Available on: https://jsshrs.ep-journals.org/ E-ISSN: 3051-3561 P-ISSN: 3051-3553

Journal of Social Science and Human Research Studies

Volume: 02 Issue: 03 March-2026

Application of the GSP Chart and ROC Method to Develop A Grade 8 Natural Science Item Bank from Empirical Data Analysis

Nguyen Phuoc Hai*

Faculty of Pedagogy, Social Sciences and Humanities, Kien Giang University, Vietnam

Trinh Thi Kim Binh

Faculty of Food Sciences and Health, Kien Giang University, Vietnam

Published Date: March 09, 2026 **Article DOI: 10.65150/EP-jsshrs/V2E3/2026-02**

ABSTRACT: The objective of this study is to apply the GSP chart and ROC method to analyze and select Grade 8 Natural Science test items, thereby constructing an item bank to assess student learning outcomes. The Grade 8 Natural Science test was designed in accordance with Official Dispatch No. 7991/BGDĐT-GDTrH, issued by the Ministry of Education and Training of Vietnam on December 17, 2024, regarding the implementation of testing and assessment for lower and upper secondary education. The study was conducted with a sample of 116 students who took the Grade 8 Natural Science test under a new format consisting of four parts: Part 1 comprised 12 multiple-choice questions (four options with one best answer); Part 2 included 8 True/False questions; Part 3 consisted of 4 short-answer questions; and Part 4 contained 6 constructed-response (essay) questions. The research results demonstrated that MATLAB software can be effectively utilized to develop a toolbox for analyzing and selecting test items based on the GSP chart and ROC method. This toolbox facilitates rapid, accurate, and easy calculations, intuitively displaying results and visualizations via a Graphical User Interface (GUI), which can support teachers in building item banks to assess student learning outcomes.

KEYWORDS: GSP Chart, ROC Method, Item Bank, Natural Science, Educational Assessment.

1. INTRODUCTION

Assessing student learning outcomes is a crucial task in educational activities, reflecting the effectiveness, value, and quality of the teaching process. Testing and assessment are essential procedures for collecting information and comparing teaching objectives with actual outcomes to identify knowledge gaps and shortcomings. This process helps learners self-adjust their study methods while enabling instructors to refine their teaching strategies, thereby ensuring educational quality. To meet the demand for fair and objective assessments, and to innovate assessment methodologies, a combination of objective test items and constructed-response (essay) formats is commonly employed. In practice, from the initial drafting of test items to their application in assessing student outcomes, questions must be evaluated at various cognitive levels, ensuring the test comprehensively covers the intended curriculum. Currently, the use of item banks is highly encouraged in schools. However, most teacher-compiled test items do not strictly follow standard psychometric procedures. Specifically, items are rarely analyzed or calibrated before and after administration, often resulting in suboptimal and low-quality assessments. The findings of this study will serve as an essential resource to improve the quality and effectiveness of test compilation. Simultaneously, it will enhance teachers' skills in designing test items, thereby facilitating the construction of robust item banks.

In recent years, numerous studies have addressed the analysis and selection of test items using various methods. Nguyen Phuoc Hai *et al.* (2015) analyzed and selected objective test items based on the S-P chart, grey relational analysis (GRA), and the ROC curve. Nguyen Phuoc Hai (2016) proposed using the GSP chart and ROC method to analyze objective test items and assess student learning outcomes with a small sample size. Subsequently, Nguyen Phuoc Hai (2017) applied these methods to a larger sample size, comparing the outcomes with an item analysis conducted using BILOG-MG software. Currently, the application of the GSP chart, ROC method, and Item Response Theory (IRT) is gradually gaining traction in Vietnam, particularly for educational diagnosis and assessment. Furthermore, Doan Hong Chuong *et al.* (2016) utilized the "*ltm*" package in the R programming environment to analyze and evaluate objective test items. Bui Ngoc Quang (2017) employed Quest software to assess the quality of objective item banks. Bui Anh Kiet *et al.* (2018) and Pham Thi Minh *et al.* (2019) used IATA software to analyze, evaluate, and improve the quality of objective test items. Nguyen *et al.* (2020) analyzed and selected 50 objective test items for the English 1 final examination administered to 798 students at Dong Thap University during the 2019–2020 academic year using IATA software. By integrating analytical methods based on Classical Test Theory (CTT) and IRT, high-quality items were identified and incorporated into an item bank. More recently, Nguyen Phuoc Hai *et al.* (2024) analyzed, evaluated, and selected objective test items for an English 1 final exam based on multiple statistical tools, including BILOG-MG, R (*ltm* package), IATA, and GSP chart software combined with the ROC method. These studies demonstrate that such software programs can be used not only to analyze and select objective test items but also to improve the overall efficiency of item design and item bank construction. Each method and software tool possesses distinct advantages, aiming to identify high-quality items while flagging problematic ones for review and adjustment. However, the aforementioned studies focused primarily on analyzing and selecting items exclusively in the multiple-choice format.

