An Measurement Method of Ancient Poetry Difficulty for Adaptive Testing

Huiping Wang
Institute of Chinese Information Processing,
Beijing Normal University
Beijing, China
18013752306@163.com

Bihua Wang
Institute of Chinese Information Processing,
Beijing Normal University
Beijing, China
282843696@qq.com

***Abstract*—As traditional Chinese culture education grows fast, adaptive testing for ancient poetry seems promising. The prerequisite of adaptive learning is question bank, while the quality of question bank depends on the rationality of question difficulty. The paper proposes a method that measures ancient poetry difficulty for objective questions and builds a measurement system. The method includes five steps: (1) Identify the verses corresponding to the question. (2) Get four indexes of the corresponding verses: search volume, correctly answered frequency, sentence length and grade of the textbook which includes the verses. (3) Use analytic hierarchy process to index system for weight assignment. (4) Compute the weighted sum of the four indexes as the measurement of difficulty. (5) Classify the question bank according to the calculated difficulty. Experimental results showed the effectiveness of this measurement method, which thereby can be used in various adaptive testing of ancient poetry.**

Keywords-Adaptive Testing; Ancient Poetry ; Difficulty

I. INTRODUCTION

Ancient poetry is a treasure of Chinese literature. As the most important cultural memory of the Chinese people, ancient poetry has always attracted the attention of researchers. With the rise of the "fever of traditional Chinese learning", the education of ancient poetry has been paid more and more attention.

Previous studies on ancient poetry learning focused on learning content [1-2], learning methods [3-6], and learning strategies [5-6]. There was relatively little research on ability evaluation. Meng Qi [7] analyzed the examination questions of ancient poetry in middle school entrance examination from both quantitative and qualitative dimensions. It was found that the main reason for the low validity of the examination papers was that there was no effective difficulty arrangement, which made the examination questions simpler. Ren Xueyun [8] analyzed the characteristics of poetry reading test questions in Shanghai College Entrance Examination, summarized the problems existing in the test questions, and proposed that the test questions should reflect the prospects of individual cultural accomplishment. These studies focus on the assessment of ancient poetry competence in primary and secondary schools, ignoring the test needs of higher stage learners. Adaptive testing for ancient poetry may be a good way to meet the needs of different level poetry learners.

Adaptive testing is a personalized testing method based on Item Reflection Theory(IRT), which can provide subjects with test questions that are suitable for their ability level, and truly achieve the goal of "testing according to people". The current application of adaptive testing is mostly embodied in foreign language proficiency testing. In the adaptive test of foreign language proficiency, the subjects will complete the test questions of phonetic notation, word filling, etc. The final answer data can be used to update the difficulty value of the question bank so as to better serve the next adaptive test, thus forming a virtuous circle. Looking back at the self-adaptive test of ancient poetry, the predecessors have not attempted to build the ancient poetry question bank and quantify the difficulty of the test questions, let alone have constantly updated the answer data to predict the difficulty of the test questions.

The most important part of the IRT-based adaptive test system for ancient poetry ability is the construction of question bank. Whether the question bank system is good or not depends to a great extent on whether the quantification of the difficulty of a poem is reasonable. In the process of adaptive testing, the computer adaptive test (CAT) system acts as the "decision maker". Whatever strategy it adopts (such as maximum information method, hierarchical method, random method, etc.), its ultimate purpose is to select the exam questions with specific difficulty value matching the ability of the subjects from the question bank system. These questions are valuable for the subjects. Apart from being decisive in the selection of questions, the difficulty value of a poem also plays a crucial role in the subsequent estimation of the poem ability of the subjects. It directly affects the final ability level of the subjects and the reliability of the test. Therefore, how to quantify the ancient poetry difficulty for objective questions and grade the difficulty of the test questions to meet the practical needs of the Adaptive Testing of Ancient Poetry is a challenge.

The difficulty measurement of objective test questions of ancient poetry can be divided into two questions: the choice of difficulty quantification indexes and the assignment of weight. At present, domestic and overseas scholars seldom involve in quantitative research on the difficulty of ancient poetry. Professor Wang Zhaopeng of Wuhan University [9] launched the Tang Poetry Ranking List and Song Ci Ranking List. By means of quantitative scientific research, four kinds of statistical indicators were collected to measure the attention and influence of the poems: selected editions of past dynasties, commentaries of past dynasties, contemporary research and selection of literary history works. After reviewing Mr. Wang Zhaopeng's research, we think that there are some problems in his research, such as too strong subjectivity of indicators, unreasonable assignment of weights and incomplete measurement indicators. Specifically, it is unavoidable and impersonal for the subjectivity of selected editions, commentaries, contemporary studies and literary history works. Secondly, the rationality of assigning equal weights to the four indicators remains to be discussed. Thirdly, these

978-1-7281-5014-7/19/$31.00 ©2019 IEEE

four indicators should be strictly classified as one indicator, and the indicators are not complete.

