A Novel Method Considering Guesses and Slips for Item Ordering: Noise-Filter Ordering Theory and Item Relational Structure Theory

Li-Wen-Yuan Zhou, Sheng-Hong Dong*, Xing-Gao Li

School of Psychology, Key Laboratory of Psychology and Cognition Science of Jiangxi, Jiangxi Normal University, Nanchang, Jiangxi, China.

* Corresponding author. Email: shdong@jxnu.edu.cn
Manuscript submitted March 23, 2022; accepted June 1, 2022.
doi: 10.17706/jsw.17.4.149-157

Abstract: Bart's ordering theory algorithm and Takeya's item relational structure algorithm are the two classical ordering algorithms, but both of them have a common problem, that is, the item ordering results depend on the setting of threshold. The reason for this problem is that the ordering index fluctuates greatly due to guessing and slipping noise. This paper presents an improved algorithm, which can effectively filter out the noise of guessing and slipping. It is more effective than the two traditional methods. This paper also develops a computer program and carries out simulation experiments to prove the advantages of this method.

Key words: ordering theory algorithm, item relational structure algorithm, item ordering index, DINA model, noise-filter ordering theory

1. Introduction

Ordering analysis of test items can clarify the relationship between items, which is not only helpful to obtain the instructional hierarchy structure, but also helpful to the cognitive diagnosis of students [1], [2]. At present, the common ordering analysis methods are mainly divided into two types: various methods extended by Bart's ordering theory (OT) and various methods extended by Takeya's item related structure theory (IRST) [1], [3]. No matter which method, the threshold parameter (tolerance level) have a great impact on the judgment of the relationship between project prerequisites [4], [5]. Bart found that the fluctuation of threshold value seriously affects the result of item ordering. Liu also proposed the method of setting the threshold for times [2], [6].

In fact, the reason why the ordering index in OT and IRST algorithm is unstable, is there are guesses and slips in student's responses. According to a research of Liu in 2013, without guesses and slips, it can perfectly reproduce the original order of items [7]. Considering guessing and slipping, and because of OT's natural defects, this paper presents a method to consider guessing and slipping to improve the accuracy of OT & IRST method.

2. OT Algorithm

2.1. OT Ordering Theory

OT algorithm provided by Airsian and Bart in 1973[1]. In a binary scoring test，if the participant answers a certain item correctly, it will be recorded as 1, and if the answer is incorrect, it will be recorded as 0. The response modes of item *i* and item *j* are (00), (01), (10) and (11).

Definition 2.1: OT algorithm

1) if (01) mode does not occur, then item *i* is a prerequisite condition for item j, denote as $X_i \rightarrow X_j$, definite this mode as disconfirmatory.
2) if (i) is false, then item i is NOT a prerequisite condition for item j, denote as $X_i \not\rightarrow X_j$.
3) if $X_i \rightarrow X_j$, and $X_j \rightarrow X_i$, then item *i* is equivalent to item j, denote as $X_i \leftrightarrow X_j$.
4) if $X_i \not\rightarrow X_j$, or $X_j \not\rightarrow X_i$, then item *i* is NOT equivalent to item j, denote as $X_i \not\leftrightarrow X_j$。

As there are guesses and slips in item response, Bart & Airsian proposed that it is necessary to set a tolerance level in determining whether the (01) response mode occurs. When the occurrence probability of the response mode is below the tolerance level, it is considered that the response mode does not occur.

Setting an appropriate tolerance level has been a problem repeatedly discussed by Bart & airsian. The tolerance levels they have used include 0.02, 0.03, 0.05, etc [2], [8], [9].

