The Item Response Theory Model for an AI-based Adaptive Learning System

Wei Cui[1], Zhen Xue[1], Jun Shen[2], Geng Sun[2], Jianxin Li[3]

[1]*Squirrel AI Learning Inc., Shanghai, China*

[2]*School of Computing and Information Technology, University of Wollongong, NSW, Australia*

[3]*School of Information Technology, Deakin University, Melbourne, Australia*

Emails: cuiwei@songshuai.com, xuezhen@mail.ustc.edu.cn, {jshen, gsun}@uow.edu.au, jianxin.li@deakin.edu.au

Abstract—**Item characteristics (e.g. item difficulty), and students' latent traits (e.g. student ability) are essential in a personalized learning system. In such system, items of different characteristics need to be recommended according to students' latent traits. The item response theory can fulfill such requirements. In this paper, we present a detailed description of the student ability and the item difficulty measurement based on an item response theory model. The model is evaluated using real data regarding the math and the English curriculum gathered from students' learning processes during the usage of the Squirrel AI Learning System. The data is regarding the math and the English curriculum. It is proved that the difficulty evaluated by teachers minus the difficulty estimated from the data approximately follows a normal distribution. The mean value of students' abilities approaches 0.50 as the numbers of their attempts on questions increase.**

Keywords—Item difficulty, Ability measurement, Item response theory, Big data, Logistic model, Maximum likelihood estimation

I. INTRODUCTION

Online tutoring, such as the delivery of Massive Open Online Course (MOOC), is popular these days. However in such tutoring systems, it is difficult for a learner to locate appropriate learning materials matching his ability and therefore he would lose engagement gradually. As a result, the retention rate is low. The retention rate for classes taught through distance education is 10–20% lower than classes taught in a face-to-face setting [1, 2]. In face-to-face classes, teachers can adjust the difficulty of learning materials according to students' performance. In this case, the difficulty should not be so high that the students cannot follow; meanwhile, the difficulty should not be so low that the students sense no challenge.

The Squirrel AI Learning System integrates artificial intelligence and adaptive learning technology to imitate a private tutor [3]. It grasps students' attention by adjusting the learning content and difficulty in time according to students' individual knowledge states. It runs on a cloud server, to which students can access at any time via internet browsers, as shown in Figure 1. We refer readers to [4, 5, 6] for its user interfaces and working mechanisms.

Figure 1. Shanghai students learning math using Squirrel AI

An item could be in the form of text, video, or animation etc. "Item" requires a test/assessment taker to respond in order to fulfill the requirements of the item. Item characteristics (e.g. item difficulty) and students' latent traits (e.g. student ability/proficiency) are essential in a personalized learning system. In this case, items of different characteristics should be recommended to students according to their latent traits. This can be realized using the item response theory (IRT). The IRT builds an item response function that computes the probability of a correct response by considering the item difficulty and the student ability. There have been many models developed on the IRT. Linden and Rambleton [7] made a thorough review of different IRT models and the role of each of their parameters. The first IRT model is a two-parameter Gaussian ogive curve model proposed by Lord [8].

The Gaussian ogive curve model is not convenient for the integral in its formula. Therefore the logistic model was first introduced by Birnbaum into IRT for the psychological measurement [9, 10, 11, 12]. There is no integral in logistic function. The logistic model takes a logistic function as the item response function of IRT.

The maximum likelihood estimation was introduced by Maxwell for measuring the parameters of the logistic model [13]. Maximum likelihood estimation is the most common technique used to measure the item parameters. We refer readers to Baker and Kim's paper for a specific implementation of the maximum likelihood estimation [14]. Maximum-likelihood-estimation based expectation-maximization (EM) was applied by Bock and Aitkin to the item parameter measurement [15]; this algorithm was again utilized in the student ability measurement [16, 17].

In this paper, we present the measurement of the item difficulty and student ability according to a one-parameter logistic model. The model is evaluated by using real response data of students learning math and English curriculums provided in the Squirrel AI Learning System, an AI-based adaptive learning system. We use the abbreviation "Squirrel AI" in this paper for convenience.

