPSS and Winsteps for analysis. The test's difficulty and discrimination are assessed using SPSS, and Winsteps can discover a deeper connection between the test and students' performance.

2. Content to be Tested

2.1. Application of Bloom's Taxonomy of Educational Objectives

Bloom's taxonomy is a system for categorizing and identifying the many levels of a student's cognitive capabilities, such as thinking, learning, and comprehending. To direct or serve as guidance for the development and improvement of evaluations like examinations and other assessments of student learning, educators have frequently employed Bloom's taxonomy [2]. It concludes with six categories: knowledge, comprehension, application, analysis, synthesis, and evaluation [7]. The student's ability for retaining information is made up of knowledge. And then a comprehension exam gauges the student's ability to understand the importance of prior knowledge. The following stage is the application, which shows how well the student can adapt abstract knowledge to a fresh situation. The analysis aims to differentiate between facts and opinions [4]. The synthesis category exemplifies the ability to put together many elements or thoughts into a strong pattern or structure that helps provide new meaning. An assessment category demonstrates the ideas' capacity for significance-based evaluation [10].

Based on the tested objectives are primary students in Grade 6, so the application of Bloom's taxonomy of educational objectives theories is mainly in three aspects: 1. Knowledge: The capacity to recall information acquired earlier. It might require recalling relevant detail points or bringing them to mind. The lowest level of learning is remembering. 2. Comprehension: The ability to understand a set of content. Understanding can be demonstrated by converting information into another format, extrapolating information, or developing views and ideas. 3. Application: The capacity to apply knowledge in a fresh, practical setting. These kinds of processes used in the application are described by the implementation of principles, methodologies, ideas, principles, rules, and theories.

2.2. Background of Subject and Assessment Object

Large-scale standardized test designers who specialize in educational assessment have recommended using a TOS as a guide to address validity data based on test content. This chart will assist teachers in planning exams by their educational objectives for a specific learning area [9]. When designing the tested content, based on the theme, time-consuming in each part, and teaching activities on TOS, the test content is created. The content I would be assessing them on is mainly in three aspects: New vocabulary including location expressions, famous scenic spots and culture expressions, Sentence structure (What..., Where..., Which...), and Grammar (past and present).

2.3. Design of TOS

Vocabulary learning is the basic part of English learning when a new unit begins. According to Bloom's theory, Knowledge is the lowest-level learning. Remembering is the foundation of knowledge cognitive [5]. To let students remember the words better, the first activity on the TOS (see Figure 1) is to show the pictures of famous cities in China and the maps of China and then to guide students read the new words combined with pictures, after that, let them apply the new words to match the location and the cities and customs. This activity was intended to help children recognize the names of various cities and customs in China and to enable them to apply the knowledge accordingly to the location model (south, north, west, and east). Understanding and application of sentence structure of "What...", "Where..." and "Which..." (see Figure 1) is another objective that students must have to achieve. Some students may easy to remember the sentence structure and the example sentences, but it may be hard for them to apply these sentence structures

in different situations. To help them deeply understand and apply the sentence structures, the activity of making conversations in groups is necessary [6] (see Figure 1). On one hand, students can help each other to finish the tasks that can let those students who fall behind have a chance to solve their problems about difficult points. On the other hand, through this activity processing, students will have a better understanding and application of how to ask and answer the question in "What...", "Where..." and "Which..." sentence structures. Furthermore, the Grammar part of the test is about the past tense and the present tense and the difficult point is distinct between the past tense and the present tense. So activities in class should be held based on that purpose. Firstly, ask two students to share the vacation on last summer holiday and next summer holiday respectively, then group students into four groups, each group should choose one representative to do the summary about the content that the two students shared. After that, the teacher advises on their answers. In this way, students can not only comprehend the grammar points deeply but also each student can have a process of thinking and apply what they've learned to finish the task [8]. The final activity of the class is to ask students to finish the task "Let's read" on page 11 of the textbook independently and then the teacher gives the right answers. In this part, students can think independently to find out whether they understand the points of the class.