Defects in difficulty calculation of ancient poetry test questions will inevitably lead to inaccuracy of the adaptive test system for ancient poetry in topic selection and ability evaluation, and its practical value will be greatly reduced. This paper designs a difficulty quantification system and proposes a method for calculating the difficulty of ancient poetry test questions. In the process of calculating difficulty, we choose objective and multi-dimensional quantitative indicators, and allocate the weight of indicators reasonably, so that the calculation results are more scientific and practical.

II. Quantitative System of Difficulty in Ancient Poetry

The paper proposes a method that measures ancient poetry difficulty for objective questions and builds a measurement system. The structure of the system is shown in Figure 1, which includes question bank module, quantitative indicators module, information acquisition module, difficulty calculation module and difficulty classification module. Among them, (1)the question bank module includes many kinds of objective questions of ancient poetry; (2)the quantitative indicators module is used to select the quantitative indicators of a poem difficulty (3)the information acquisition module is used to identify the verses corresponding to the question, and obtain the information of search volume, correctly answered frequency, sentence length and grade of the textbook which includes the verses; (4)the difficulty calculation module is used to calculate the difficulty of the test questions; (5)the difficulty classification module can classify the question bank according to the calculated difficulty.

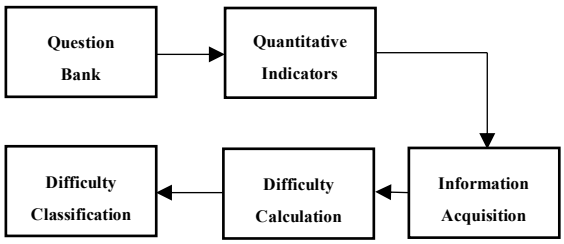

Figure 1. Quantitative System of Difficulty in Ancient Poetry.

A. Question Bank Module

In this paper, the ancient poetry question bank is derived from the previous research results of intelligent proposition of ancient poetry. The question bank contains selected labels, selected sentences, sorted sentences, selected style and crossword questions, as shown in the following examples:

- Selected Labels:

Select the following poem labels: “看朱成碧思纷纷，憔悴支离为忆君。不信比来长下泪，开箱验取石榴裙。”

A.思念　　B.惜时　　C.爱国　　D.离别

- Selection Style:

Select the following poetic style: “兰叶春葳蕤，桂华秋皎洁。欣欣此生意，自尔为佳节。谁知林栖者，闻风坐相悦。草木有本心，何求美人折?”

A.五言古诗　　B.五言律诗
C.五言绝句　　D.五言乐府

- Selected Sentence:

Choose the next sentence of the following poems: “感时花溅泪，_____。”

A.恨别鸟惊心　　B.昔有鹡鸰心
B.昔有鹡鸰心　　D.冥冥鸟去迟

- Sorted Sentences:

Choose the correct order of the following poems: “岱宗夫如何　齐鲁青未了　造化钟神秀　阴阳割昏晓。”

A.1234　　B.2314　　C.1324　　D.2413

- Crossword questions:

Choose the words to fill in the blank: “兰__春葳蕤，桂华秋皎洁。”

A.叶　　B.花　　C.树　　D.枝

B. Quantitative Indicators Module

The difficulty of a poem depends on many factors, such as: language, rhythm, genre, rhetoric, allusion, rhythm and ideological value. These factors can be summarized as internal factors; such as: attention, awareness, influence, transmission, etc. These factors can be summarized as external factors. This paper focuses on the difficulty of the poem in the CAT question bank. We believe that this difficulty can be understood as the tester's familiarity with poetry. In the face of familiar ancient poetry test questions, the probability of the tester to answer it is far greater than that of unfamiliar poem. For example, Li Bai's <蜀道难> and the Book of Songs <摽有梅>, from the internal factors such as metrics, genre, allusions, language, etc., the former is more difficult than the latter. But because the former is more familiar to the public, its attention, awareness and influence are far greater than the latter. In the adaptive test, the difficulty of <蜀道难> is less than that of <摽有梅>.