2.2. Liu's Ordering Theory, LOT

Liu theorized sorting into a more mathematical language, and pointed out that the algorithm of sorting theory does not take into account the independence of correctly answered item i and incorrectly answered item j. if these two events are independent of each other, it can be directly considered that item I is not a prerequisite for item J [10]. In this research, Liu et al. also gave examples of independent responses to two items, and using OT would mistakenly believe that there is a prerequisite relationship between the two items. Based on the above reasons, Liu et al. Proposed an improved sorting theory, called Liu's ordering theory, with the following algorithm:

Definition 2.2: LOT algorithm

Define the response vector as $X_T = (X_1, X_2, \ldots, X_n)$, $X_i = 0$ represents wrong answer to item i, $X_j = 1$ represents correct answer to item j, Then the joint response probability is $P(X_i = 0, X_j = 1)$. The following indicators can be calculated:

$$\gamma_{ij}^{(LOT)} = \begin{cases} 0 & if\ Ind = 1 \\ 1 - P(X_i = 0, X_j = 1) & if\ Ind \neq 1 \end{cases}, \tag{1}$$

where $Ind = \dfrac{P(X_i = 0, X_j = 1)}{P(X_i = 0)P(X_j = 1)}$.

Denote ε as the tolerance level, then:

(i) if $\gamma_{ij}^{(LOT)} > 1 - \varepsilon$, then item *i* is a prerequisite condition for item j, denote as $X_i \rightarrow X_j$, otherwise, item i is NOT a prerequisite condition for item *j*, denote as $X_i \not\rightarrow X_j$.

(ii)if $X_i \rightarrow X_j$ and $X_j \rightarrow X_i$, then item *i* is equivalent to item j, denote as $X_i \leftrightarrow X_j$.

For the value of ε, Liu takes ε=0.03 in LOT subjectively. He supposed that the value is basically between 0.02 and 0.04.

2.3. Improved Liu's Ordering Theory, ILOT

Liu believes that the subjective acquisition of tolerance level value lacks statistical characteristics, so he improves the tolerance level on the basis of OT algorithm and obtains the critical value from the distribution of as the tolerance level value [10]. Liu's ILOT with the following algorithm:

$\gamma_{ij}^{(LOT)}$ is the same as in LOT, let $\gamma_{ij}^{(ILOT)} = \gamma_{ij}^{(LOT)}$, then:

$$\gamma_c^{(ILOT)} = \arg_x[1 - \int_{-\infty}^{x} f(\gamma_{ij}^{(ILOT)}) d\gamma_{ij}^{(ILOT)} = 0.05], \tag{2}$$

where *a* is the probability density function of the random variable, if the test length is n, You can first calculate the $\gamma_{ij}^{(ILOT)}$ of any two items in the test, the empirical distribution or empirical pdf of $\gamma_{ij}^{(ILOT)}$ can be obtained from the $n(n-1)$ $\gamma_{ij}^{(ILOT)}$s.

Then, taking the correlation coefficient as 0 as the null hypothesis, the t-test of $\gamma_c^{(ILOT)}$ is carried out. If the original hypothesis is rejected, it can be used as an effective critical value. If $\gamma_{ij}^{(ILOT)} > \gamma_c^{(ILOT)}$, then item *i* is a prerequisite condition for item *j*, denote as $X_i \to X_j$, otherwise, item *i* is NOT a prerequisite condition for item *j*, denote as $X_i \not\to X_j$.

3. IRS Algorithm

Japanese researcher Takeya pointed out that there is an irrationality in OT algorithm, that is, follow the rule "joint probability must be less than the marginal probability", if $P(X_i = 0) \le \varepsilon$ or $P(X_j = 1) \le \varepsilon$, there must be, which will lead to item *i* becoming a prerequisite for all items in the whole test, or all items in the whole test becoming a prerequisite for item j [3].

Although $P(X_i = 0) \le \varepsilon$ stands for the low difficulty of item *i* and $P(X_j = 1) \le \varepsilon$ stands for the high difficulty of item *j*, this is still irrational: imagine that when $P(X_j = 1) \le \varepsilon$, and there is a item k is more difficult than item *j*, in this case, through OT we will get the equivalence of two items. Based on the above reason, Takeya proposed a new ordering analysis method and named it item relational structure theory.

3.1. Item Relational Structure Theory, IRST

Takeya in 1980 proposed a new definition of preconditions.

