論文 / 著書情報
Article / Book Information

Title	Peak Shift Estimation: A Novel Method to Estimate Ranking of Selectively Omitted Examination Data
Author	Satoshi Takahashi, Masaki Kitazawa, Ryoma Aoki, Atsushi Yoshikawa
Journal/Book name	TALE2021 IEEE International Conference on Engineering Technology, and Education Conference Proceedings, , ,
Pub. date	2021, 12
DOI	https://doi.org/10.1109/TALE52509.2021.9678753
Copyright	(c)2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.
Note	This file is author (final) version.

Peak Shift Estimation: A Novel Method to Estimate Ranking of Selectively Omitted Examination Data

Satoshi Takahashi
College of Science and Engineering
Kanto Gakuin University
Yokohama, Japan
satotaka@kanto-gakuin.ac.jp

Masaki Kitazawa
Graduate School of Artificial Intelligence and Science
Rikkyo University
Tokyo, Japan
Kitazawa Tech
Fujisawa, Japan
masaki.kitazawa@rikkyo.ac.jp

Ryoma Aoki
Department of Computational Intelligence and Systems,
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology
Yokohama, Japan
aoki.r.ae@m.titech.ac.jp

Atsushi Yoshikawa
Department of Computational Intelligence and Systems,
Interdisciplinary Graduate School of Science and Engineering,
Tokyo Institute of Technology
Yokohama, Japan
Graduate School of Artificial Intelligence and Science
Rikkyo University
Tokyo, Japan
at_sushi_bar@dis.titech.ac.jp

***Abstract*—This paper focuses on examination results when examinees selectively skip examinations to compare the difficulty levels of these examinations, e.g., university entrance examinations, certification examinations, and outcomes of students' job-hunting activities. The resultant data is referred to as "selectively omitted examination data." For instance, high school students select and apply to universities in consideration of the difficulty levels of the entrance examinations and their academic abilities. Based on their academic abilities, students skip university entrance examinations that are too challenging or too easy. In that case, the results include selectively omitted data. Existing methods, such as IRT and CTT, cannot estimate the difficulty levels of such examinations because of the omitted data. Understanding the difficulty level of these examinations can facilitate the formulation of a new index to assess organizational ability, number of students who pass, and difficulty of the examinations. This index would reflect the outcomes of the organizations' education, corresponding with perspectives on examinations. Therefore, we propose a novel method, Peak Shift Estimation, to estimate an examination's difficulty level based on selectively omitted examination data. We apply Peak Shift Estimation to the simulation data and demonstrate that it estimates the rank order of the difficulty level of examinations.**

Keywords—university entrance examination, simulation, linking, item response theory, educational organization rankings

I. INTRODUCTION

Many educational organization rankings have been developed to assess educational organizations' environment [1, 2], for example the 100 Best Public High Schools in the U.S. [3], World University Rankings [4], and the Shanghai Jiao Tong Academic Ranking of World Universities [5]. They are based on multiple perspectives: teaching, research, citations, industry income, international outlook, opinion survey, SAT/ACT scores, AP test, IB test, and so on. However, we cannot refer to them when we assess educational organizations from an educational outcome perspective for particular domains, such as results of certification examinations, outcomes of students' job-hunting activities, and entrance examination because examinees skip these examinations selectively.

The purpose of this study is to propose a method to compare the difficulty levels of such examinations, which is referred to as "selectively omitted examination data." Such data comprises the results of certification examinations and outcomes of students' job-hunting activities, and are often archived. Universities usually track students' job-hunting activities and document which company the students join. However, extant literature has not focused on this type of data. Knowledge of the examinations' difficulty level can facilitate the formulation of a new index to assess the organization's standards, pass rates, and examinations' difficulty levels. This index would reflect the outcomes of student education from an examination perspective.

The entrance examination data is a part of the selectively omitted examination data. High school students select universities to which they apply, considering the difficulty levels of entrance examinations and their academic abilities. Based on their academic abilities, students skip university entrance examinations that are too challenging or easy. Usually, high schools only share the number of students accepted by each university; meaning, they do not release the number of unsuccessful students. Consequently, one can collect information on the number of students accepted by different universities from different high schools without knowing their names.

Fig. 1 depicts the selectively omitted examination data. The results are sorted according to the organization's ability levels. Each examination has a peak, that is, a point at which acceptance is the highest. The peak corresponds with an organization's higher ability level when the examination's difficulty level increases. This implies that during the peak, the examination's difficulty levels correspond with the organization's ability levels. When the number of examinees

978-1-6654-3687-8/21/$31.00 ©2021 IEEE

is lower or higher than the number at the peak, the organization's ability level is lower or higher than the examination's difficulty level. In this figure, the organizations and examinations are arranged in the order of their ability and difficulty levels. Their consequent shapes are very regular; therefore, they can be easily sorted. However, the ability and difficulty levels are not known in real life, so they cannot be sorted without ingenuity.

Methods such as Classical Test Theory (CTT) [6] and Item Response Theory (IRT) [7] have been used to compare the examinations' difficulty levels or their items. IRT suggests that the relationship between the accuracy rate of each item and a person's ability level can be expressed as a function. Therefore, different examinations' difficulty levels can be estimated based on the function.

Many studies have connected different tests using these methods, which is referred to as "linking" [8,9,10,11,12]. Liu and Walker [8] demonstrated a connection between the National Assessment of Educational Progress [13], the International Assessment of Educational Progress [14], the Armed Services Vocational Aptitude Battery [15], and the North Carolina End-of-Grade Tests [16]. Kolen and Brennan [9] linked American College Testing [17] and the Iowa Tests of Educational Development [18].

