

IRT Analysis of the Mental Health Test for Elementary School Students

Siyu Luo, Yangying Li, Yi-Jiun Lin*

School of Educational Science, Yulin Normal University, Yulin, China

Email: *linyijiun@qq.com

How to cite this paper: Luo, S. Y., Li, Y. Y., & Lin, Y.-J. (2024). IRT Analysis of the Mental Health Test for Elementary School Students. *Open Journal of Social Sciences*, 12, 611-622.

<https://doi.org/10.4236/jss.2024.125033>

Received: April 5, 2024

Accepted: May 28, 2024

Published: May 31, 2024

Abstract

This paper utilized the two-parameter logistic IRT model to analyze survey data of the Mental Health Diagnostic Test (MHT) administered to students in grades 4 - 6 at an elementary school. The findings indicate that the data aligns with the basic assumptions of IRT, and the parameter estimation for most items meets theoretical expectations. Since deleting certain items does not lessen the test's information, it is suggested that 10 or 15 items be considered for revision.

Keywords

Elementary School Students, Mental Health Diagnostic Test (MHT), IRT Item Analysis

1. Introduction

According to the World Health Organization (WHO), health is “not merely the absence of disease or physical weakness, but the physical, mental, and social well-being of the individual”. Mental health, as an integral component of health and well-being, refers to an individual's capacity to adjust their psychological state to cope with stress while adhering to social norms and cultural constraints. It also refers to the ability to take action to enhance one's own life (Zheng et al., 2004). Mental health is a lifelong process that varies across different life stages.

Developing a positive mindset during adolescence has a significant impact on an individual's future life trajectory and is associated with their accurate perception of their personal life, family, and society. However, the academic community is more concerned with the mental health of college students, high school pupils, and mature adults (Zheng et al., 2008). Until the Guidelines for Mental Health Education in Elementary and High School were released in 2002 by the

Ministry of Education, the mental health of elementary school students went largely unnoticed. The goals of these initiatives were to foster strong character and mental health, as well as to enhance the well-being of students who have experienced psychological distress or problems by providing timely and effective interventions.

Appropriate and effective interventions require accurate evaluation as a necessary precondition. There are two categories of mental health assessment instruments: those that evaluate mental health from either a positive or negative perspective. For example, the Child and Adolescent Depression Inventory (CDI), and the Child and Adolescent Anxiety Scale (RCMAS) are intended to address specific emotions and symptoms. Those that measure mental health holistically take into account both positive and negative feelings, including the Mental Health Diagnostic Test (MHT), the Chinese Middle School Students' Mental Health Inventory, and Symptom Checklist 90 (SCL-90).

Such assessment instruments used in China, however, have a few issues (Liu, 2003; Zheng et al., 2008). First, scales imported from abroad do not take into account the moral beliefs and behavioral traits valued by Chinese culture, making it challenging to accurately measure the intended constructs for Chinese elementary school students. Second, the imported scales may be outdated since their time of introduction and revision may have been many years ago, which could be detrimental to their applicability. Third, most of these scales focus on negative psychological traits like depression, anxiety, and compulsion, which may not accurately reflect the real-life experiences and developmental challenges faced by elementary school students. Therefore, scholars (Liu, 2003; Zhang & Su, 2015; Zhang et al., 2004; Zheng et al., 2008) have developed mental health measures for elementary school students.

The Mental Health Diagnostic Test compiled by Zhou (1991) serves as a common instrument to measure the mental health status of primary and high school students in China (Lin et al., 2010; Li et al., 2017; Xi et al., 2021). The scale has been developed over many years, and only Chen (2002), Zheng et al. (2004), Zheng et al. (2005), and Wei & Ren (2009) have proposed revisions in their studies on high school or elementary school students, respectively. Chen (2002), Zheng et al. (2004), and Zheng et al. (2005) considered it unrealistic to limit responses to just "yes" or "no." Therefore, they modified the original 2-point scoring system to a 5-point Likert-type scale. They also revised the wording of several items to administer a survey to high school students. After analyzing the data, they concluded that the items in MHT could be reduced to 83, 69, and 69 items, respectively. By contrast, Wei & Ren (2009) maintained the 2-point scoring system in their investigation of elementary school students and suggested 59 items in their conclusion.