Definition 3.1: IRST ordering relationship

There exists an ordering relationship $X_i \to X_j$, if and only if,

$$P(X_i = 0, X_j = 1) \le \mu P(X_i = 0) P(X_j = 1), \tag{3}$$

otherwise $X_i \not\to X_j$. The meaning of μ is just like the tolerance level in OT, is a constant, $0 < \mu < 1$, it's usually taken as 0.5.

Follow definition 2, we can get the relationship of items:

1) if and only if $X_i \to X_j$ and $X_j \not\to X_i$, exist a prerequisite relationship $X_i \Rightarrow X_j$, means item i is the prerequisition condition of item j.
2) if and only if $X_i \to X_j$ and $X_j \to X_i$, exist an equivalence relationship $X_i \Leftrightarrow X_j$, means item *i* and item j are equivalence.
3) if and only if $X_i \not\to X_j$ and $X_j \not\to X_i$, item *i* is independent to item *j*.
4) if and only if there exists an equivalence or prerequisite relationship between item *i* and item *j*, we say a direct relationship between item *i* and item *j*.

3.2. Improved Item Relational Structure Theory, IIRST

Liu *et al*. point out that the threshold in IRST is a constant value and lacks statistical significance, so they put forward IIRST [10]. Let $r_{ij} = \mu$, the threshold value as follow:

$$r_c = \arg[1 - \int_{-\infty}^{x} f(r_{ij}) dr_{ij} = 0.05] \quad (4)$$

where $f(r_{ij})$ is pdf of r_{ij}, $r_{ij} \sim t \quad or \quad N$, then:

(i) if and only if $r_{ij} > r_c$ and $r_{ji} \le r_c$, exist a prerequisite relationship $X_i \Rightarrow X_j$.

(ii) if and only if $r_{ij} > r_c$ and $r_{ji} > r_c$, exist an equivalence relationship $X_i \Leftrightarrow X_j$.

3.3. Liu's Item Relational Structure Theory, LIRST

Liu *et al*. found that IRST could not well meet the principles of completeness, normalization and consistency. Therefore, Liu's project related structure theory is put forward [6]. This model is calculated based on the parameter μ of IRST, the ordering index is as follows:

$$\gamma_{ij}^{(LIRST)} = \begin{cases} 0.1 & if\ P(X_i = 0)P(X_j = 1) \\ 0.1 + 0.9\mu & otherwise \end{cases} \quad (5)$$

If and only if $\gamma_{ji}^{(\mathrm{LIRST})} > 0.5$, there exists $X_i \to X_j$, otherwise $X_i \not\to X_j$. Then make decision of the prerequisite relationship by the definition of IRST.

4. Noise-Filter Item Relational Structure Algorithm

Based on Liu's research results in 2013, guessing and slipping led to the instability of item ordering results [7]. Therefore, this paper proposes a new algorithm to improve the accuracy of IRST algorithm by eliminating the influence of guessing and slipping. We follow the definition of guessing and slipping parameters as "deterministic input noisy 'and' gate (DINA)" model, and use this model to estimate guess and slip parameters [12].

4.1. Formulation of the DINA Model

Let $i(=1,2,\ldots,I)$ denote a student, $j(=1,2,\ldots,J)$ denote an item, $k(=1,2,\ldots,K)$ is an attribute that denotes cognitive elements, and $l(=1,2,\ldots,L)$ is an attribute master pattern.

Definition 4.1: definite *attribute mastery pattern vector* as $\alpha_{ij} = (\alpha_{l1},\ldots,\alpha_{lk},\ldots,\alpha_{lK})$ as a combination of the binary variables

$$\alpha_{lk} = \begin{cases} 1, & mastering\ kth\ attribute \\ 0, & otherwise \end{cases}. \quad (6)$$

Definition 4.2: if a student has all required attributes for item j, definite a probability of obtaining an incorrect answer as s_j. On the other hand, definite g_j as a probability of the correct response when a student lacks at least one attribute required for item j. We call s_j and g_j item parameters.