One of the prerequisites of using IRT and CTT is that examinees are challenged to solve all examination items; therefore, they require data on who can or cannot solve what items. However, the data used in this study is not suitable for IRT and CTT because it includes examinees who skip examinations selectively. Further, information on the examinations that are skipped and why examinees skip them remains veiled. Thus, named after the shapes obtained in Fig. 1, we propose a novel method, that is, Peak Shift Estimation.

First, we introduced the algorithm for Peak Shift Estimation. Then, we apply it to the simulation data and demonstrate that it estimates the rank order of the examinations' difficulty levels.

II. Peak Shift Estimation

Peak Shift Estimation step by step estimates examination ranks and organizations around their peak from top-rank examinations based on the acceptance number. The Peak Shift Estimation algorithm is mentioned below. For readability, examinations are referred to as university entrance examinations and organizations as high schools. The total number of high schools is designated by m, the total number of universities by n, the loop count by i, a set of universities by U_i, and a set of high schools by H_i. The application of Peak Shift Estimation requires information on the top universities according to the difficulty level ranking. These are designated by U_1 in Step (0); Tokyo University corresponds to U_1 in Japan.

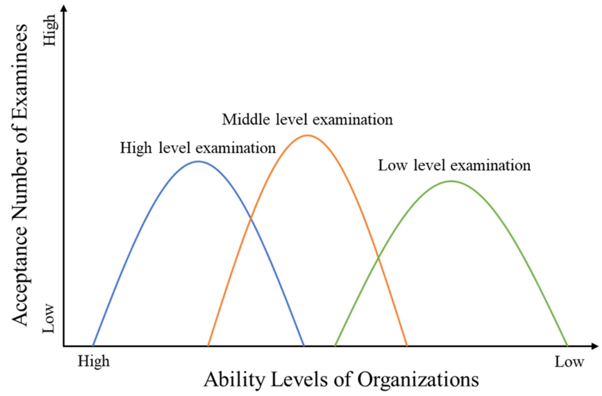

Fig. 1. Selectively omitted examination data.

First, Peak Shift Estimation standardizes input data by considering the difference between the total number of students in high schools and the total acceptance number between universities. Table I depicts the procedure. Then, Peak Shift Estimation estimates the university entrance examination ranks and Table II depicts the procedure.

Peak Shift Estimation determines H_i at Estimating steps (1) and (2). This corresponds with determining high schools around the U_i's peak of the standardized acceptance number of examinees (Fig. 2). The size of H_i is relative to the size of U_i.

The tendency of accepted students from $\bigcup_1^i H_j$ would be similar at U_i and universities that have difficulty levels close to U_i. Hence, Peak Shift Estimation clusters universities based on the acceptance rate of $\bigcup_1^i H_j$ and defines the top university cluster as U_{i+1} (Fig. 3) at Estimating steps (3), (4), and (5).

Peak Shift Estimation uses X-means, and the temporary difficulty level ranking is stochastically variable. The temporary difficulty level ranking was calculated more than once, and then their average was used. Then, the average was sorted and defined as estimating the difficulty level ranking.

TABLE I. Standardizing Step

Step	Processing
Step (1)	Divide each acceptance number by the corresponding total number of high school students.
Step (2)	Divide each number from step (1) by the total number of universities.

TABLE II. Estimating Step

Step	Processing
Step (0)	Rank U_1 number one of temporary difficulty level ranking.
Loop steps (1)–(6).	
Step (1)	Sort high schools according to the summation of acceptance rate of U_i in descending order.
Step (2)	Define the top $\left\lfloor \frac{\lvert U_i \rvert}{n} \times m \right\rfloor$ high schools from step (1) ranking as H_i.
Step (3)	Cluster universities that are not ranked based on the acceptance rate of $\bigcup_1^i H_j$, with X-means [19].
Step (4)	Sort the university clusters by an average acceptance rate of $\bigcup_1^i H_j$ in descending order.
Step (5)	Define the top university cluster from step (4) ranking as U_{i+1} and rank U_{i+1} as $i+1$ of the temporary difficulty level ranking.
Step (6)	Stop the algorithm after all the universities are ranked.

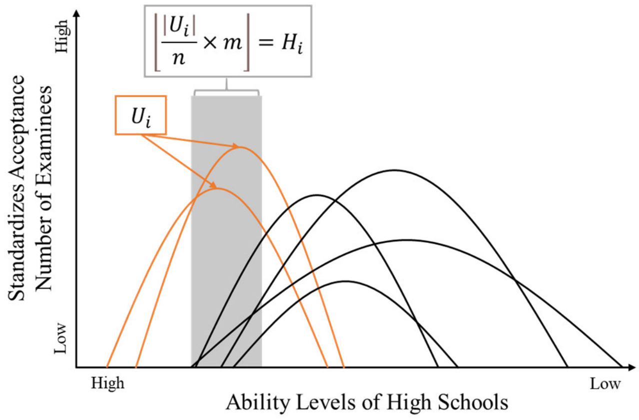

Fig. 2. Estimating steps (1) and (2).

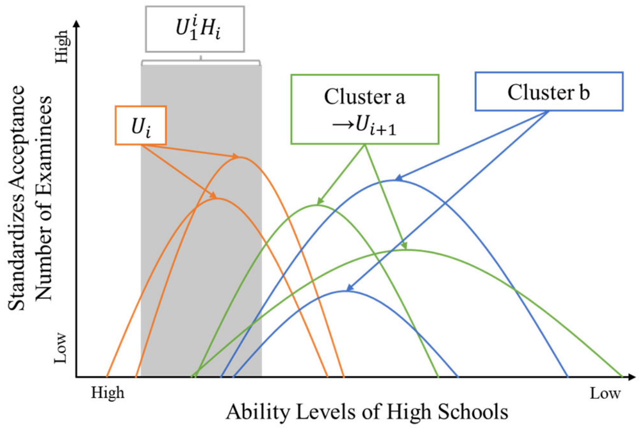

Fig. 3. Estimating steps (3), (4), and (5).