These revisions from the previous studies are mainly based on the item analysis of Classical Test Theory (CTT). But the CTT approach suffers from three major theoretical limitations in practical application (Rusch et al., 2017). First, CTT assumes a linear relationship between the latent variable and observed

scores, which rarely reflects the empirical reality of behavioral constructs. Second, parameters such as item discrimination and item difficulty depend on the sample being used. In other words, varying results are obtained when the identical test is given to different samples. Third, the reliability of CTT is closely linked to parallel measures. However, the true score cannot be estimated directly, or it can only be determined by making unrealistic assumptions. Comparing these two methods, in a study by Zhu and Zhang (2003), it was found that Item Response Theory (IRT) yielded more accurate results for personality tests than CTT method.

Thus, the present study aims to use the IRT approach to analyze the Mental Health Diagnostic Test (MHT) and make applicable suggestions for revising the items of it.

2. Method

2.1. Measures

The Mental Health Diagnostic Test (MHT) (Zhou, 1991), designed for students from Grade Four in elementary school to Senior Three in high school, is a widely used mental health assessment scale for primary and high schools in China. The full scale consists of one validity sub-scale and eight sub-scales, including learning anxiety, sensitive tendency, loneliness, self-blame, allergy tendency, physical symptoms, horror tendency, and impulsive tendency, totaling 100 items. The scale uses a two-point scoring system, where each question requires the subject to choose either “yes” (1 point) or “no” (0 points) based on their actual situation. The higher the total score, the worse the mental health status. 65 is the cut-off point in the total score, above which the presence of certain psychological problems is considered.

2.2. Sampling and Subjects

A total of 1497 students in 33 classes from grade 4 to 6 in an elementary school were selected using convenience sampling and were given a questionnaire with the MHT test. After excluding incomplete responses, garbled answers, and questionnaires with validity scale scores greater than or equal to 8, there were 1183 valid questionnaires (see **Table 1**), resulting in a response rate of 79%. With a threshold score of 65, the positive screening rate was 1.69% (20 out of 1183), which closely aligns with the 1.7% screening rate reported by Yu et al. (2019) for 7672 students in grades 3 - 6 in Guangxi, China. This suggests that the study

Table 1. Respondents' sex and grade.

Sex	Grade 4	Grade 5	Grade 6	Total
Boy	217	177	237	631
Girl	201	159	192	552
Total	418	336	429	1,183

sample is somewhat representative of the target population.

2.3. Statistical Model

The MHT test is a two-point scale, which is not affected by random or guessing responses under normal circumstances. Therefore, the two-parameter logistic model (2PL) of the IRT method was utilized to conduct statistical analysis using the Stata 15.0 program.

3. Results

3.1. Test for IRT Assumptions

Before estimating the parameters of the IRT model, hypothesis testing should be conducted on the data to prevent drawing biased conclusions. Item response theory has three fundamental assumptions: unidimensionality, local independence, and monotonicity.

Unidimensionality means that only one of the potential traits measured by all the items in the scale plays a determining role. If this assumption is violated, the items are not homogeneous, making it difficult for the IRT model to accurately estimate the parameters of the items and their ability. Among the methods related to testing unidimensionality, the eigenvalue ratio is the most commonly used (Wang & Zhou, 2018). Since the items of the MHT test are all on 2-point scales, it is appropriate to use the tetrachoric correlation matrix for factor analysis (Mislevy, 1986; Muthén, 1989; Uebersax, 2000). The results (see **Table 2**) showed that after excluding the validity scale, the first eigenvalue of each subscale was substantially larger than the second eigenvalue. The eigenvalue ratios for all subscales and the total scale, except for physical symptoms, were greater than 3, conforming to the general criterion for unidimensionality that the first-factor eigenvalue is 3 to 5 times the second-factor eigenvalue. Nonetheless, one subscale had a slight violation of this assumption. According to Guo et al.

Table 2. The factor eigenvalues of MHT subscales.