II. DETERMINE ABILITY VALUE WITH ITEMS OF KNOWN DIFFICULTY

IRT has been widely used in computer-supported tests such as GRE and TOEFL. Some IRT-implemented systems require the examinees to finish learning in a strict time. In this case, examinees might not have sufficient time for the entire items, which would affect the measurement of the difficulty of individual items and the student's ability. Squirrel AI adopts the pedagogy of mastery-based learning, which is also called competency-based learning. More specifically, it is only when students have mastered the

current learning objectives can they commence the new learning objectives [18]. Therefore, they have enough time to finish every item.

In Squirrel AI, the learning content of a curriculum is divided into many fine-grained knowledge components, which are also called skills in some literature. These components constitute a knowledge graph. Every knowledge component has at least 20 items with varying difficulty levels. To better engage students, the items' formats are diverse in their formats. The typology of items could be in the form of text, slides, lecturing video, or animation. When a student enters Squirrel AI for the first time, he is set up with a pre-test to diagnose his knowledge states: what he has mastered. Then he is recommended learning contents on his weak knowledge components, in other words, what he has not mastered. The difficulty levels of items were initially determined by education experts. According to his performance on the items from each knowledge component a student has different ability levels on different knowledge components.

The types of math items include blank-filling, single-choice, multiple-choice, calculation, application, etc. The types of English language items include blank-filling, single-choice, multiple-choice, translation, reading comprehension, listening comprehension, etc. Squirrel AI adopts a binary logistic response model: the response is labelled as correct only when a student's answer is totally correct. In this model, a student's probability of correctly answering the ith item is computed using the following logistic function:

$$P = \frac{1}{1+e^{-D(\theta-b_i)}},$$

in which D is a constant 1.7, b_i is the difficulty value of the ith item and θ is the student's ability. Assume that a student's response to the ith item is u_i. If the student's response to the item is right, then $u_i = 1$. If the student's response to the item is incorrect, then $u_i = 0$. The student's response vector to n items is $u=(u_1, u_2, u_3 ... u_n)$. Thus, according to probability theory, the probability of observing the student's responses u is:

$$\prod_{i=1}^{n} P_i^{u_i} Q_i^{1-u_i}$$

Which is called likelihood function, noted as $L(u\,|\,\theta,b)$,

$$L(u\,|\,\theta,b) = \prod_{i=1}^{n} P_i^{u_i} Q_i^{1-u_i} \qquad (1.1)$$

In which， $P_i = \dfrac{1}{1+e^{-D(\theta-b_i)}}$ \qquad (1.2)

$$Q_i = 1 - P_i = 1 - \frac{1}{1+e^{-D(\theta-b_i)}} \qquad (1.3)$$

We applied the maximum likelihood estimation to find optimal values for parameters θ and b that maximizes the likelihood function L.

In Squirrel AI, the initial difficulty levels of items were manually set up by expert teachers. Therefore, the parameter b is known from the beginning, and $L(u|\theta,b)$ can be denoted as $L(u|\theta)$. When L's first-order partial derivative with respect to θ is 0, L takes the maximum value, namely

$$\frac{\partial L(u\,|\,\theta)}{\partial\theta} = 0, \qquad (2)$$

$$\frac{\partial^2 L(u\,|\,\theta)}{\partial^2\theta} < 0 \qquad (3)$$

Equation (2) is called the likelihood equation. The solution of equation (2) is the student's ability value.

In some cases, it is more convenient to maximize the logarithm of $L(u\,|\,\theta)$. The following equation is called the logarithmic likelihood function:

$$\ln L(u\,|\,\theta) = \sum_{i=1}^{n} [u_i \ln P_i + (1-u_i) \ln Q_i] \qquad (4)$$

According to calculus theory, we have that

$$\frac{\partial \ln L(u\,|\,\theta)}{\partial\theta} = \frac{\partial \ln L(u\,|\,\theta)}{\partial L(u\,|\,\theta)} \cdot \frac{\partial L(u\,|\,\theta)}{\partial\theta} = \frac{1}{L(u\,|\,\theta)} \cdot \frac{\partial L(u\,|\,\theta)}{\partial\theta} \qquad (5)$$

From equation (1) we know $\dfrac{1}{L(u\,|\,\theta)} > 0$. Therefore from equation (2) and (5) we have that

$$\frac{\partial \ln L(u\,|\,\theta)}{\partial\theta} = 0 \qquad (6)$$

Equation (6) is called logarithmic likelihood equation. Equation (6) is equivalent to equation (2).