How can we effectively and scientifically obtain the public's familiarity with a poem through surveys and statistics? We then transform the abstract concept of familiarity into the influence of poem based on the degree of attention and dissemination. The higher the degree of attention and dissemination of the poem, the greater its influence, and the higher the public's familiarity with the poem. Based on the big data, we define the following indexes to measure the difficulty of poem:

1) Internet Search Volume

Nowadays, the Internet is an important medium for the dissemination of ancient poetry. It is also an important way for ordinary readers to understand, read and comment on ancient poetry. The more transcribed by the web page, the more popular the poem is and the more attention it receives. Therefore, we can use the link rate of the network to measure the attention of different works in the general readership.

2) Frequency of Occurrence

Frequency of occurrence can be regarded as a statistical feature of the difficulty of the poem, that is, the number of repeated occurrences of the poem in a certain number of real corpus. It determines the degree of common use of the poem

and the degree of familiarity of poetry learners, so it is an important factor to quantify the difficulty of a poem. Generally speaking, the higher the frequency of the use of a poem, the higher the public's daily perception, the less difficult it will be in memory.

3) Single Sentence Length

The difficulty of memorizing often affects the judgment of difficulty. The easy-to-remember sentences are easy to pass from mouth to ear. Therefore, the length of sentences is regarded as one of the difficulty indicators of quantitative topics in this paper.

4) Textbook Appearance Level

Using the above three indicators, we can roughly get the ideal difficulty of ancient poetry, but as mentioned above, this study focuses on the difficulty of the poem in the adaptive test bank. Most of the people's poetry reserve comes from textbooks of primary and secondary schools. In view of this situation, this paper regards the distribution level of ancient poetry in textbooks as a penalty factor for difficulty discrimination. The lower the grade in textbooks, the greater the difficulty value.

C. Information Acquisition Module

1) Internet Search Volume (S)

Internet Search Volume is obtained by using the precise search and fuzzy search mode of Baidu search engine. There are two steps to obtain it:

Step 1, the question is searched as a keyword, and the result is a fuzzy search quantity.

Step 2, search<"question ">as a key word, and the result is precise search quantity.

Step 3, the sum of the results of the two modes is the internet search volume of the poem.

2) Frequency of Occurrence (F)

The frequency of inscription stems comes from the frequency of visits in the knowledge map of ancient poetry constructed in the earlier period. The higher the frequency, the more familiar the public is with it, the simpler the title is.

3) Single Sentence Length(L)

The length of a single sentence is to calculate the number of Chinese characters in a given sentence. For example, the single sentence of the poem "两只黄鹂鸣翠柳，一行白鹭上青天" is 7.

4) Grade of Textbook Distribution(R*)

We compiled the poems included in the different versions of elementary, middle, and high school textbooks, and used the distribution level of the questions in the textbook as the penalty factor for difficulty discrimination. The primary school textbook is marked as 12–7 by grade, 6–4 in junior high school, 3–1 in high school, and 0 in non-existing.

D. Difficulty Calculation Module

In this study, the indexes that affect the difficulty of a poem are Internet search volume, frequency, sentence length and textbook grade. The four kinds of indexes have different influence on the determination of a poem difficulty. Therefore, we use linear weighted summation method to preliminarily define the formula for calculating the difficulty of ancient poetry titles (1):

$$*b = S \times W_1 + F \times W_2 + L \times W_3 + R* \quad (1)$$

Among them, the internet search volume (S) and the frequency of occurrence(F) are negatively correlated with the difficulty value; the single sentence length(L) and the grade of textbook distribution are positively correlated with the difficulty value.

W_1、W_2、W_3 are the weights of S、F、and L respectively.How to determine the weight of data, there are no precedents in literary measurement research, but in bibliometrics, there are two ways of subjective weighting and objective weighting. Among them, the subjective weighting method is mainly from the perspective of qualitative analysis, according to the size of each indicator to determine the impact of the object to determine the weight of the corresponding indicators, but also can be divided into expert evaluation method and analytic hierarchy. Objective weighting mainly calculates the weight of each index from the perspective of quantitative analysis, and there are two methods of coefficient of variation and correlation coefficient. Considering the specific situation of literary research, we choose the analytic hierarchy process(AHP) in subjective weighting [8] to determine. The basic steps of the analytic hierarchy process are as follows:

Step 1, we invited 10 experts in this field to grade the importance of the indexes and construct the judgment matrix as follows:

$$A = \begin{bmatrix} 1 & 3 & 5 \\ \frac{1}{3} & 1 & 4 \\ \frac{1}{5} & \frac{1}{4} & 1 \end{bmatrix}$$

Step 2, according to the matrix, W1, W2 and W3 are computed as 0.62, 0.28 and 0.10 respectively.

Step 3, we tested the consistency of the three weights and obtained the consistency ratio 0.075<0.1, which shows that the weights we got are effective and reliable.