3. Test Making Process

Teachers can determine the kinds of assignments include in their examinations by using TOS to help them connect the instructional time allocated to each goal with the level of cognition at which each goal was taught. Experts in measuring recommend a variety of methods for creating and using a TOS [8] (e.g., Anderson, Krathwohl, Pintrich, Raths, & Wittrock, 2001, Gronlund, Livingston, & Wilson, 2006). This paper contains five types of questions: fill-in-the-blank, multiple choice, reading comprehension, sentence rewrite, and short essay writing. The exam lasts 30 minutes and has a maximum score of 100. The score and scoring standard with each question are marked on the test paper (see Figure 1). The goal of the chapter's exam is to assess how well the students have mastered the material from Unit 1. The student's memory, comprehension, and application skills are explored by the requirements of the teaching objectives. Because the subjects are students at the beginning of learning English, they are not required to reach the ability level of analysis, evaluation, and creation. The order of the test questions ranges from simple to complex. Fill-in-the-blank questions are the foundation of test paper design. Students can enter the exam state more quickly with the aid of simple question formats. Additionally, the topic is properly enhanced with tables and images, which not only assists students in comprehending the context of the topic but also raises their level of interest in the exam paper.

There are thirteen fill-in blanks (Q1-Q5, Q12-Q15) and six multiple-choice questions (Q6-Q11), each with 2 points (Q1-Q11) and 3 points (Q12-Q15). Memory and comprehension are the ability levels that should be investigated. These questions are combined with real-life images to assess student's ability to connect real life and knowledge. Q1-Q5 and Q10-Q11 assess students' memory of the new vocabulary from Unit 1, while Q6-Q9 assess their comprehension of sentence structure (what, where, and which). The similar shape and style of words can test students' ability to distinguish the usage of these words. Understanding of Past tense and Present tense can be tested by Q12-Q15. It can detect whether students have mastered the knowledge and prevent students from doing the right thing by guessing. The design of questions in Q12-Q15 also assesses students' memory and understanding of knowledge. A task Q&A is developed for the reading comprehension questions (Q16–Q19) of the test paper to measure how students can construct complete sentences and comprehend the information being read. The article doesn't show students the same words of stems directly which requires students to have an understanding ability to turn articles into answers.

Additionally, sentence rewrite questions (Q20–Q23) ask students to rewrite sentences from statement sentences to special question sentences to assess their application skills. In Q24, it checks students' ability level to the application of the grammar of past tense and the present tense to create an essay.

Table of specifications of Unit1 In China

Theme	Specific learning content -	Learning activity (Teaching plan)	Time (spent in learning activity)	Objectives achieved in learning activity (Highlight learning difficulties & high order thinking)	Objectives to be tested (Item type)						Weight (Score)
					Remember	Understand	Apply	Analyze	Evaluate	Create	
Cities in China	New Vocabulary Learning: 1.famous scenic spots traditional culture in China 2.position: south, north, west, east	1.Use PPT to show students Chinese cities and their characteristics, and introduce new words about famous scenic spots in China. 2.Lead students to read new words and list them on the board. 3.Show the map of China with location and ask the question: Where is.... (a city on the map)? And let the students think. Then, according to the answers of the students, lead to the location words. 4.List the words on the board.	5mins	Remember new words:Guangzhou-Morning tea, soup; Suzhou- garden; Lhasa-the Potala Palace; Beijing-the Summer Palace; Xi'an-the Terracotta Army, south, west, east, north	5 fill-in-blanks and 2 multiple choices						14%
	Sentences Learning: Where did you go on your summer vacation? I went to ..... What's famous for? It's famous for... Which city do you want to visit in China? What do you want to do in...? I want to...	1.Based on the first step of city introduction, ask students questions and introduce new sentence patterns: "Where did you go on your summer vacation? I went to Guangzhou with my parents.What did you do there? We ate seafood, had morning tea and tasted soup." 2.Ask students to make a conversations by using the new sentence structure and ask two groups of students to show the conversation. 3. Give feedback about the conversation.	12mins	Understand the new sentence structures. Apply the new structure and past tense in different situation.		4 multiple choices	4 sentences rewrite				32%
	Grammar: Distinct the past voice and present voice	1. Ask two students to share the vacation on last summer holiday and next summer holiday respectively, then group students into four groups, each group should choose on representative to do the summary about the content that two students shared. 2. After that, teacher gives advice on their answer.	13mins	Understand and Apply the grammar point in different situations.	8 fill-in-blanks		1 writing				34%
	Reading comprehension	1. Ask students to finish the task 'Let's read' on page 11 of the textbook independently. 2. Teacher gives the right answers.	8mins			4 short questions					20%