Equation (2) or (6) is usually nonlinear equation. In such case, numerical method is used to find the maximum of L or $\ln L$ to obtain the solution of equation (2) or (6).

Conventionally, the item difficulty value b is in the range of $-\infty$ and $+\infty$. In Squirrel AI, the expert teachers tend not to design extremely difficult and easy questions. As a result, extremely difficult items are much fewer than moderately difficult items. For easier understanding and expression, we normalize the difficulty range from the natural difficulty ($-\infty,+\infty$) to a normalized difficulty (0,1) by the function

$$normalizedDiff\text{iculty} = \frac{e^{naturalDifficulty}}{e^{naturalDifficulty}+1}$$

Figure 2 shows the relationship between the natural difficulty and the normalized difficulty. We divide the normalized difficulty interval (0,1) into 100 equidistant intervals. The expert teachers determine the initial difficulty b of each question to be 0.00, 0.01, 0.02, ..., 1.00 according to their experience.
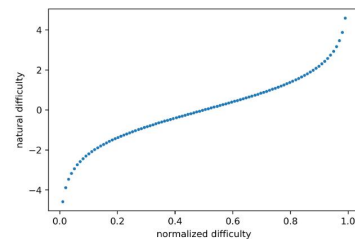

Figure 2. Natural difficulty versus normalized difficulty

For example, in Squirrel AI, there are 5 math items whose difficulty values were initially set by expert teachers to be $b_1 = 0.95$, $b_2 = 0.77$, $b_3 = 0.83$, $b_4 = 0.89$ and $b_5 = 0.89$. According to the one-parameter logistic model (1PLM), the probability of the jth student with the ability θ correctly answering the ith item with the difficulty b_i is

$$P_i = P(\theta, b_i) = \frac{1}{1 + e^{-D(\theta - b_i)}} .$$

The probability of the student answering the item incorrectly is $1-P_i$, denoted as Q_i. For instance, suppose that the student responded to the above 5 items and the response vector $u=<0,1,1,0,1>$, in this case, the likelihood function is:

$$L(\theta) = \prod_{i=1}^{5} P_i^{u_i} Q_i^{1-u_i}$$

$$= Q_1 P_2 P_3 Q_4 P_5 = (P_1^0 Q_1^1)(P_2^1 Q_2^0)(P_3^1 Q_3^0)(P_4^0 Q_4^1)(P_5^1 Q_5^0)$$

$$= (P_1^{u_1} Q_1^{1-u_1})(P_2^{u_2} Q_2^{1-u_2})(P_3^{u_3} Q_3^{1-u_3})(P_4^{u_4} Q_4^{1-u_4})(P_5^{u_5} Q_5^{1-u_5})$$

With equation (1.1) and (1.2), we have

$L(\theta)$

$$= \frac{e^{-1.7(\theta-b_1)}}{1+e^{-1.7(\theta-b_1)}} \cdot \frac{1}{1+e^{-1.7(\theta-b_2)}} \cdot \frac{1}{1+e^{-1.7(\theta-b_3)}} \cdot \frac{e^{-1.7(\theta-b_4)}}{1+e^{-1.7(\theta-b_4)}} \cdot \frac{1}{1+e^{-1.7(\theta-b_5)}}$$

$$= \frac{e^{-1.7(\theta-0.95)}}{1+e^{-1.7(\theta-0.95)}} \cdot \frac{1}{1+e^{-1.7(\theta-0.77)}} \cdot \frac{1}{1+e^{-1.7(\theta-0.83)}} \cdot \frac{e^{-1.7(\theta-0.89)}}{1+e^{-1.7(\theta-0.89)}} \cdot \frac{1}{1+e^{-1.7(\theta-0.89)}}$$

The likelihood function $L(\theta)$ is shown in Figure 3. By using the numerical method, we can solve the likelihood equation, and get the solution $\theta = 1.11$. This means that the jth student's natural ability is estimated to be 1.11. Furthermore, his normalized ability should be 0.75 according to the relationship shown in Figure 5.