Thus, the DINA model can be expressed as follows:

$$P(X_{ij} = 1 | \alpha_i) = (1 - s_j)^{\eta_{ij}} g_j^{1-\eta_{ij}}, \quad (7)$$

where $\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}$, q_{jk} is an element of the Q-matrix, it means if item *j* test the attribute *k*, $q_{jk} = 1$, otherwise $q_{jk} = 0$.

4.2. Noise-Filter OT Algorithm & IRS Algorithm

According to Liu's research in 2013, if there are no guesses and slips in the subjects' responses, the original item prerequisite (i.e. order relationship) can be 100% restored through OT or other ranking methods. However, many studies have proved that there must be some guesses and slips in the test. Therefore, this study restores a more accurate response probability by considering guesses and slips, so as to improve the accuracy of the OT or IRS algorithm.

All the definitions in DINA are followed, so we can estimate item parameters s_j and g_j through response matrix and DINA model. Thus, noise-filter improved Liu's ordering theory algorithm (NFILOT) as follows:

Definition 4.3: Noise-Filter Improved Liu's Ordering theory

Define the response vector as $X_T=(X_1,X_2,\ldots,X_n)$, $X_i=0$ represents wrong answer to item *i*, $X_j=1$ represents correct answer to item *j*, Then the joint response probability is $P(X_i=0,X_j=1)$.

(i) The ordering index from X_i to X_j is defined as below:

$$\gamma_{ij}^{NFILOT}=\tilde{P}(X_i=0,X_j=1)=\frac{(1+s_i)P(X_i=0,X_j=1)}{(1+g_j)}. \tag{8}$$

(ii) Let the threshold limit value of NFILOT denoted as γ_c^{NFILOT}, be defined as

$$\gamma_c^{NFILOT}=\arg_x[1-\int_{-\infty}^{x} f(\gamma_{ij}^{NFILOT})d\gamma_{ij}^{NFILOT}=0.05], \tag{9}$$

where $f(\gamma_{ij}^{NFILOT})$ is the pdf of random variable γ_{ij}^{NFILOT}.

(iii) There exists an ordering relationship $X_i \to X_j$, if and only if,

$$\tilde{P}(X_i=0,X_j=1)\le\gamma_c^{NFILOT}, \tag{10}$$

otherwise $X_i \not\to X_j$.

Similarly, Noise-filter relational structure algorithm as follows:

Definition 4.4: Noise-Filter Improved Item Relational Structure Theory (NFIIRST)

Response vector as $X_T=(X_1,X_2,\ldots,X_n)$, $X_i=0$, $X_j=1$, $P(X_i=0,X_j=1)$ was defined the same as given above.

(i) The ordering index from X_i to X_j is defined as below:

$$\gamma_{ij}^{NFIIRST}=1-\frac{\tilde{P}(X_i=0,X_j=1)}{\tilde{P}(X_i=0)\tilde{P}(X_j=1)}, \tag{11}$$

where $\tilde{P}(X_i=0,X_j=1)$ is defined the same as NFOT, and $\tilde{P}(X_i=0)$ can be calculated as follows:

$$\tilde{P}(X_i=0)=\frac{P(X_i=0)}{1+s_i}/[\frac{P(X_i=0)}{1+s_i}+\frac{P(X_i=1)}{1+g_i}] \tag{12}$$

$$=\frac{(1+g_i)P(X_i=0)}{1+s_i+(g_i-s_i)P(X_i=0)}, \tag{13}$$

$\tilde{P}(X_j=1)$ can be calculated as follows:

$$\tilde{P}(X_j=1)=\frac{P(X_j=1)}{1+g_j}/[\frac{P(X_j=1)}{1+g_j}+\frac{P(X_j=0)}{1+s_j}] \tag{14}$$

$$=\frac{(1+s_j)P(X_j=1)}{1+g_j+(s_j-g_j)P(X_j=1)}, \tag{15}$$

(ii) Let the threshold limit value of NFILOT denoted as $\gamma_c^{NFIIRST}$, be defined as

$$\gamma_c^{NFIIRST}=\arg_x[1-\int_{-\infty}^{x} f(\gamma_{ij}^{NFILOT})d\gamma_{ij}^{NFILOT}=0.05]$$

where $f(\gamma_{ij}^{NFIIRST})$ is the pdf of random variable $\gamma_{ij}^{NFIIRST}$.