Scale	First-factor eigenvalue	Second-factor eigenvalue	First-factor % of Variance	Eigenvalue ratio
Total scale	18.20	3.59	0.84	5.06
Learning Anxiety	3.92	1.09	0.78	3.59
Interpersonal Anxiety	2.59	0.50	0.84	5.19
Loneliness Tendency	2.90	0.48	0.86	6.04
Self-blame Tendency	3.16	0.83	0.79	3.81
Allergic Tendency	2.70	0.46	0.85	5.82
Physical Symptoms	4.31	2.56	0.63	1.68
Horror Tendency	4.10	0.54	0.88	7.65
Impulsiveness Tendency	3.59	0.52	0.87	6.90

(2005), there is no absolute criterion for unidimensionality. Thus, they suggested that the model could still be robust and provide accurate estimates even if the assumption of unidimensionality is slightly violated.

Local independence refers to the phenomenon where the subjects' responses to an item are solely based on their trait level, unaffected by the test content, item order, or test procedure of other items (Henning, 1989). In conducting the test, on the one hand, it is to determine whether the subjects answered truthfully; on the other hand, it is to determine whether there is content dependence or correlation between the items (Luo, 2012). The methods to test this assumption include residual correlation, chi-square test, G2 test, Q3 test, etc. (Wang & Zhou, 2018). However, these methods are not suitable for a 2-point scale. For dichotomous items, Liu and Maydeu-Olivares (2013) found that methods with higher test power typically required the calculation of the information matrix. However, when applied to tests with many items (more than 30 questions), the ability to identify poor performance is reduced. Therefore, according to Henning (1989), if the data fit the assumption of unidimensionality, local independence can be considered equivalent in practical applications.

Monotonically increasing refers to the concept that as the potential trait θ increases, the probability of the subject answering "1" also increases. It can be shown as the item characteristic curve, which is an S-shaped curve with an increasing trend from left to right (Yang et al., 2012; Yang & Kao, 2014). Since the content of the MHT items is negatively formulated, higher total scores indicate that mental health is likely to be poorer. This is reflected in the item characteristic curve, where the greater the latent trait θ (indicating poorer mental health), the more likely it is that the response to an item will be "yes." From the item characteristic curve (see next section), it can be intuitively observed that all 89 items meet this assumption, except for Item 9, which displays a horizontal line. The data used in this paper collectively adhere to the assumptions of Item Response Theory (IRT) and are appropriate for estimating item parameters, as well as for analyzing item information and test information.

3.2. Estimation of Parameters

For item analysis based on IRT, item quality can be assessed in terms of item discrimination and item difficulty (see **Table 3**). Item discrimination (a) is generally considered acceptable when it falls within the range of $0.3 < a \leq 3$ (Han et al., 2013). It is commonly believed that items with negative discrimination should be eliminated (Yang & Kao, 2014). In the MHT test, no items with negative discrimination but four items with a discrimination estimate of less than 0.3, suggested that these four items are unable to effectively differentiate the mental health status of the subjects:

- Item 1: "When you are going to sleep at night, do you always think about tomorrow's homework?"
- Item 9: Do you want to excel in every exam?

Table 3. Item parameters of MHT.