TABLE III. HIGH SCHOOL DATA STEP

Step	Processing
Step (1)	Generate a number of high school students and allocate students' academic ability following a normal distribution.
Step (2)	Generate a number of high schools and allocate their average academic ability following a normal distribution.
Step (3)	Considering the number of students in each high school, assign them stochastically to high schools whose average academic ability is close to students' academic ability.
Step (4)	The standard deviation of students' academic abilities for each high school was calculated. When at least one high school's standard deviation exceeds the limitation, go back to Step (3).
Step (5)	Rank the high schools based on the average of their students' academic abilities.

TABLE IV. UNIVERSITY DATA STEP

Step	Processing
Step (1)	Generate universities and allocate examinations' difficulty levels to them following a normal distribution. Each university is allocated one examination.
Step (2)	Each high school student chooses universities to take entrance examinations, considering that the examinations' difficulty levels are not over or under the student's academic ability ± limitation of entrance examination ability. Each high school student chooses a number of universities and thus, becomes a candidate for the chosen universities.
Step (3)	In the order of the examinations' difficulty levels, each university accepts candidates up to the number of university students according to the candidates' academic ability. Universities cannot accept students who have already been accepted by other universities. Students become "accepted students" at universities that enroll them.
Step (4)	Each university considers candidates who have higher academic ability as acceptance students rather than entrance students with the lowest academic ability

III. SIMULATION DATA

A. Generating Data

Procedure of generating data: the procedure of high school data generation is described in table III and the procedure for generating university data is described in table IV.

In real life, students take and pass several university examinations and then choose to enroll in one university. Using previously collected data, universities predict how many students would refuse to accept and then accept students more than the number of pre-planned entrance students. Even if universities cannot gather enough students, they accept additional students to match their planned number. Consequently, the number of students who chose the university became close to the number of entrance students planned by the university. University data steps (2), (3), and (4) express the results of this process.

1) Parameters of generating data: The simulation parameters are listed in Table V. The relational expression between σ of students' academic ability in each high school and σ of high schools' average academic ability can be determined based on other parameters.

Each high school is defined as i, number of high schools as k, each student in each high school as j, number of students at each high school as m, distribution of students' academic ability as $N(0,1^2)$, distribution of high schools' average academic ability as $N(0,\sigma_a^2)$, distribution of students' academic ability in each high school as $N(\mu_{ei},\sigma_e^2)$, a_i as the constant for each high school, $\mu_{ei} = a_i\sigma_a$, $b_{i,j}$ as constants for each student, and each student's academic ability as $x_{i,j} = a_i\sigma_a + b_{i,j}\sigma_e$.

From $N(0,1^2)$: distribution of high school students' academic ability,

$$0 = \frac{1}{km}\sum_i^k \sum_j^m \left(a_i\sigma_a + b_{i,j}\sigma_e\right) \text{ (1).}$$

$$1^2 = \frac{1}{km}\sum_i^k \sum_j^m \left(a_i\sigma_a + b_{i,j}\sigma_e\right)^2 \text{ (2).}$$

From $N(0,\sigma_a^2)$: distribution of high schools' average academic ability,

$$0 = \frac{1}{k}\sum_i^k (a_i\sigma_a) \quad \text{(3)},\ \sigma_a^2 = \frac{1}{k}\sum_i^k (a_i\sigma_a)^2 \text{ (4).}$$

From (3) and (4),

$$\sum_i^k a_i = 0 \text{ (5)},\ \sum_i^k a_i^2 = k \text{ (6).}$$

From $N(a_i\sigma_a,\sigma_e^2)$: distribution of students' academic ability in each high school,

$$a_i\sigma_a = \frac{1}{m}\sum_j^m \left(a_i\sigma_a + b_{i,j}\sigma_e\right) \text{ (7)},\ \sigma_e^2 = \frac{1}{m}\sum_j^m \left(b_{i,j}\sigma_e\right)^2 \text{ (8).}$$

From (7) and (8),

$$\sum_j^m b_{i,j} = 0 \text{ (9)},\ \sum_j^m b_{i,j}^2 = m \text{ (10).}$$

From (2), (5), (6), (9), and (10),

$$1^2 = \sigma_a^2 + \sigma_e^2 \quad \text{(11).}$$

Then, σ_a and σ_e can be expressed by θ as shown below.

$$\sigma_a = \sin\theta \quad \text{(12)},\ \sigma_e = \cos\theta \quad \text{(13).}$$

TABLE V. SIMULATION SETTINGS

Item	Simulation Pattern 1	Simulation Pattern 2
μ of students' academic ability	0	
σ of students' academic ability	1	
Number of high schools	1,000	
Number of students at each high school	100	
μ of high schools' average academic ability	0	
θ	$\frac{1}{12}\pi, \frac{2}{12}\pi, \frac{3}{12}\pi, \frac{4}{12}\pi, \frac{5}{12}\pi$	
σ of high schools' average academic ability	$\sin\theta$	
σ of students' academic ability in each high school	$\cos\theta$	
Limitation of σ of students' academic ability in each high school	1.96 x 0.20	
Number of universities	100	
Number of entrance students in each university	500	100
μ of the difficulty levels of entrance examinations	0	
σ of the difficulty levels of entrance examinations	1	
Limitation of entrance examination ability	1	

B. Results

Data were generated by conducting 1,000 simulations (Table V). The number of entrance students in each university was set to 500 and 100, and the σ of high schools' average academic ability was changed from $\sin 1/12\pi$ to $\sin 5/12\pi$ to evaluate their effect. The results are presented in Figs. 4, 5, 6, 7, 8, and 9.