(iii)There exists an ordering relationship $X_i \to X_j$, if and only if,

$$1-\frac{\tilde{P}(X_i=0,X_j=1)}{\tilde{P}(X_i=0)\tilde{P}(X_j=1)}\le\gamma_c^{NFIIRST}, \tag{16}$$

otherwise $X_i \not\to X_j$.

Then make decision of the prerequisite relationship by the definition of IRST.

5. Experiments and Results

For comparing the performance of bart's OT, Takeya's IRS algorithm, Liu's IIRS algorithm and our new method, we simulation a test with 7 items and 5 attributes by DINA mode, number of participants set as 500.

The simulation were repeated 10 times, to calculate the average value of validity index for OT, IRS and NFIRS, the validity index was proposed by Liu in 2013 [13]:

$$Val(I'|I)=0.5\left[1+\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(e_{ij}-\bar{e})(e'_{ij}-\bar{e})}{\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(e_{ij}-\bar{e})^2}\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(e'_{ij}-\bar{e})^2}}\right], \tag{17}$$

where, I and I' were matrixes of item ordering, e_{ij} was element of I, e'_{ij} was element of I', $\bar{e}$ and $\bar{e}'$ could calculated as follow:

$$\bar{e}=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}e_{ij}, \tag{18}$$

$$\bar{e}'=\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}e'_{ij}, \tag{19}$$

$Val(*)\in\{0,1\}$, the larger the value is, the better the validity of ordering algorithm method is.

Table 1 list the result of the experiment, the validity index takes the average of ten results:

Table 1. Validity Index of Different Algorithm

	algorithm				
	OT	IRST	IIRST	NFILOT	NFIIRST
Mean of Validity index	0.765	0.756	0.756	0.822	0.804

Above results showed that the new method — NFILOT & NFIIRST are obviously better than OT & IRST.

6. Conclusion and Prospect

At present, the two primary item ordering methods — OT and IRST, and their variants, item ordering result of them was rely on the threshold, that is because the ordering index fluctuates greatly. The method developed in this paper solves the problem of ordering index fluctuation fundamentally, and develops a computer program for this method. It can also be seen from the simple experimental results that the method in this paper is better than the previous methods, this shows the advantages of the new algorithm.

The NFILOT and NFIIRST algorithms are only applicable to binary scoring items, which makes them have certain limitations in practical application. In the future, researchers can consider developing NF ordering analysis method with multi-level scoring.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

LZ conducted the research, made the simulation experiment and wrote the paper; SD provided guidance and revision for paper writing; XL assisted in the experimental study; all authors had approved the final version.

Acknowledgment

This paper is supported by Graduate innovation foundation of Jiangxi Provincial Department of Education (YC2020-B084).

References

[1] Airasian, P. W., & Bart, W. M. (1973). Ordering theory: A new and useful measurement model. *Educational Technology, 13(5)*, 56-60.

[2] Liu, H. C., Yu, Y. K., & Tsai, H. C. (2013). Item relational structure theory based on Liu's improved nonparametric IRT theory. *Applied Mechanics and Materials, 479-480*, 1193-1196.

[3] Takeya, M. (1980). Relational Structure analysis among test items on performance scores. *Japan Society for Science Education, 4(4)*,183-192.

[4] Bart, W. M., & Mertens, D. M. (1979). The hierarchical structure of formal operational tasks. *Applied Psychological Measurement, 3(3)*, 343-350.

[5] Bart, W. M., Rothen, W., & Read, S. (1986). An ordering-analytic approach to the study of group differences in intelligence. *Educational and Psychological Measurement, 46(4)*, 799-810.