Learning Anxiety			Interpersonal Anxiety			Loneliness Tendency			Self-blame Tendency		
item	a	b	item	a	b	item	a	b	item	a	b
q1	0.27	-0.09	q16	1.09	0.39	q26	0.24	1.11	q36	1.04	-0.02
q2	0.91	0.71	q17	1.20	0.35	q27	0.52	2.02	q37	0.98	-0.53
q3	0.83	0.09	q18	0.67	3.31	q28	1.33	0.98	q38	0.94	0.42
q4	0.68	-1.56	q19	0.68	-0.55	q29	0.94	2.24	q39	0.74	-0.97
q5	0.77	-0.41	q20	0.59	2.61	q30	0.26	0.82	q40	0.76	-0.64
q6	0.67	-1.06	q21	0.68	1.97	q31	1.55	0.70	q41	0.69	0.46
q7	0.60	-0.89	q22	0.79	1.29	q32	0.78	1.96	q42	0.85	0.45
q8	0.89	-0.32	q23	0.91	0.82	q33	0.78	1.50	q43	0.93	0.12
q9	0.03	80.81	q24	0.84	-1.24	q34	0.64	2.18	q44	0.88	-0.83
q10	0.85	-0.13	q25	1.11	-0.16	q35	0.94	1.38	q45	0.46	0.86
q11	1.05	-0.03									
q12	0.76	2.45									
q13	0.54	0.93									
q14	0.72	1.13									
q15	1.10	1.03									
Allergic Tendency			Physical Symptoms			Horror Tendency			Impulsiveness Tendency		
item	a	b	item	a	b	item	a	b	item	a	b
q46	0.31	-5.24	q56	0.77	-0.74	q71	0.54	0.12	q81	1.11	0.99
q47	0.77	1.03	q57	0.32	2.00	q72	1.29	0.61	q83	1.33	0.79
q48	0.72	0.01	q58	1.13	1.28	q73	1.17	1.18	q85	1.21	0.79
q49	1.04	-0.93	q59	1.11	-0.37	q74	0.98	0.81	q87	1.10	0.87
q50	0.79	-1.71	q60	0.88	1.24	q75	0.91	1.39	q89	0.98	0.35
q51	0.90	0.90	q61	1.28	0.74	q76	1.05	1.74	q91	0.71	1.38
q52	0.67	1.23	q62	0.58	1.32	q77	0.86	1.58	q93	0.99	1.22
q53	1.30	-0.36	q63	0.99	0.48	q78	0.68	2.02	q95	0.42	4.33
q54	1.13	-0.96	q64	1.25	0.56	q79	1.00	0.44	q97	1.18	1.19
q55	1.04	-0.72	q65	1.33	0.20	q80	1.42	0.39	q99	1.29	0.73
			q66	0.79	1.79						
			q67	0.55	2.46						
			q68	0.80	3.78						
			q69	1.23	1.72						
			q70	0.73	3.03						

Note: a: Item Discrimination; b: Item Difficulty.

- Item 26: Do you not always laugh when your classmates are laughing?
- Item 30: Do you choose not to participate when your classmates are talking?

In addition, the theoretical range of item difficulty (b) is $[-\infty, +\infty]$, which typically falls between $[-3, +3]$ (Luo, 2012). The lower the difficulty value, the more likely subjects answer “yes” to the item. For example, if Item 9 has a difficulty value of $b = -80.81$, 89.01% of the subjects chose “yes” for this item. This suggests that the item is too common, and most people tend to answer positively even if they do not have a mental health condition. On the other hand, a difficulty value that is too high indicates that only a minority of subjects answered “yes” to agree with the item content. For example, Item 95 with a difficulty value of $b = 4.33$, only 14.45% of the subjects chose “yes” for this item.

As a result, six items with a difficulty estimate greater than 3 indicated that only a few respondents had thoughts or experiences with these topics. The items are listed as follows.

- Item 9: the same as listed above.
- Item 18: If you lose a game or a competition with someone else, then you don’t want to do it anymore?
- Item 46: Do you always sincerely want to do something good for the class?
- Item 68: Do you often lick your fingers?
- Item 70: Do you go to the toilet more often than others?
- Item 95: Do you have to get what you want?

Furthermore, the item characteristic curves also illustrate the probability of responding “yes” to the latent trait. As depicted in **Figure 1**, the estimated probabilities of the sub-scales, such as Loneliness Tendency, Self-blame Tendency, Horror Tendency, and Impulsiveness Tendency, exhibit consistent patterns across trait levels. This indicates that the curves of items within a certain sub-scale are more similar. In contrast, the curves of the items are more dispersed in subscales such as Learning Anxiety, Interpersonal Anxiety, Allergic Tendency, and Physical Symptoms. This suggests significant differences in the estimation of items at a specific level of the latent trait. Among these items, items 9, 46, and 57 were undifferentiated, suggesting that most respondents were likely to answer “yes.”

- Item 9, Item 46: the same as listed above.
- Item 57: Do you blush when you are shy?