Figs. 4 and 7 depict the candidates heatmaps between universities and high schools. The candidates are distributed between the two arcs and the distance between them decreases when the σ of high schools' average academic ability increases. The σ of high schools' average academic ability is not related to the high school students' choice of taking entrance examinations. Figs. 4 and 7 are the same.

Figs. 5 and 8 depict the entrance students' heatmap between universities and high schools. The students of top- and bottom-ranked universities are concentrated in top- and bottom-ranked high schools. The students of middle-ranked universities are distributed between the two arcs. They choose to enroll in the highest-ranked university out of those whose examination they have cleared. The distance between the arcs decreases when the σ of high schools' average academic ability increases, as is the case for candidates. In addition, the entrance students were widely scattered when their number in each university was large.

The standard deviations of high school students' academic abilities and examinations' difficulty levels are the same; however, the number of students is far greater than the number of universities. Thus, high school students' academic abilities are more widely distributed than examinations' difficulty levels. Therefore, some of the highest-ranked high school students have no choice but to enroll in top-ranked universities, even if their difficulty levels are lower than the student's academic abilities. The students who enroll in top-ranked universities are concentrated in top-ranked high schools. High school students enroll in a university according to the difficulty levels. As the lowest-ranked universities lose their candidates, the arcs suddenly rise.

Figs. 6 and 9 depict the heatmaps of accepted students between universities and high schools. The students who enroll in top-ranked universities are concentrated in top-ranked high schools, respectively. Students at middle-rank universities are distributed between the two arcs. The distance between them decreases when the σ of high schools' average academic ability increases, as is the case for candidates. The relative degree of concentration around high-ranked high schools and universities increased.

Each university considers candidates with higher academic ability as accepted students compared to the entrance students with lower academic ability. Accordingly, the students who are between the two arcs in Figs. 5 and 8 and above the lower arc in Figs. 6 and 9 become accepted students.

IV. ESTIMATION

Peak Shift Estimation was applied to the simulation data, the temporary difficulty level ranking was calculated 1,000 times, and an average of the values was obtained. The average was sorted and defined as the difficulty level ranking. The simulation data included the results of 1,000 patterns; therefore, the difficulty level ranking was obtained for 1,000 patterns. It is a prerequisite for Peak Shift Estimation to know the top-level universities; then, the universities are ranked according to the entrance examinations' difficulty levels.

In addition, IRT of one-parameter and two-parameter logistic models with marginal maximum likelihood were applied. IRT requires data about which student passed or failed each item. Consequently, IRT could not be applied to "selectively omitted examination data" without ingenuity. Hence, the data were processed as follows:

i. Each high school generates students and assigns academic ability ranking.

ii. Each university assesses students based on the number of accepted students from each university, considering their academic ability ranking.

Tables VI and VII present examples of this process. High school A generates students a, b, c, d, and e and assigns academic ability ranking in an alphabetical order. University A determines the top three students (a, b, and c) as accepted.

TABLE VI. ACCEPTED STUDENTS OF HIGH SCHOOL A.

High School	Number of Students	University A	University B	University C
High School A	5	3	2	1

TABLE VII. THE PROCESSED DATA OF HIGH SCHOOL A.

Student	University A	University B	University C
Student a	succeeded	succeeded	succeeded
Student b	succeeded	succeeded	failed
Student c	succeeded	failed	failed
Student d	failed	failed	failed
Student e	failed	failed	failed

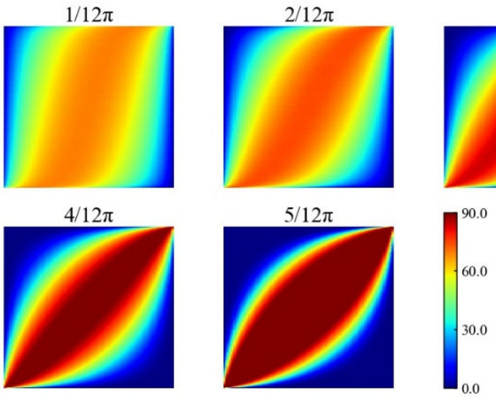

Fig. 4. Candidates' Heatmap of Simulation Pattern 1, X-axis is the level of entrance examination and Y-axis is the level of high school; the bottom left high school and entrance examinations are top-rank.

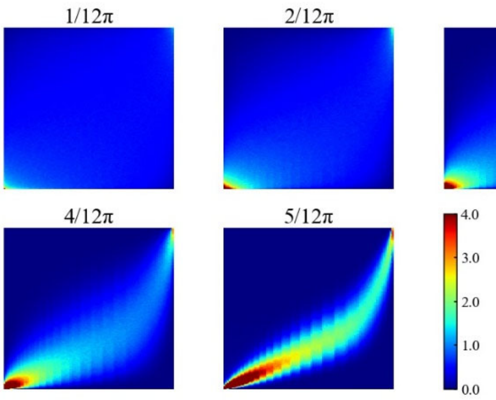

Fig. 5. Entrance Students' Heatmap of Simulation Pattern 1, X-axis is the level of entrance examination and Y-axis is the level of high school; the bottom left high school and entrance examinations are top-rank.