In Figure 3, it can be found that the curves of P3 and P2 have different influences on the curve of $L(\theta)$. More specifically, if the item 2 (for P2) is more difficult than the item 3 (for P3), the $L(\theta)$ curve will shift rightward. This results in that the optimal θ value obtained with the curve $L(\theta)$ becomes bigger.
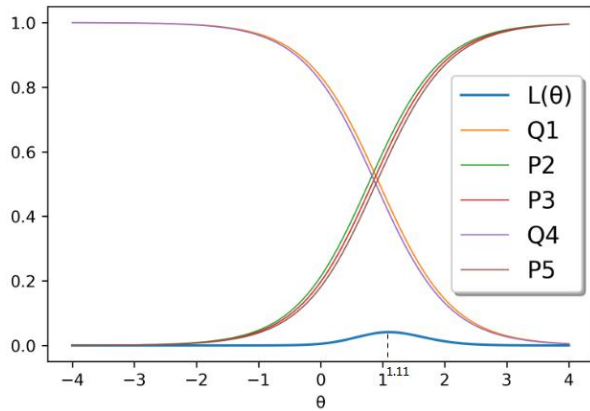

Figure 3. The likelihood function

We can obtain the same ability value by using the logarithmic likelihood function, as shown in Figure 4. The aforementioned example describes the calculation method of

the ability value. In this way, a student's ability value on a knowledge component is updated at every time he finishes an item from that knowledge component. Thereafter the difficulty value of the topic is also updated.

As shown in Figure 3, if a student answers only one item and correctly, his ability is estimated as $+\infty$. If he answers all the n items correctly, his ability is also estimated to be $+\infty$. If he answers none of the n items correctly, his ability is estimated to be $-\infty$. In these extreme cases, the students would be recommended with other knowledge components. For example, if he fails on 3 consecutive items from the same knowledge component, he will be recommended the items of prepositive knowledge component. Prepositive knowledge component is more fundamental than the current knowledge component.
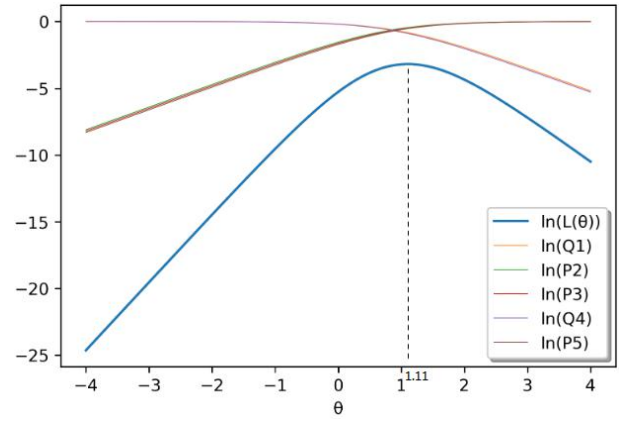

Figure 4. The logarithmic likelihood function

For the same reason of normalizing the item difficulty, we normalize the student ability value θ from $(-\infty, +\infty)$ into the range $(0,1)$ using the same function shown in Figure 2

$$normalizedAbility = \frac{e^{naturalAbility}}{e^{naturalAbility} + 1}$$

Figure 5 shows the relationship between the natural ability and the normalized ability. We also divide the normalized ability interval $(0,1)$ into 100 equidistant intervals.
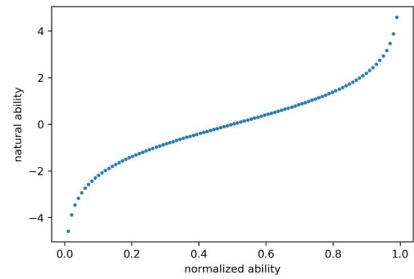

Figure 5. Natural ability versus normalized ability

III. DETERMINE ITEM DIFFICULTY VALUE ACCORDING TO STUDENTS' PERFORMANCE ON IT

The difficulty value is determined using the maximum likelihood estimation method. Suppose that the jth item has been responded by n students with known ability values.

For example, in Squirrel AI, the jth item has been attempted by 3 students (we take 3 students as example for the simplicity of illustration), whose ability values are

$\theta_1=0.07$, $\theta_2=0.25$ and $\theta_3=0.09$ respectively. Next we need to update the difficulty value of the jth item, denoted as b.