3.3. Item Information Functions

The amount of information provided by each item was further analyzed. The more information an item provides, the more precise the measurement will be. Items measured with greater precision offer more information and are graphically represented as longer and narrower compared to those with less information. For instance, when the level of discrimination is too low (<0.3), the items may offer insufficient information for estimating the latent traits of the subjects. As can be seen from the item information curves (see **Figure 2**), some items

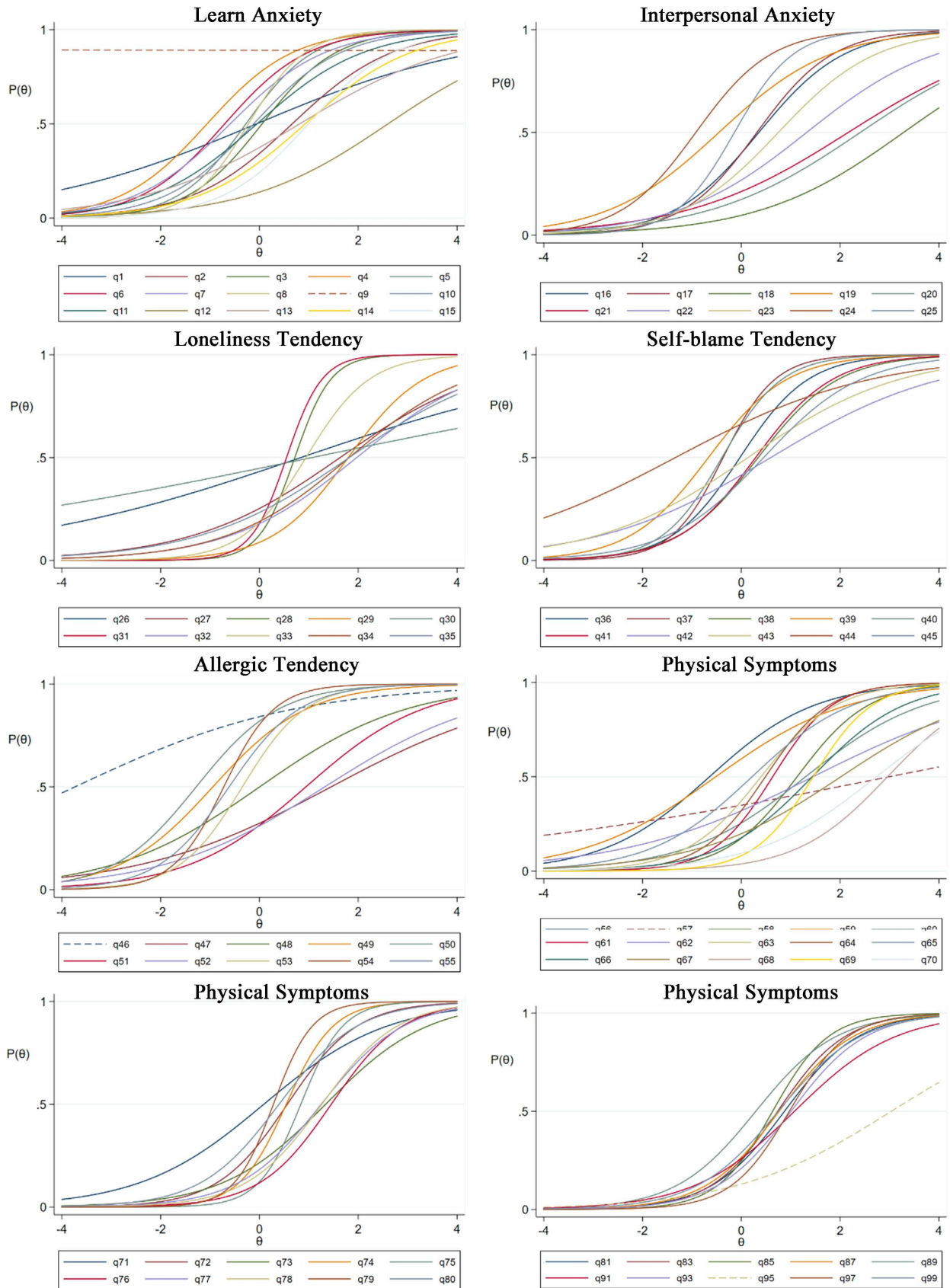


Figure 1. Item characteristic curves.