Figure 1: Table of specification of Unit1 in China

4. Results of Statistical Analyses

```
ENG6
-------------------------------------------------------------------------------
| student      33 INPUT      33 MEASURED              INFIT        OUTFIT    |
|          TOTAL      COUNT      MEASURE   REALSE    IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN      80.4       24.0        30.73     1.55      .95    -.3    .86    -.3|
| P.SD      11.8         .0         2.37      .15      .50    1.8    .34    1.0|
| REAL RMSE   1.55 TRUE SD   1.79  SEPARATION  1.16  studen RELIABILITY   .57|
|-----------------------------------------------------------------------------|
| Item      24 INPUT      24 MEASURED              INFIT        OUTFIT    |
|          TOTAL      COUNT      MEASURE   REALSE    IMNSQ   ZSTD  OMNSQ   ZSTD|
| MEAN     110.5       33.0        50.00     1.38      .86    -.9    .86   -1.0|
| P.SD      53.4         .0        14.57      .27      .69    3.0    .65    2.7|
| REAL RMSE   1.41 TRUE SD  14.51  SEPARATION 10.32  Item   RELIABILITY   .99|
-------------------------------------------------------------------------------
```

Figure 2: Rasch analysis of student's ability

The collected data were processed with IBM SPSS Statistics and Winsteps Rasch Model to analyze mainly in three aspects: the reliability of items and students, difficulty factor, and level of

discrimination [1]. It is common to assess the exam items' reliability using Cronbach's alpha. In general, if the internal consistency is higher than 0.8, it is excellent, between 0.6 and 0.8, it is good, and lower than 0.6, it is poor. Cronbach's alpha ought to be greater than 0.5, ideally greater than 0.7, in practice. In Figure 2, the student's ability as measured by Rasch analysis is 0.53—nearly good—and the reliability as measured by item quality is 0.99—high.

For the level of difficulty, we can see in Figure 3, the difficulty factor can be calculated by SPSS, the higher the difficulty factor, the easier the item. The general difficulty factor should be between 0.3-0.7 if we want to know the performance of different students. In Figure 3, we can find that 20 items of the difficulty factor of the test are over 0.7, meaning that the item is easy, and 13 items of it are over 0.8 (which means very easy). The remaining four items of difficult factors are moderate.

DI can show whether the items are differentiated, which means that the students to be tested are divided into higher groups and lower groups and calculate the ability to discriminate (compare). From Figure 3, DI (compare) shows that 14 test items have a DI of less than 0.19, indicating that they have very little discrimination and cannot fully describe students' true ability levels. Since the DI of FB1 and FB2 is at 0, it can be assumed that students of any ability level can provide accurate responses. Additionally, 5 items have a DI greater than 0.4, demonstrating the high quality of the items and their ability to clearly separate students with high ability from those with low ability. The discrimination index (correlation) in Figure 3 indicates that with 14 items, the DI index is less than 0.19, implying that the consistency between students' scores on these items and their overall test scores cannot be reflected. With 8 items, however, there is consistency between students' scores on these items and their overall test scores. Rasch analysis of the item reveals that six low-quality items cannot maintain consistency between the item score and total score. This is shown in Figure 3. We can see that MNSQ greater than 1.5 for some items, such as R19, C20-23, indicates that when students answer these items, students with low ability correctly answer the items, while students with high ability incorrectly answer the items, indicating that discrimination ability is a little low.

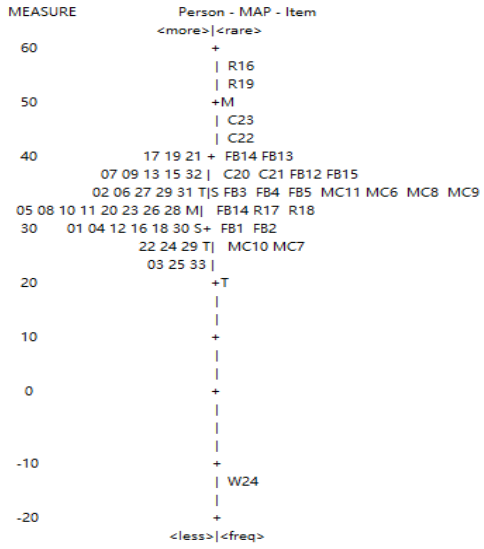
In Figure 4, the Student-Item Map (variable map made by Winsteps) shows that students No.17, No.19, and No.21 showed the best performance on this test, as shown in Figure 4, earning the highest score, No.7, No.9, No.13, No.15 and No.32 are behind them with the second-best ability. The five students with the lowest scores are No.3, No.25, and No.33. The ability level is comparable, and there is not much of a gap between the students overall.