According to the one-parameter logistic model (1PLM), the probability of a student with the ability θ_1 answering correctly item of difficulty b, is:

$$P_i = P(\theta_i, b) = \frac{1}{1+e^{-D(\theta_i - b)}}$$

The probability of the student with the ability θ_i answering the item of difficulty b incorrectly, is $1-P_i$, denoted as Q_i. The vector of the three students' response against the item is $u=<0,1,1>$. Then the likelihood function is the following:

$$L(b) = Q_1 P_2 P_3 = (P_1^0 Q_1^1)(P_2^1 Q_2^0)(P_3^1 Q_3^0)$$

$$= (P_1^{u_1} Q_1^{1-u_1})(P_2^{u_2} Q_2^{1-u_2})(P_3^{u_3} Q_3^{1-u_3})$$

$$= \prod_{i=1}^{3} P_i^{u_i} Q_i^{1-u_i}$$

With equation (1.1) and (1.2), we have

$L(b)$

$$= \frac{e^{-1.7(\theta_1-b)}}{1+e^{-1.7(\theta_1-b)}} \cdot \frac{1}{1+e^{-1.7(\theta_2-b)}} \cdot \frac{1}{1+e^{-1.7(\theta_3-b)}}$$

$$= \frac{e^{-1.7(0.07-b)}}{1+e^{-1.7(0.07-b)}} \cdot \frac{1}{1+e^{-1.7(0.25-b)}} \cdot \frac{1}{1+e^{-1.7(0.09-b)}}$$

The likelihood function $L(b)$ is shown in Figure 6. Using the numerical method, we obtain the solution $b = -0.27$ for the likelihood equation. Therefore, the jth item's natural difficulty value is -0.27 and the normalized difficulty value is then 0.43 according to the relationship shown in Figure 5.

It is shown in Figure 6 that the curves of P_3 and P_2 have different influence on the curve of $L(b)$. If the item is answered incorrectly by a student with the ability θ_2 and θ_2 bigger than θ_3, the $L(b)$ curve will shift more rightward. This means that the optimal value b obtained by curve $L(b)$ is bigger. In this case, not only whether a student answers correctly but also the student's ability matters. This is different from the classical test theory, in which the difficulty is calculated by the false rate of all the students, not considering the students' abilities.
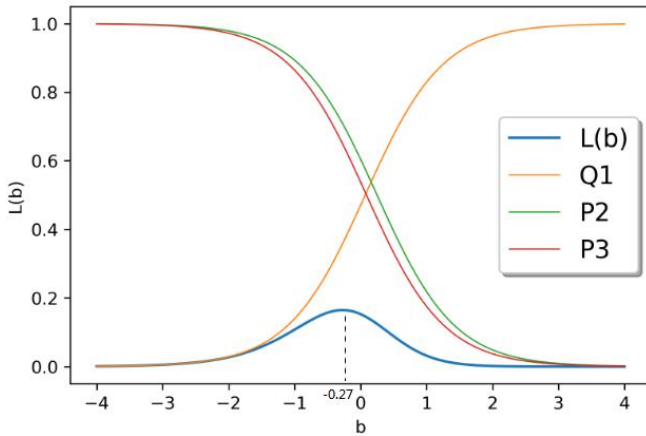

Figure 6. The likelihood function

We can obtain the same ability values by using the logarithmic likelihood function, as shown in Figure 7.
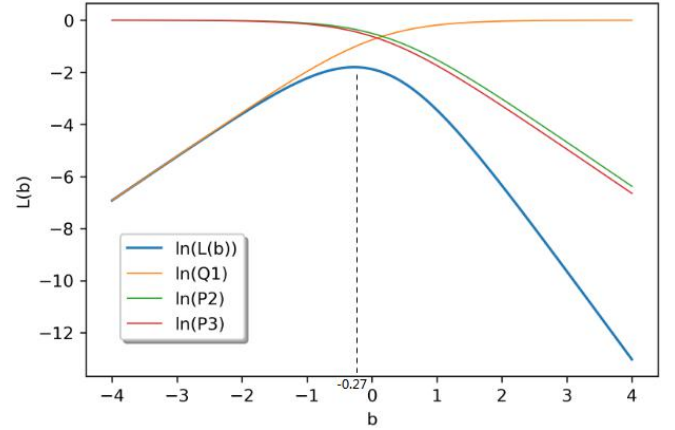

Figure 7. The logarithmic likelihood function

As shown in Figure 6, if an item is answered by only one student and the answer is correct, the item difficulty is estimated to be -∞. If the item is answered correctly by all of the n students, its difficulty is also -∞. If none of the n students answers it correctly, its difficulty is estimated to be +∞. These extreme cases are rare in practice as the expert teachers avoid designing the items to be very difficult or easy. The difficulties of most items are moderate as shown in Figures 10 and 14.