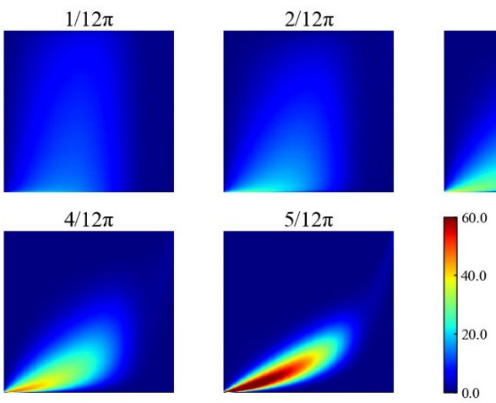

Fig. 6. Accepted Students' Heatmap of Simulation Pattern 1, X-axis is the level of entrance examination and Y-axis is the level of high school; the bottom left high school and entrance examinations are top-rank.

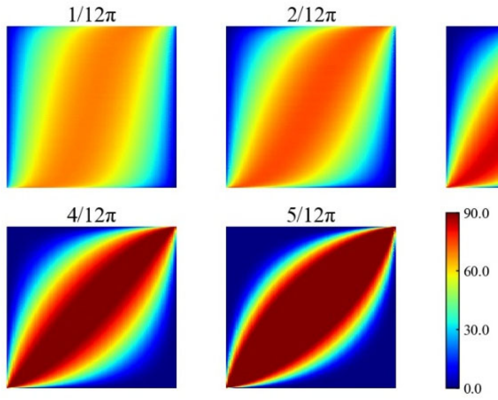

Fig. 7. Candidate Students' Heatmap of Simulation Pattern 2, X-axis is the level of entrance examination and Y-axis is the level of high school; the bottom left high school and entrance examinations are top-rank.

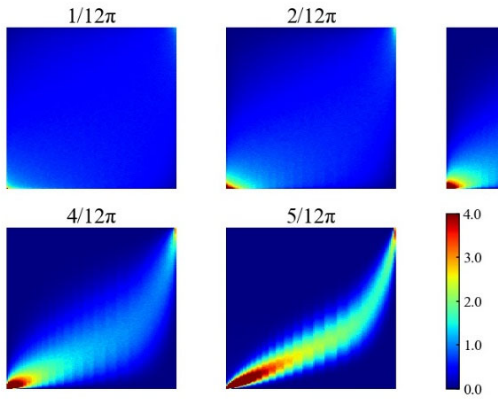

Fig. 8. Entrance Students' Heatmap of Simulation Pattern 2, X-axis is the level of entrance examination and Y-axis is the level of high school; the bottom left high school and entrance examinations are top-rank.

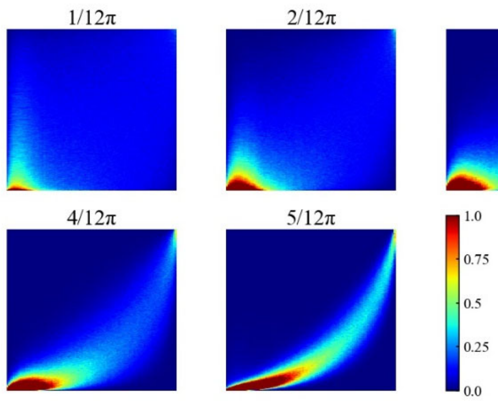

Fig. 9. Accepted Students' Heatmap of Simulation Pattern 2, X-axis is the level of entrance examination and Y-axis is the level of high school; the bottom left high school and entrance examinations are top-rank.

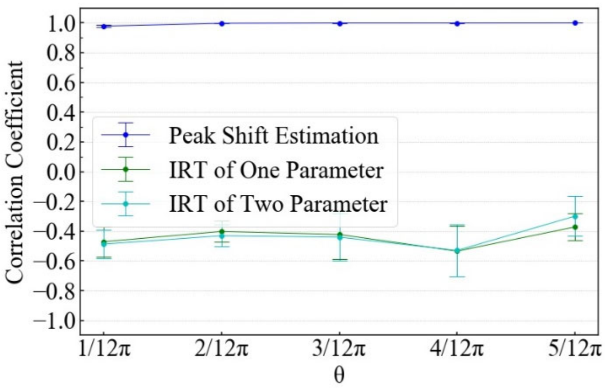

Fig. 10. Spearman's Rank Correlation Coefficient in Pattern 1.

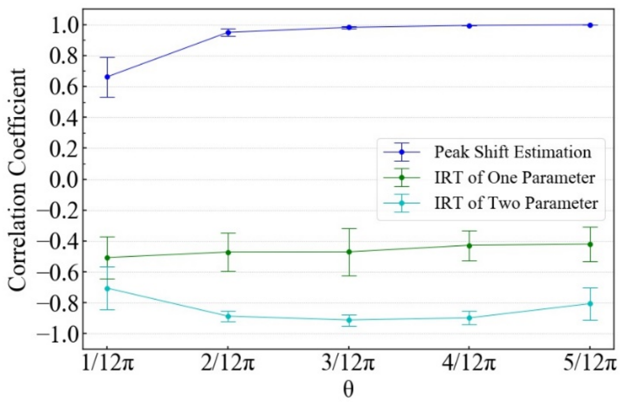

Fig. 11. Spearman's Rank Correlation Coefficient in Pattern 2.

A. Spearman's Rank Correlation Coefficient

Figs. 10 and 11 depict the Spearman's rank correlation coefficient between the difficulty level ranking and true ranking. The rank correlation coefficients of Peak Shift Estimation are high and stable, except in the case of $\theta = 1/12\pi$ in pattern 1. On the other hand, the rank correlation coefficients of IRT depicted negative values.