	SPSS						Winsteps			
Item	Full score	Mean	MeanH-MeanL	Difficulty Factor	DI (Correlation)	DI (Compare)	Point measure correlation	Rasch measure	In-fit MNSQ	out-fit MNSQ
FB1	2.000	1.580	0.000	0.790	0.029	0.000	0.2758	59.59	0.498	0.5328
FB2	2.000	1.820	0.000	0.910	0.008	0.000	0.294	58.13	0.2796	0.2965
FB3	2.000	1.880	0.364	0.940	0.299	0.182	0.2992	57.79	0.16	0.1953
FB4	2.000	1.940	0.182	0.970	0.200	0.091	0.3047	57.47	0.1344	0.1543
FB5	2.000	1.940	-0.182	0.970	-0.130	-0.091	0.3047	57.47	0.1822	0.1799
MC6	2.000	1.940	0.182	0.970	0.185	0.091	0.3047	57.47	0.1362	0.1562
MC7	2.000	1.640	-0.182	0.820	-0.179	-0.091	0.2799	59.2	0.5156	0.5186
MC8	2.000	1.940	-0.182	0.970	-0.130	-0.091	0.3047	57.47	0.1822	0.1799
MC9	2.000	1.940	0.182	0.970	0.080	0.091	0.3047	57.47	0.1492	0.1677
MC10	2.000	1.820	0.000	0.910	-0.008	0.000	0.294	58.13	0.2835	0.2933
MC11	2.000	1.940	0.182	0.970	0.200	0.091	0.3047	57.47	0.1344	0.1543
FB12	6.000	5.270	1.019	0.878	0.311	0.170	0.2638	44.6	0.9861	0.9287
FB13	6.000	4.450	0.273	0.742	0.058	0.046	0.374	48.33	0.9262	1.0489
FB14	6.000	5.550	0.273	0.925	0.128	0.046	0.2016	42.39	1.2542	1.3074
FB15	6.000	5.450	-0.273	0.908	-0.112	-0.046	0.2245	43.25	1.2582	1.9874
R16	5.000	4.240	2.273	0.848	.486**	0.455	0.3896	49.07	0.8288	0.699
R17	5.000	3.480	3.636	0.696	.585**	0.727	0.4087	51.54	1.2225	1.2616
R18	5.000	3.940	1.818	0.788	.425*	0.364	0.4039	50.07	1.1148	1.0134
R19	5.000	2.580	4.545	0.516	.807**	0.909	0.3634	54.66	1.5738	1.7386
C20	6.000	4.000	3.818	0.667	.633**	0.636	0.4018	49.88	1.911	1.7578
C21	6.000	3.820	3.818	0.637	.627**	0.636	0.4071	50.47	1.9683	1.8675
C22	6.000	5.090	2.182	0.848	.386*	0.364	0.2965	45.64	2.1314	1.6258
C23	6.000	4.730	1.636	0.788	0.343	0.273	0.3468	47.29	2.2241	1.7976
W24	9.000	7.390	1.727	0.821	.831**	0.192	0.2144	-14.85	0.6892	0.6904

Figure 3: The relationship between items and student’s ability

For fitting tests, the Rasch model employs Outfit MNSQ and Infit MNSQ in Figure 4. As the picture seen in Figure 4, there are nine students—Students No. 1, 2, 12, 15, 16, 20, 28, 29, and

31—whose Outfit MNSQ is less than 0.5. This finding suggests that their test data are too perfect and the model is over-fitted, however, the issue is not severe because of the small population and the value's general proximity to 0.5. Additionally, some students—such as No. 3, 4, 7, 10, 21, and 23—have marginally higher MNSQ scores, suggesting that they may have engaged in questionable behavior like guessing, cheating, or being careless, among other things.