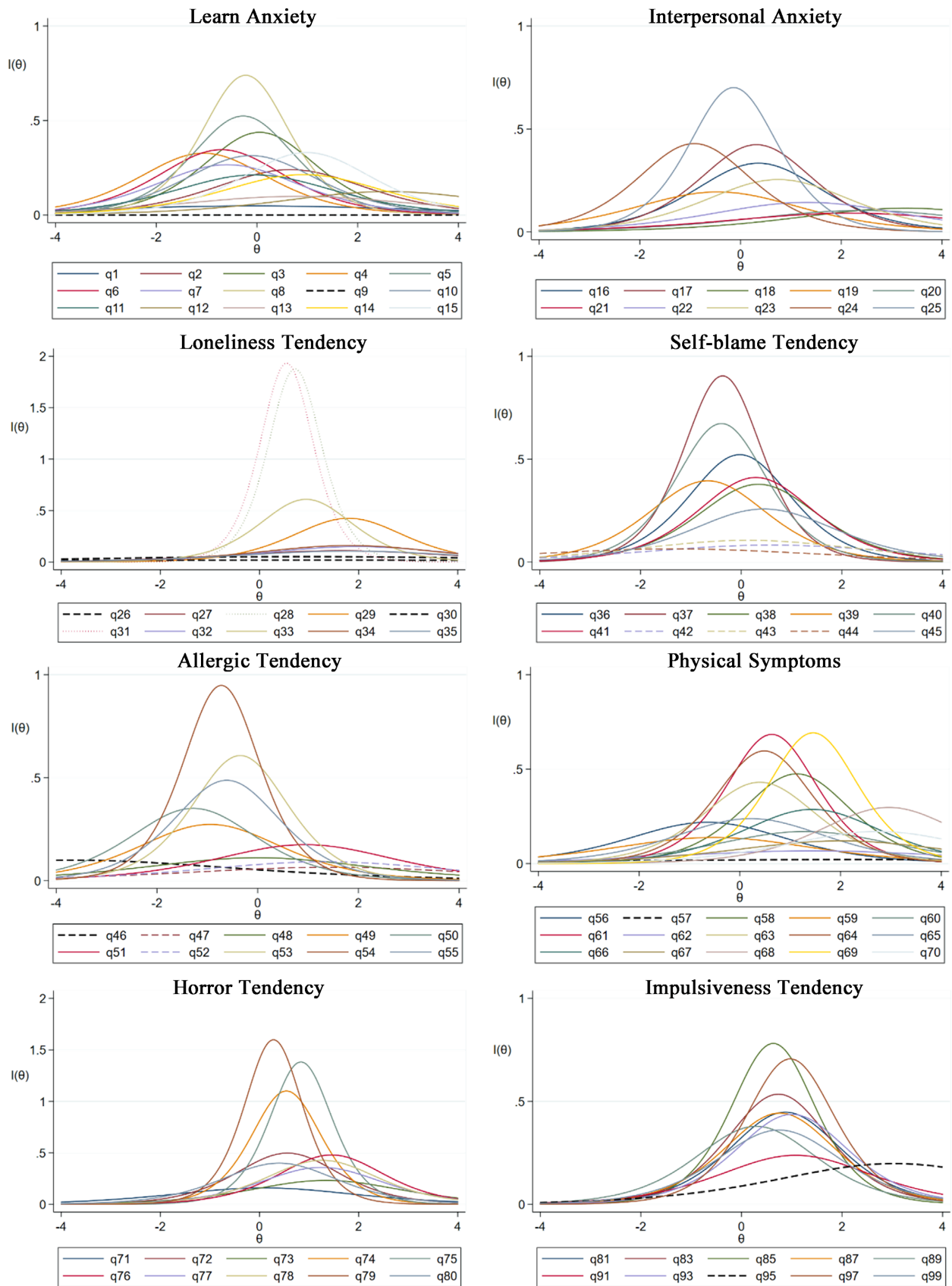


Figure 2. Item information function curves.

exhibited flat curves, suggesting that these items provided minimal information regardless of the respondents' inclination toward mental health issues. The items are listed as follows.

- item 9, item 26, item 30, Item 46, Item 57: the same as listed above.
- item 42: Do you participate in sports competitions such as table tennis, badminton, and gymnastics, paying special attention whenever you make a mistake?
- Item 43: Do you find it difficult to cope with challenging situations when you encounter them?
- Item 44: Do you sometimes regret not doing that thing?
- Item 52: Are you particularly sensitive to the sound of the radio and cars?