B. Peak Shift Estimation

1) $\theta = 5/12\,\pi$ in Pattern 1: The case of $\theta = 5/12\pi$ in pattern 1 is selected as a stable example. Figs. 12 and 13 depict the difficulty level ranking and true ranking, one simulation data, and results of Peak Shift Estimation conducted 1,000 times. Fig. 13 depicts the estimation process of one simulation data and results of Peak Shift Estimation conducted only once. In Fig 14, polygonal lines represent the number of accepted students for each university from each high school. The lines are color-coded according to the clustering group, wherein the red line represents the top university cluster (U_{i+1}) and the markers are $\bigcup_1^i H_j$, according to Estimating Steps (3) and (4).

The top true-ranking universities are clustered into small size groups from Fig. 13; however, the middle and low true-ranking universities are clustered into the middle-size group. In addition, the mean of 1,000 estimated rankings is ordered according to the real ranking from Fig. 12.

Peak Shift Estimation succeeded in extracting $\bigcup_1^i H_j$ from the high schools with the highest academic abilities and small-size group of the top university (U_{i+1}) from Fig. 14.

2) $\theta = 5/12\,\pi$ in Pattern 2: The case of $\theta = 5/12\pi$ in pattern 2 was selected as an example of lower acceptance rate. Figs. 15 and 16 depict the difficulty level ranking and true ranking, one simulation data, and results of Peak Shift Estimation conducted 1,000 times. Fig. 17 depicts the estimation process of one simulation data, and Peak Shift Estimation conducted only once.

From Fig. 16, the universities are clustered into larger groups compared to $\theta = 5/12\pi$ in pattern 2 (Fig. 13). In addition, the mean of 1,000 estimated rankings is ordered according to the real ranking from Fig. 15.

The gap in the number of accepted students between high schools is small. This caused $\bigcup_1^i H_j$ to include high schools with lower academic ability, as shown in Fig. 17. Similarly, the gap in the number of accepted students between universities is also small, which produces large-clustered groups.

3) $\theta = 1/12\,\pi$ in Pattern 2: The case of $\theta = 1/12\pi$ in pattern 2 is selected as an unstable example. Figs. 18 and 19 depict the difficulty level ranking and true ranking, one simulation data, and results of Peak Shift Estimation conducted 1,000 times. Fig. 20 depicts the estimation process of one simulation data and results of Peak Shift Estimation conducted only once.

The universities are clustered into larger groups and sometimes, U_{i+1} includes universities with lower true ranking than U_i from Fig. 19. However, the means of estimated ranking conducted 1,000 times and the real ranking are positively correlated from Fig. 18.

As shown in Fig. 20, the gaps in the number of accepted students between high schools and universities are very small, which produced large-clustered groups.

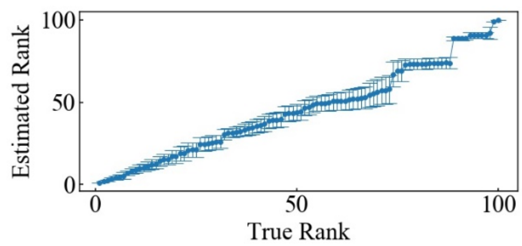

Fig. 12. θ=5/12 π in pattern 1; difficulty level ranking and true ranking, one simulation data, and Peak Shift Estimation conducted 1,000 times; error bar depicts standard deviation.

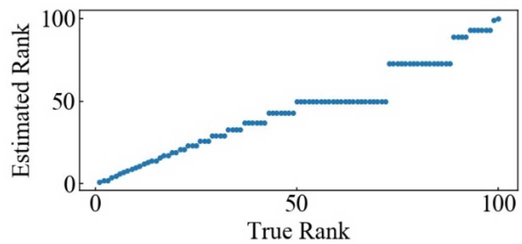

Fig. 13. θ=5/12 π in pattern 1; difficulty level ranking and true ranking, one simulation data, and results of Peak Shift Estimation conducted only once.

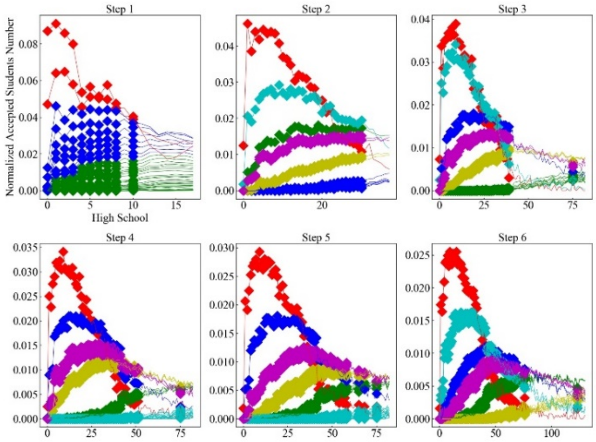

Fig. 14. θ=5/12 π in pattern 1; the estimation process of one simulation data, and results of Peak Shift Estimation conducted only once, the red line represents the top university cluster (U_{i+1}).

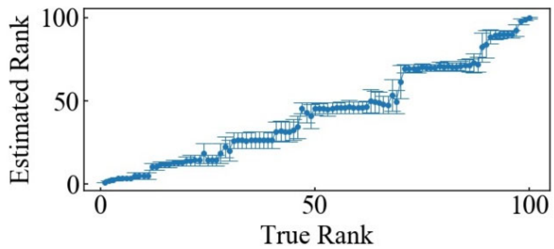

Fig. 15. θ=5/12 π in pattern 2; difficulty level ranking and true ranking, one simulation data, and Peak Shift Estimation conducted 1,000 times; error bar depicts standard deviation.

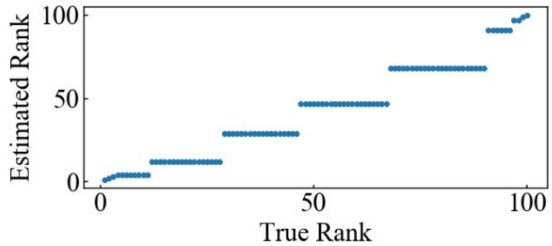

Fig. 16. θ=5/12 π in pattern 2; difficulty level ranking and true ranking, one simulation data, and results of Peak Shift Estimation conducted only once.