IV. THE ABILITY AND DIFFICULTY DISTRIBUTION IN MATH CURRICULUM

The data used in the empirical study was collected from 60,415 students from grade 7 to 9 in middle schools. According to the algorithm described in section II, the ability distributions are calculated from the real data of students learning the math curriculum. Figure 8 shows these distributions. Totally, there are 72,649 items which have been attempted 8,647,156 times.

In Figure 8(a), the entire set of data was applied. In this case, the mean value of the ability estimates is 0.64. In Figure 8(b), we only consider students who have attempted a knowledge component for at least 4 times. In this case, the mean value of the ability estimates is 0.54. In Figure 8(c), we only take into account those who have attempted a knowledge component at least 8 times. In this case, the mean value of the ability estimates is 0.51. Note that a student may have different ability values on different knowledge components. Therefore, the name of y-axis should not be "Number of students". Instead, we call it "No. of abilities". Comparing the Figure (a), (b) and (c), it can be observed that the mean ability becomes closer to 0.50 along with the number of attempts increasing. This is in accordance with the hypothesis that if the number of attempts is large enough, the mean value of students' ability estimates should approach 0.50, as their natural abilities tend to be random.
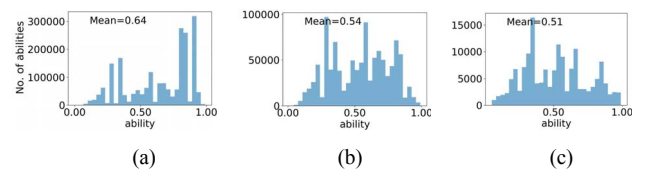

| (a) | (b) | (c) |

Figure 8. Ability value obtained through data of math curriculum

In Squirrel AI, the difficulty value of each question was initially set by some expert teachers, and then peer-reviewed independently by another teacher. The difficulty distribution is shown in Figure 9.
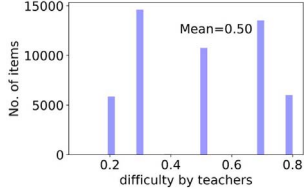

Figure 9. Difficulty values evaluated by expert teachers who designed the math curriculum

After students' usage of the system, the difficulty distribution of the items is shown in Figure 10. In this case, the difficulty values are computed by the method described in section 3. In Figure 10(a), the entire data is taken into account. In Figure 10(b), we only consider the data of items that have been attempted at least 10 times. We can see that the histogram distribution in Figure 10(a) is spikier than the one in Figure 10(b) even with fewer samples. In Figure 10, the regression curves in deep blue are obtained by kernel density estimation. The outliers in Figure 10(a) are more distant from the regression curve in deep blue. This is because the more times an item is attempted, the more precisely the difficulty value can be measured.
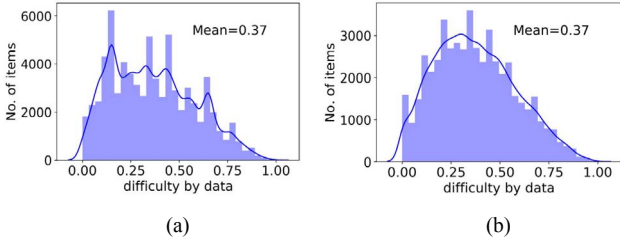

(a)　　　　　　　　　(b)

Figure 10. Difficulty values obtained from data of the math curriculum

Figure 11 shows the distribution of difficulty by teachers minus difficulty estimated from the data, i.e. difficulty shift. In Figure 11(a), all data was taken into account. In Figure 11(b), we only take the data of items that has been attempted for at least 10 times. We can see that the distribution complies with a normal distribution. This means that the difficulty by teachers and difficulty by data are in accordance generally. The mean value of this distribution is 0.12. This means the math teachers tend to overestimate the item difficulty.
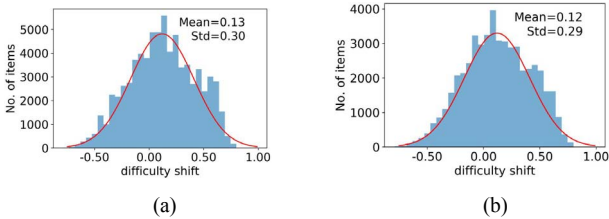