3.4. Test Information Functions

The test information function (see **Figure 3**) reflects the relationship between the contribution of information to the overall test in evaluating different levels of traits. The richer the information provided by the test, the more accurate the test is in evaluating those levels of traits. The test information curve, as shown in **Figure 3**, indicates that the MHT test is a reliable measure for subjects with a trait level of $\theta = 0.5$. After excluding 10 ineffective questions (1, 9, 18, 26, 30, 46, 57, 68, 70, 95) or 15 questions (adding 42, 43, 44, 47, 52 to the previous 10), there was only a minor change in the test information function, indicating that the excluded items might have minimal impact on the assessment of mental health status.

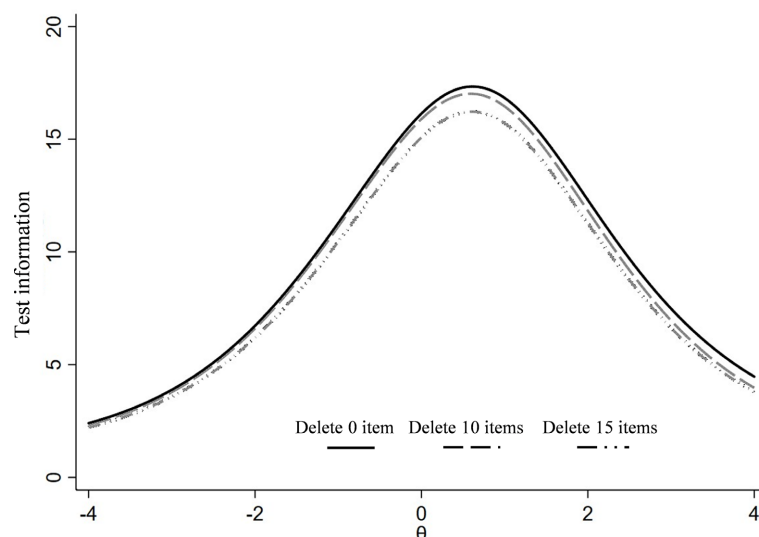


Figure 3. Test information function.

4. Conclusion

In this study, we analyzed the MHT test using the 2PL model to examine the performance of each item and subscales in terms of item discrimination, item difficulty, item information function, and test information function. The results

showed that most of the items met the theoretical requirements in terms of item difficulty and item discrimination. After removing undifferentiating or easy items, the test information curve rarely changed significantly. The findings suggest that deleting these items could be considered to reduce the questionnaire length and prevent biased evaluation or intervention on the subjects.

This conclusion is supported by the opinions of research subjects. During data collection in the field, many participants mentioned feeling fatigued after completing the test, which consisted of 100 items and took approximately 25 minutes. Therefore, it was feasible to reduce the number of items without compromising the quality of the test. According to interviews with teachers and students, the 5-point scale, in comparison to the original dichotomous choice, can more accurately reflect the various levels of responses to each item and align better with the actual circumstances of the subjects.

Currently, most psychological tests and assessments utilize CTT-based methods for item analysis. Using the IRT method to screen items and revise scales is also a beneficial strategy for enhancing the accuracy of such tests. If the two methods can be effectively combined and complement each other, a more accurate scale could be obtained.

Acknowledgments

This work was supported by a grant from the Research Initiation Fund Project of Yulin Normal University (G2023SK06).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- Chen, C. Y. (2002). Exploring the Structure of Mental Health Diagnostic Tests for High school Students. *Journal of Ningbo University (Educational Science Edition)*, 3, 10-13.
- Guo, Q. K., Chen, Y. M., & Meng, Q. M. (2005). Feasibility of Applying IRT to Self-Report Tests. *Journal of Psychology*, 2, 275-279.
- Han, X., Wu, R., & Chen, Y. X. (2013). Unidimensionality Test of College Students' Low Self-Esteem Questionnaire under IRT Framework. *Journal of Chifeng College (Natural Science Edition)*, 21, 91-92.
- Henning, G. (1989). Meanings and Implications of the Principle of Local Independence. *Language Testing*, 6, 95-108. <https://doi.org/10.1177/026553228900600108>
- Li, F. L. et al. (2017). Meta-Analysis of the Mental Health Diagnostic Test for Left-Behind Children in China in the Past Ten Years. *Chinese Journal of Child Health*, 25, 493-495.
- Lin, X. H., Shen, M., Wang, L., & Wang, Y. J. (2010). Meta-Analysis of Mental Health Status of Left-Behind Children in Rural China. *Journal of Huazhong University of Science and Technology, Medical Edition*, 2, 228-231.
- Liu, Y. R. (2003). Development of a Self-assessment Scale for Mental Health of Primary and High School Students. *Journal of Psychological Science*, 3, 515-516.
- Liu, Y., & Maydeu-Olivares, A. (2013). Local Dependence Diagnostics in IRT Modeling of Binary Data. *Educational and Psychological Measurement*, 73, 254-274.