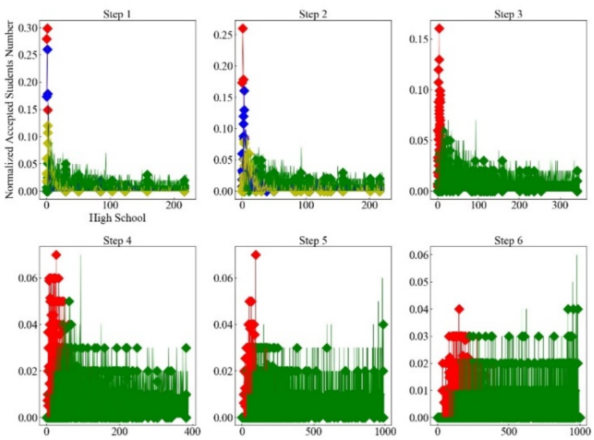

Fig. 17. θ=5/12 π in pattern 2; the estimation process of one simulation data and results of Peak Shift Estimation conducted only once, the red line represents the top university cluster (U_{i+1})

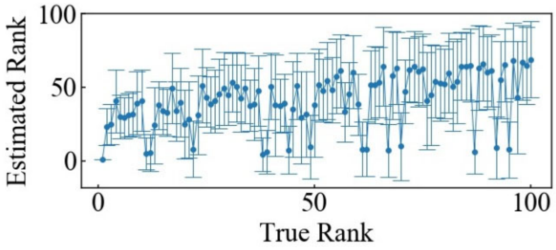

Fig. 18. θ=1/12 π in pattern 1; difficulty level ranking and true ranking, one simulation data, and Peak Shift Estimation conducted 1,000 times; error bar depicts standard deviation.

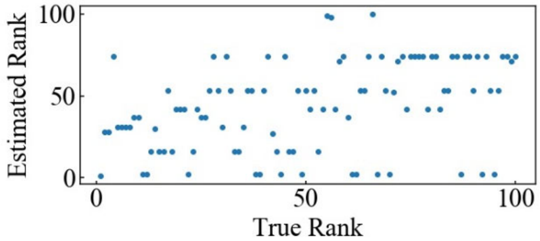

Fig. 19. θ=1/12 π in pattern 2; difficulty level ranking and true ranking, one simulation data, and one-time peak shift estimation.

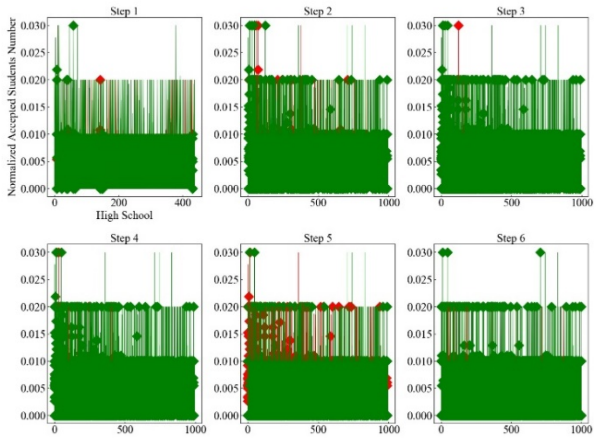

Fig. 20. θ=1/12 π in pattern 2, the estimation process of one simulation data and results of Peak Shift Estimation conducted only once, the red line represents the top university cluster (U_{i+1})

C. Item Response Theory

Figs. 21 and 22 depict the results of IRT of the two-parameter logistic models with marginal maximum likelihood. Figs. 21(a) and 22(a) depict the standardized number of accepted students and true ranking. Figs. 21(b) and 22(b) depict the difficulty level ranking and true ranking.

The middle-ranking universities accept a larger number of students, and their difficulty level ranking is estimated to be low (Fig. 21). The examinations' difficulty levels and high schools' average academic ability are allocated following a normal distribution; therefore, the difficulty levels of middle-ranking universities and academic ability of students of middle-ranking high schools are close. Consequently, a lot of students choose middle-ranking universities and get accepted. IRT estimates that universities that accept many students have a low difficulty level.

The peak of the standardized number of accepted students moves toward the left (Fig. 22). The number of entrance

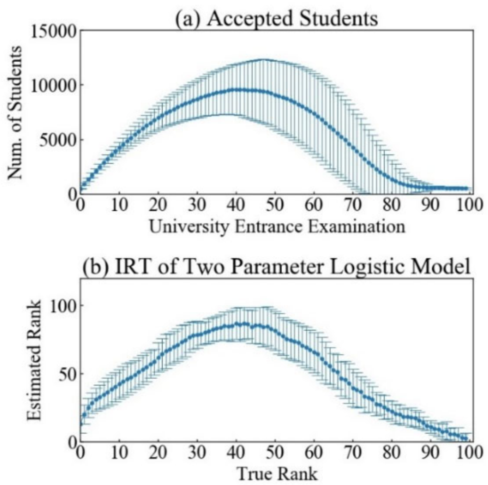

Fig. 21. θ=5/12 π in pattern 1. (a) number of accepted students, (b) Difficulty-level ranking and true ranking; error bar depicts standard deviation.