(a)　　　　　　　　　(b)

Figure 11. Difficulty value shift of math curriculum

V. THE ABILITY AND DIFFICULTY DISTRIBUTION IN THE ENGLISH LANGUAGE CURRICULUM

Same as the analysis on the math data described in the section IV, the distribution of student abilities, in this case, is obtained from the real data of students learning the English curriculum. Figure 12 shows this distribution. The data was collected from 27,040 students from grade 7 to 9 in middle school. The students attempted 44,291 items 2,468,661 times in total.

In Figure 12(a), the entire data is examined. In Figure 12(b), we again consider students who have attempted a knowledge component for at least 4 times. In Figure 12(c), only those who have attempted a knowledge component for at least 8 times are considered. The more attempts, the closer the mean value of the ability estimates approaching 0.50. In this case, the ability distribution is in accordance with the hypothesis that if the number of attempts is sufficient, the mean value of students' ability estimates should approach 0.50, as the natural abilities tend to be random.
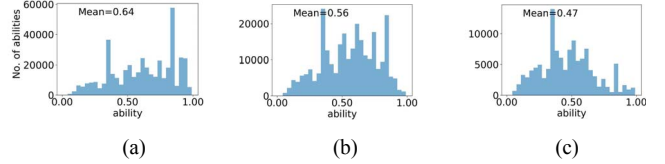

(a)　　　　　　(b)　　　　　　(c)

Figure 12. Ability value obtained through data of English curriculum

Again, the difficulty values of the English language items were firstly determined by expert teachers, and then reviewed independently by other teachers. The difficulty distribution is shown in Figure 13.
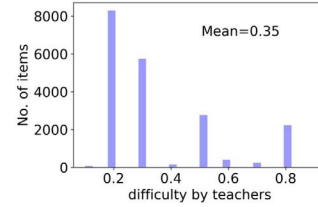

Figure 13. Difficulty value set by teachers of English curriculum

After students' usage of the system, the difficulty distribution of these items is shown in Figure 14. They were obtained by the method described in section III. In Figure 14(a), the entire data was taken into account. In Figure 14(b), we only consider the data of items that have been attempted at least 10 times. The same as in Figure 8, Figure 14(a) is spikier than Figure 14(b).
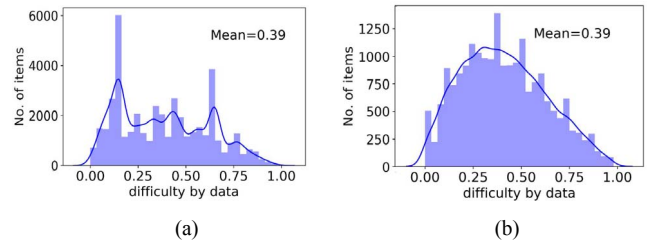

(a)　　　　　　　　　(b)

Figure 14. Difficulty value obtained through data of English curriculum

Figure 15 shows the distribution of the difficulty shift on English items. In Figure 15(a), the entire data was investigated. In Figure 15(b), we only take the data of items that have been attempted for at least 10 times. The facts shown in Figure 15(a) and 15(b), especially Figure 15(b), comply with a normal distribution. Its standard deviation is approximately the same as that of the distribution for the math items. This means that the consensus between teachers and data on the English items are the same as that on the math items. In this case, the mean value of the distribution is -0.05. This means that the English language teachers tend to underestimate the item difficulty, but the bias is smaller than that from the math teachers.
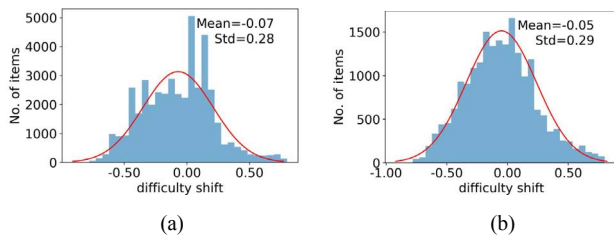
Figure 15. Difficulty value shift of English curriculum