Rasch

id1	id2	Rasch measure	In-fit MNSQ	Out-fit MNSQ
0	1	30.09	0.3946	0.4497
0	2	35.49	0.2161	0.2551
0	3	24.65	1.5539	1.1646
0	4	30.09	1.1765	1.0059
0	5	32.27	0.5845	0.7125
0	6	35.83	1.4558	1.1336
0	7	38.25	1.6953	1.4531
0	8	32.25	1.0467	1.0786
0	9	38.65	1.102	0.9342
1	0	32.25	1.6795	1.2518
1	1	32.28	0.6476	1.2285
1	2	30.09	0.3274	0.3468
1	3	38.35	1.3369	0.992
1	4	28.05	1.4141	1.1517
1	5	38.68	0.2714	0.282
1	6	30.09	0.3063	0.3757
1	7	40.45	1.2895	0.8976
1	8	30.07	0.5759	0.5303
1	9	40.65	1.0955	0.8184
2	0	32.68	0.3295	0.8348
2	1	40.65	1.504	1.1549
2	2	26.27	1.1035	0.962
2	3	32.25	1.9016	1.4022
2	4	26.27	0.7006	0.5339
2	5	24.65	1.4178	1.1062
2	6	32.24	1.2486	0.9757
2	7	35.28	0.4277	0.8238
2	8	32.28	0.4181	0.8551
2	9	35.27	0.3779	0.3539
3	0	30.09	1.1545	0.9872
3	1	35.68	0.2714	0.282
3	2	38.65	1.095	0.8548
3	3	24.65	1.2641	1.0721

Figure 4: Student-Item Map and Outfit MNSQ and Infit MNSQ

5. Suggestions and Summary

After designing the test paper, peer review is concluded, recommendations like do not limit too much to the content of the textbook and do not go into too much detail to make a point, but should dig out the knowledge points behind the textbook, such as grammar of past tense and present tense. At the same time, the setting of question stems needs to be clearer and specific so I revised the question stem and added examples.

On the whole, it was discovered that the test was not too challenging for the students. Regardless of their level of ability, most students were able to correctly answer the majority of the questions. This is mainly due to the integration of examination and classroom based on the help of TOS. For example, the reasonable arrangement of the time of the classroom link corresponds to the score of the exam, and the theme of the classroom activities corresponds to the knowledge points of the exam. On the other hand, the theory of scientific measurement is used to sort through the test questions on the test paper, selecting the best and discarding the worst. After the test, Winsteps and SPSS are utilized for process analysis, and it is possible to get a preliminary evaluation of the testing results and papers. In future teaching career, scientific measurement and TOS should be included in evaluation and teaching. In this way, teachers can have a better understanding of students' learning ability and test quality and can promote students better.

References

[1] Arsad N., Kamal N., Ayob A., Sarbani N., Tsuey C. S., Misran N., & Husain H. (2013). Rasch model analysis on the effectiveness of early evaluation questions as a benchmark for new students ability. International Education Studies, 6(6), 185-190.

[2] Arievitch I. M. (2020). Reprint of: The vision of Developmental Teaching and Learning and Bloom's Taxonomy of educational objectives. Learning, culture and social interaction, 27, 100473.

[3] Berry R. (2008). Assessment for learning (Vol. 1). Hong Kong University Press.

[4] Betts S. C. (2008). Teaching and assessing basic concepts to advanced applications: Using Bloom's taxonomy to inform graduate course design. Academy of Educational Leadership Journal, 12(3), 99.
[5] Bloom B. S., Krathwohl D. R., & Masia B. B. (1984). Bloom taxonomy of educational objectives. In Allyn and Bacon. London: Pearson Education.
[6] Butler Y. G. (2011). The implementation of communicative and task-based language teaching in the Asia-Pacific region. Annual review of applied linguistics, 31, 36-57
[7] Conklin J. (2005). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives complete edition.
[8] Di Donato-Barnes N., Fives H., & Krause E. S. (2014). Using a Table of Specifications to improve teacher-constructed traditional tests: an experimental design. Assessment in Education: Principles, Policy & Practice, 21(1), 90-108.
[9] Fives H., & Di Donato-Barnes N. (2013). Classroom test construction: The power of a table of specifications. Practical Assessment, Research, and Evaluation, 18(1), 3.
[10] Jain M., Beniwal R., Ghosh A., Grover T., & Tyagi U. (2019). Classifying question papers with bloom's taxonomy using machine learning techniques. In Advances in Computing and Data Sciences: Third International Conference, ICACDS 2019, Ghaziabad, India, April 12–13, 2019, Revised Selected Papers, Part II 3 (pp. 399-408). Springer Singapore.