- Luo, Z. S. (2012). *Fundamentals of Item Response Theory*. Beijing Normal University Press.
- Mislevy, R. J. (1986). Recent Developments in the Factor Analysis of Categorical Variables. *Journal of Educational and Behavioral Statistics*, 11, 3-31.
<https://doi.org/10.3102/10769986011001003>
- Muthén, B. (1989). Dichotomous Factor Analysis of Symptom Data. *Sociological Methods and Research*, 18, 19-65. <https://doi.org/10.1177/0049124189018001002>
- Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking Free from the Limitations of Classical Test Theory: Developing and Measuring Information Systems Scales Using Item Response Theory. *Information & Management*, 54, 189-203.
<https://doi.org/10.1016/j.im.2016.06.005>
- Uebersax, J. S. (2000). *Estimating a Latent Trait Model by Factor Analysis of Tetrachoric Correlations*. <http://www.john-uebersax.com/stat/irt.htm>
- Wang, W. L., & Zhou, Y. Q. (2018). Current Status and Outlook of the Application of Item Response Theory in Health-Related Scales. *China Health Statistics*, 4, 633-636.
- Wei, G. L., & Ren, J. (2009). Revision of the Mental Health Diagnostic Test (MHT) in an Elementary School Student Population. Abstracts of the 12th National Conference on Psychology. *The Twelfth National Conference on Psychology*.
- Xi, Y., Xu, W. W., & Zhang, N. (2021). A Cross-sectional Historical Study on the Change of Mental Health Level of Rural Left-behind Children. *Psychological Technology and Application*, 5, 283-292.
- Yang, F. M. & Kao, S. T. (2014). Item Response Theory for Measurement Validity. *Shanghai Archives of Psychiatry*, 26, 171-177.
- Yang, J. Y., He, J. Z., & Zhao, S. Y. (2012). Item Response Theory Analysis of Eysenck Personality Questionnaire's Item Quality. *Journal of Shandong Normal University (Natural Science Edition)*, 2, 40-44.
- Yu, X. X., Wang, J. Y., & Yang, J. (2019). Analysis of the Current Situation of Psychological Health of Elementary School Students in Grades Three to Six in Guangxi. *China Health Education*, 12, 1089-1093.
- Zhang, D. J., & Su, C. (2015). The Development of Psychological Quality Scale for Elementary School Students. *Journal of Southwest University (Social Science Edition)*, 3, 89-95+ 190.
- Zhang, Y. M., Wang, Y. L., Zeng, P. P., & Yu, G. L. (2004). Confirmatory Factor Analysis of the Mental Health Scale for Elementary School Students. *Chinese Journal of Mental Health*, No. 4, 278-280.
- Zheng, Q. Q., Wen, J., Xu, F. Z., & Zhu, J. H. (2004). Exploring and Modifying the Structure of the Mental Health Diagnostic Test for High school Students. *Applied Psychology*, 2, 3-7.
- Zheng, R.C., Zhang, Y., & Liu, S.H. (2008). The Structure and Scale Development of Elementary School Students' Mental Health. *Educational Measurement and Evaluation*, 9, 30-34.
- Zheng, S. S., Wen, J., & Zheng, Q. Q. (2005). Construction and Administration of the Mental Health Diagnostic Questionnaire for High school Students. *Chinese Science of Behavioral Medicine*, 2, 21-23.
- Zhou, B. C. (1991). *Mental Health Diagnostic Test (MHT)*. East China Normal University Press.
- Zhu, N. N., & Zhang, H.C. (2003). A Comparative Study of CTT and IRT Methods for Processing Personality Test Results. *Psychological Inquiry*, 23, 48-51.