On this basis, considering the numerical differences among internet search volume, frequency of occurence and single sentence length, we get the final formula for determining the difficulty of ancient poetry (2):

$$*b = -log_{100}(S + 0.001) * 0.62 - log_{100}(F + 0.001) * 0.28 + L * 0.1 - R* \quad (2)$$

It is worth pointing out that this difficulty is aimed at self-adapting test of the difficulty of a poem. It considers the attention and influence of the relevant poems, and cannot be used to evaluate the artistic and ideological value of each poem. Poems with low difficulty value can only be said to have greater influence in the long process of communication and acceptance, which does not mean that they have little artistic value and little ideological significance.

E. Difficulty Classifition Module

In order to improve the applicability of the question bank to various test methods, we calculate the difficulty of all the questions according to the difficulty calculation formula defined above, and classify the question bank. The graded question bank can also be used for poetry test of other modes (such as breaking through customs), and can also be decomposed into Test Bank of different levels of ancient poetry level test.

1) Ancient Poetry Pass-through Game

The whole ancient poetry question bank is divided into 60 levels. According to the grade of textbook distribution(R*), the test questions with non-zero textbook grades (corresponding to the ancient poems in the textbook) are arranged in ascending order of difficulty and then data is divided into 15 levels, representing low difficulty; the test questions marked with a textbook grade of 0 (corresponding to the textbook's external poems) are arranged in ascending order of difficulty and then data is divided into 45 levels, representing moderate difficulty.
tThe flow chart of the game is shown in Figure 2.

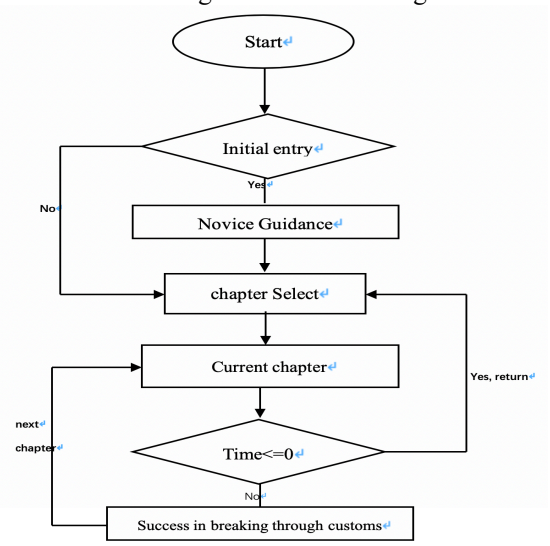

Figure 2. Ancient Poetry Pass-through Game.

2) Ancient Poetry Ability Test

The entire ancient poetry question bank is divided into 25 grades. The overall difficulty level of the first 15 grades is low. Within the scope of textbooks in primary and middle schools, the number of questions and the number of ancient poems involved in each level are small, and the applicable population is wide. The last 10 levels are medium to high difficulty, and the number of questions and the number of ancient poems per level are large, aiming at refining the ancient poetry of middle and high level people. The flow chart of the ancient poetry ability test is shown in Figure 3.

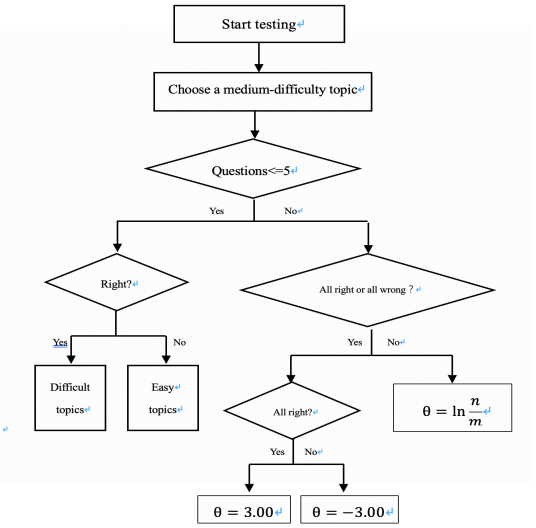

Figure 3. Ancient Poetry Ability Test.

III. Validation of Difficulty Effectiveness

We use the following methods to evaluate the effectiveness of the difficulty calculation of the question stem:

1) 100 questions were randomly extracted from the question bank, and the difficulty of these 100 questions was calculated as the sample X to be tested.

2) Using min-max standardization to normalize the difficulty of questions to 1-10, the calculation formula of min-max standardization is as follows:

$$f(b)_{\text{scaled}} = \frac{b - b_{\min}}{b_{\max} - b_{\min}} \cdot (\max - \min) + min \quad (3)$$

Among them, $b_{\max}$ and $b_{\min}$ correspond to the maximum and minimum difficulty values calculated in the whole question bank respectively, max and min correspond to the maximum and minimum values after normalization respectively, where 10 and 1 are taken respectively.