[6] Liu, H. C., Shia, B. C., Ju, J. M., Wang, T. L., Su, C. H., & Lin, Y. T. (2016). A novel generalized item relational structure theory based on Liu's normalization and consistency criteria. *ICIC Express Letters, 10(12)*, 2957-2962.

[7] Liu, H. C. (2013). A novel validity index for evaluating the item ordering structure based on Q-matrix theory. *Research Journal of Applied Science, Engineering and Technology, 6(1)*, 53-56.

[8] Bart, W. M., & Airsian, P. W. (1972). *The generation of item hierarchies by an ordering theoretic method on a Piagetian example*. Paper presented at the meeting of American Educational Research Association, Chicago.

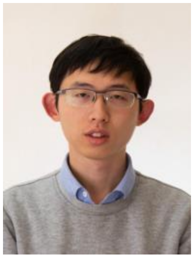
[9] Bart, W. M., & Airsian, P. W. (1974). Determination of the ordering among seven Piagetian tasks by an ordering-theoretic method. *Journal of Educational Psychology, 66(2)*, 277-284.

[10] Liu, H. C., Shia, B. C., & Cheng, D. Y. (2014). *An improved Liu's ordering theory based on empirical distributive critical value.* Paper presented at the meeting of the 27th of International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems, Kaohsiung, Taiwan.

[11] Liu, H. C., Wu, S. N., & Chen, C. C. (2011). Item relational structure algorithm based on empirical distribution critical value. *Journal of Software, 6(11)*, 2106-2113.

[12] de la Torre, J. (2009). DINA model and parameter estimation: a didactic. *Journal of Educational and Behavioral Statistics, 34(1)*, 115-130.

[13] Liu, H. C. (2013). A novel validity index for evaluating the item ordering structure based on Q-matrix theory. *Research Journal of Applied Science, Engineering and Technology, 6(1)*, 53-56.



Li-Wen-Yuan Zhou was born in 1991, Jiangxi Province, China. He received the master degree in psychology from Jiangxi Normal University, Jiangxi, China. His main research areas are classical measurement and cognitive diagnosis.

He is a Ph.D. candidate of psychology; Ph.D. research direction is psychological statistics and measurement. Since 2017, he has taught psychology course in Yichun early childhood teachers college. He presided over a provincial project and participated in the moral education monitoring in the 2019 basic education quality monitoring in Jiangxi Province.

Mr. Zhou is a member of Jiangxi psychological society and Jiangxi Social Psychological Society

Sheng-Hong Dong was born in 1971, Jiangxi Province, China. He received the Ph.D. degree in psychology from Jiangxi Normal university, Jiangxi, China. professor of psychology at the Jiangxi Normal University. Main research areas are psychological statistics and measurement, cognitive psychology and social psychology.

He has been engaged in the applied research of education and psychological measurement for a long time, and has successively presided over the research of more than 20 topics, including China National Social Science Foundation, provincial major commissions and key topics. As an official document of the Ministry of education, the statistical measurement specification for online marking of national education examination drafted by him has been applied in Chinese college entrance examination. Participated in the revision of Wechsler children's intelligence scale 4th Edition (Chinese version), adaptive behavior assessment scale and other famous scales. He also presided over the moral education monitoring in the 2019 basic education quality monitoring in Jiangxi Province. He once served as the vice president of the school of education, the vice president of the school of psychology and the director of the Department of Social Sciences of Jiangxi Normal University. Now he is the executive vice president of the Graduate School of Jiangxi Normal University.

Prof. Dong is currently the director of Jiangxi Key Laboratory of psychology and cognitive science, Editor in chief of the Journal of psychological exploration, the vice president of statistical measurement branch of China Education Society, the president of Jiangxi psychological society and the president of Jiangxi Social Psychological Society.

Xing-Gao Li was born in China. His main research areas are classical measurement and cognitive diagnosis.

He is a graduate student of psychology, and is currently studying at Jiangxi Normal University. He participated in the moral education monitoring in the 2019 basic education quality monitoring in Jiangxi Province and a provincial project.

Mr. Li is a member of Jiangxi psychological society and Jiangxi Social Psychological Society.