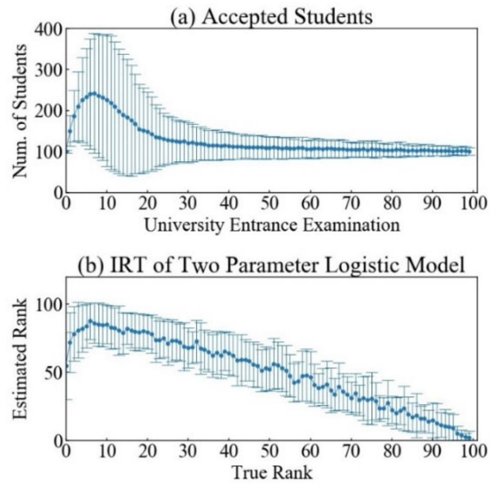

Fig. 22. θ=5/12 π in pattern 2. (a) number of accepted students, (b) Difficulty-level ranking and true ranking; error bar depicts standard deviation.

students in each university was low, and students with low academic ability could not pass the examination. In this case the peak of the difficulty level ranking also moved toward the left. Consequently, the rank correlation coefficients of IRT embody negative values.

V. Conclusion

This paper proposes Peak Shift Estimation as a novel method to estimate an examination's difficulty level based on "selectively omitted examination data." It also verified the accuracy of Peak Shift Estimation using the simulation data. This estimated difficulty level of the examination reflects the outcomes of the organization's education, corresponding with perspectives on examinations, and the Peak Shift Estimation can be applied to university entrance examinations, certification examinations, and outcomes of students' job-hunting activities.

However, this study is not without its limitations. The application of Peak Shift Estimation requires information on the top examinations as a precondition and in certain cases, for example, new category certification examinations, this information is not known. The accuracy and robustness of Peak Shift Estimation had to be evaluated under various situations; for example, the different number of entrance students between universities, different number of students between high schools, and biased distribution of high schools' average academic ability or distribution of students' academic ability in each high school. Therefore, future research should focus on applying Peak Shift Estimation to real data and verify its applicability.

References

[1] P. Taylor and R. Braddock, "International university ranking systems and the idea of university excellence," Journal of Higher Education Policy and Management, vol. 29, issue 3, pp. 245–260, 2007.

[2] R. Lukman, D. Krajnc and P. Glavič, "University ranking using research, educational and environmental indicators," Journal of Cleaner Production, vol. 18, issue. 7, pp. 619–628, 2010.

[3] The Best Schools. The 100 Best Public High Schools in the U.S. https://thebestschools.org/rankings/best-public-high-schools-us/ (accessed February 13, 2021).

[4] Times Higher Education. World University Rankings. https://www.timeshighereducation.com/content/world-university-rankings (accessed February 13, 2021).

[5] N. C. Liu and Y. Cheng, "The academic ranking of world universities," Higher education in Europe, vol. 30, issue 2, 127–136, 2005.

[6] M. R. Novick, "The axioms and principal results of classical test theory," Journal of Mathematical Psychology, vol. 4, issue 1, pp. 1–18, 1966. DOI=https://doi.org/10.1016/0022-2496(66)90002-2

[7] R. K. Hambleton, H. Swaminathan and H. J. Rogers, Fundamentals of Item Response Theory. Newbury Park, CA: Sage Press, 1991.

[8] J. Liu and M. E. Walker, "Score linking issues related to test content changes," in Linking and Aligning Scores and Scales, N. J. Dorans, M. Pommerich, and P. W. Holland, Eds. New York, NY: Springer, 2007, pp. 109–134. https://doi.org/10.1007/978-0-387-49771-6_7

[9] M. J. Kolen and R. L. Brennan, Test Equating, Scaling, and Linking: Methods and Practices. Germany: Springer Science & Business Media, Berlin, 2014.

[10] M. Orlando, C. D. Sherbourne and D. Thissen, "Summed-score linking using item response theory: application to depression measurement," Psychological Assessment, vol. 12, issue 3, pp. 354–359, 2000. DOI=10.1037//1040-3590.12.3.354

[11] S. G. Sireci, "Problems and issues in linking assessments across languages," Educational Measurement: Issues and Practice, vol. 16, issue 1, pp. 12–19, 1997. DOI=https://doi.org/10.1111/j.1745-3992.1997.tb00581.x

[12] S. H. Kim and A. S. Cohen, "A comparison of linking and concurrent calibration under item response theory," Applied Psychological Measurement, vol. 22, issue 2, pp. 131–143, 1998. DOI=https://doi.org/10.1177/01466216980222003

[13] Institute of Education Sciences. National Assessment of Educational Progress. https://nces.ed.gov/nationsreportcard/ (accessed February 13, 2021).

[14] National Research Council, "International assessment of educational progress," in *International Comparative Studies in Education: Descriptions of Selected Large-Scale Assessments and Case Studies*, Washington, DC: The National Academies Press, 1995, pp. 48–57.

[15] Military Advantage. The ASVAB Test. https://www.military.com/join-armed-forces/asvab (accessed February 13, 2021).

[16] North Carolina Department of Public Instruction. End-of-Grade (EOG). https://www.dpi.nc.gov/districts-schools/testing-and-school-accountability/state-tests/end-grade-eog (accessed February 13, 2021).

[17] Act, Inc. http://www.act.org (accessed February 13, 2021).

[18] The University of Iowa. Iowa Testing Programs. https://education.uiowa.edu/centers/iowa-testing-programs (accessed February 13, 2021).

[19] D. Pelleg and A. Moore, "X-means: Extending K-means with efficient estimation of the number of clusters," in Proceedings of the Seventeenth International Conference on Machine Learning (Stanford, CA, June 29-July 2, 2000). ICML'00. San Francisco, CA: Morgan Kaufmann Publishers Inc., 2000, pp. 727–734. DOI=https://dl.acm.org/doi/10.5555/645529.6578