3) Twenty postgraduates majoring in Chinese were asked to grade the difficulty of 100 questions, with the score limited to 1-10, and the average of 20 scoring tables was taken as the final target sample Y.

4) The Pearson correlation coefficient is used to measure the correlation between the tested sample X and the target sample Y. The Pearson formula selected in the present invention is as follows:

$$\rho_{X,Y} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum(X-\bar{X})^2 \sum(Y-\bar{Y})^2}} \quad (4)$$

Generally, the correlation strength of the judgment variable can be based on Table 1.

TABLE I. Pearson Correlation Coefficient Meaning Table

Coeffcient	Correlations
0.8-1.0	Extremely strong
0.6-0.8	Strong
0.4-0.6	intermediate
0.2-0.4	Weak
0.0-0.2	Extremely weak

Finally, the correlation coefficient between the test sample X and the target sample Y is 0.862 (three decimal places are reserved), and the difficulty calculated by the difficulty quantization method is highly correlated with the difficulty judged by the professional. It can be proved that the method for quantifying the difficulty of the ancient poetry proposed by the present invention is reasonable.

IV. Conclusion

This paper proposes a method to quantify the difficulty of a poem, which gives a score of difficulty for each ancient poetry in the adaptive test bank. We chose Internet search volume, frequency of occurrence and single sentence length as the quantitative criteria of difficulty, and assign the weight for each criteria with analytic hierarchy process (AHP). At the same time, we used the distribution of textbooks as a penalty to improve the effectiveness of system. Furthermore, we divided the difficulty-calculated question bank into several levels so that it can be adapted to more test forms. In order to validate method, we selected subjects with different educational levels to conduct

experiments. The results show that the difficulty of ancient poetry calculated by our method is reasonable and feasible. Of course, there are still many shortcomings in our research, such as more quantitative indicators are not considered, the level of subjects needs to be refined and so on. We expect that these shortcomings will be better remedied in the future work.

REFERENCES

[1] Z. Ting, "A Study on the Joining Points of Ancient Poetry in Textbooks of Grade Four to Seven of Shanghai Educational Press", Shanghai Normal University, 2018.

[2] X. Kexiong, "What Should Students Learn in Ancient Poetry Teaching", Inner Mongolia Education, no. 10, pp. 273-297, 2010.

[3] C. Ximmin, "Exploration of Learning Methods for Ancient Poetry Literature", China Educational Technology & Equipment, no. 27, pp. 63-63, 2010.

[4] W. Xianwen, "Teaching Practice of Five-step Learning Method of Ancient Poetry", Bulletin of Chinese Language Teaching, no. 6, pp. 11-11, 2001.

[5] C Jianwei, "Research on Teaching Strategies of Chinese Ancient Poetry in Primary Schools under the Background of New Curriculum Reform", Education for Chinese After-school(Theory), no. 1, 2014.

[6] C. Luyi, "The High School Language Ancient Poetry Teaching Innovation Method Research", Hunan Normal University, 2013.

[7] M. Qi, "Research on the Testing Objectives of the Poem Test Question in Shanghai Entrance Examination for Senior Middle School", Shanghai Normal University, 2015.

[8] W. Rufei, "An Analysis of Strategies in Teaching Chinese Ancient Poetry and Ci in Junior Middle School", China Educational Technology & Equipment, no. 27, pp. 63-63, 2010.

[9] W. Zhaopeng, *Ranking List of Tang Poetry.* Shanghai,CN: Shanghai Classics Publishing House, 2011.

[10] L. Peng, "Research on Several Key Technologies of Computer Adaptive Testing", Dongbei Normal University, 2012.

[11] L. Liping, W. Wenjie and G. Shining, "Design and Realization of the Item Pool System for CAT", Computer Systems & Applications, vol. 22, no. 9, pp. 10-12, 2006.

[12] L. Zhaosheng, *Theoretical Basis of Item Response.* Beijing,CN: Beijing Normal University Publishing Group, 2012.

[13] H. Yueh-Min, L. Yen-Ting and C. Shu-Chen, "An Adaptive Testing System for Supporting Versatile Educational Assessment", Computers & Education, vol. 52, no. 1, pp. 53-67, 2008.

[14] M. Barla, M. Bieliková, A. Bou Ezzeddinne, "On the impact of adaptive test question selection for learning efficiency", Computers & Education, vol. 52, no. 1, pp. 53-67, 2008.