VI. CONCLUSION

In this paper, we present a logistic model used in the Squirrel AI learning system. With this model, we can measure the student ability and the item difficulty from the real data of students learning the math and the English curriculums. The learning system follows the pedagogy of mastery-learning, so that the students have sufficient time to finish every item, which avoids the behavior of random guessing caused by time limitation. We obtain the likelihood equation with the maximum likelihood estimation. We found that the difficulty evaluated by expert teachers minus difficulty estimated from the data approximately follows a normal distribution. We also found that the mean value of students' abilities approaches 0.50 as the number of attempts increases. This proves our hypothesis that the students' natural abilities tend to be random. In the future, we will take into account the response time to more accurately estimate the student ability and the item difficulty.

ACKNOWLEDGMENT

REFERENCES

[1] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé III, and L. Getoor, "Modeling learner engagement in MOOCs using probabilistic soft logic." *NIPS Workshop on Data Driven Education*, Vol. 21, p. 62, December, 2013

[2] L. Angelino and D. Natvig, "A Model for Engagement of the Online Learner." *Journal of Educators Online*, 6(1), 2009, 1-19. Retrieved from http://www.thejeo.com/Archives/Volume6Number1/Angelinoetalpaper.pdf

[3] Z.-M. Qi, Y. Zhao, *The People's Daily,* April 17th, 2019.

[4] H. Li, W. Cui, Z. Xu, Z. Zhu and M. Feng, 2018. Yixue adaptive learning system and its promise on improving student learning. *Proceedings of the 10th International Conference on Computer Supported Education*. http://www.scitepress.org/Papers/2018/66898/66898.pdf

[5] M. Feng, W. Cui, and S. Wang, 2018, Adaptive learning goes to China. *Artificial Intelligence in Education*, 10948, 89-93.

[6] W. Cui, Z. Xue, K.-P. Thai, "Performance comparison of an AI-based Adaptive Learning System in China." *Chinese Automation Congress*, 2018, https://ieeexplore.ieee.org/document/8623327/keywords#keywords

[7] W. van der Linden, R. Hambleton (Eds.). *Handbook of Modern Item Response Theory*. New York, NY: Springer-Verlag , 1997.

[8] Frederic M. Lord, "A theory of mental test scores." *Psychometric Monograph*, No. 7, 1952, Psychometric Society

[9] A. Birnbanm, "Probability and statistics in item analysis and classification problems: Efficient design and use of tests of mental ability for various decision-making problems." Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[10] A. Birnbaum, "Probability and statistics in item analysis and classification problems: Further considerations of efficiency in tests of a mental ability." *Tech. Rep. No. 17, Project No. 7755-23*, Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[11] A. Birnbanm, "Probability and statistics in item analysis and classification problems. On the estimation of mental ability." *Series Rep. No. 15, Project No. 7755-523*, Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[12] A. Birnbaum, Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968. Chapters 17‑20.

[13] A. E. Maxwell, "Maximum likelihood estimates of item parameters using the logistic function." *Psychometrika*, 24(3), 221-227. doi:10.1007/bf02289844, 1959

[14] F. Baker and S.-H. Kim, "Item Response Theory: Parameter Estimation Techniques." New York, NY: Marcel Dekker, Inc. (2004).

[15] R. Bock, and M. Aitkin, "Marginal Maximum Likelihood Estimation of Item Parameters: Applications of an EM Algorithm." *Psychometrika*, 46, 443-459, 1981.

[16] J. Johns, S. Mahadevan, B. Woolf, "Estimating Student Proficiency Using an Item Response Theory Model." In: M. Ikeda, K.D. Ashley, TW. Chan (eds) *Intelligent Tutoring Systems*. ITS 2006. Lecture Notes in Computer Science, vol 4053. Springer, Berlin, Heidelberg

[17] S. Kim, T. Moses and H. Yoo, "Effectiveness of item response theory (IRT) proficiency estimation methods under adaptive multistage testing.", *Research Report No. RR-15-11,* 2015b, Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12057

[18] Ok-choon Park, Jung Lee, "Adaptive instructional systems." In D. H. Jonassen & M. Driscoll (Eds.), *Handbook of research for educational communications and technology*, 2nd ed., pp. 651‑684, 2003, Mahwah, New Jersey: Laurence Earlbaum. Retrieved from http://www.aect.org/edtech/ed1/25.pdf