In this study, the researcher proposes combining the GSP chart and ROC method to analyze and evaluate Grade 8 Natural Science test items, thereby constructing a reliable item bank. The test was designed by integrating both objective and constructed-response formats (objective formats include multiple-choice, True/False, and short-answer questions). Furthermore, the researcher utilizes MATLAB software to develop a specialized toolbox that facilitates the analysis and selection of test items based on the GSP chart and ROC method, fulfilling the current demands for educational assessment innovation. This toolbox enables rapid, accurate, and seamless calculations, intuitively displaying numerical results and visualizations via a Graphical User Interface (GUI). Consequently, it effectively supports teachers in building and managing item banks to assess student learning outcomes.

2. BASIC THEORIES AND RESEARCH METHODS

2.1 GSP Chart and ROC Method

GSP chart can make the analysis more concrete and accurate, and the uncertain factors in the studies can also be analyzed. In recent years, GSP chart has been used a lot in the field of education (Nguyen, 2016; Nguyen, 2017, Nguyen *et al*., 2017). Its description is shown in Table 1, X is a matrix with m rows and n columns, where $X_{ij} = 1$ if student answers item correctly and $X_{ij} = 0$ if the student answers the item incorrectly.

Student ID is $S_i, i = 1,2,...,m$; item ID is $P_j, j = 1,2,...,n$.

The ROC curve has a vertical axis that is the true positive ratio (sensitivity) and the horizontal axis is the false positive ratio (1 minus for specificity). Both ratios use probabilities to calculate and they have values that range from 0 to 1. According to many studies, the area under the ROC curve (*AUC*) is considered to distinguish well between positive and negative states when $AUC \geq 0.7$ (Kumar & Indrayan, 2011; Nguyen *et al*.,, 2018; Nguyen & Trinh, 2022; Tavakol & Dennick, 2012).

Table 1. The GSP chart and ROC method

	Item ID $P_j,\ j = 1,2,\cdots,n$	**Total Score**	***CS***	***GS***	***AUC***
Student ID $S_i,\ i = 1,2,\cdots,m$	$X = [x_{ij}]_{m \times n}$	High ↕ Low	CS_i	GS_i	$AUCi$
Number of Correct Answers	More ↔ Less	—	—	—	—
CP	CP_j	—	—	—	—
GP	GP_j	—	—	—	—
AUC	$AUCj$	—	—	—	—

This study focuses on the application of the GSP chart and ROC method to analyze and evaluate Grade 8 Natural Science test items under a new format consisting of four parts: Part 1 comprises 12 multiple-choice questions (four options with one best answer); Part 2 includes 8 True/False questions; Part 3 consists of 4 short-answer questions; and Part 4 contains 6 constructed-response questions. In this study, the researcher employed a combination of the following methods:

Data Collection and Statistical Methods: Collecting literature and data related to the analysis and evaluation of test items in Natural Science education, as well as classifying and statistically analyzing the empirical results.

Analytical and Synthesis Methods: Based on the collected literature, empirical data, and expert opinions, the researcher analyzed and synthesized information to propose the application of the GSP chart and ROC method for analyzing and evaluating Grade 8 Natural Science test items.

Expert Consultation Method: During the application process of the GSP chart and ROC method, the researcher consulted with experts in educational measurement and evaluation, alongside lecturers and teachers at various educational institutions.

Empirical Research Method: Data was collected from administered Grade 8 Natural Science tests. Subsequently, the GSP chart and ROC method were applied to analyze and evaluate the test items, thereby constructing an item bank to assess student learning outcomes.

2.2 Building MATLAB Toolbox

Software Specifications and Requirements

a) Windows 10 or upgrade versions.

b) Screen resolution: 1920 × 1080 pixels.

c) MATLAB R2021b version or upgrade versions.

To facilitate the item analysis process, a specialized MATLAB-based program was developed. This program integrates advanced scientific computing functions to create a reliable environment for data processing. Operating this tool requires basic proficiency in MATLAB (Nguyen & Trinh, 2022; Nguyen & Trinh, 2023; Nguyen *et al*., 2017; Nguyen *et al*., 2015).

The Program for Test Item Analysis

The main algorithms of the program are described in the flowchart (Figure 1). This program consists of 6 basic steps:

Step 1: Input data. Data is the response matrix of students that is numerical and written in *.csv file or *.xlsx file.

Step 2: Test data. The program processes and tests the valid input data before proceeding to the next analysis phases.

Step 3: GSP Chart. First, calculate the total number of students and items. Next, compute the Student and Item Caution Indices (CS_i and CP_j). Then, determine the Student and Item Gamma Indices (GS_i and GP_j) and design the results.

Step 4: ROC Method. Initially, determine the positive and negative states. Next, calculate the values *a, b, c* and *d*. Then, compute the Area Under the ROC Curve (AUC) and design the results and graphs.

Step 5: Output Results and Graphs. The program outputs the designed numerical results and visual graphs for further evaluation.

Step 6: Continue or exit program. If new data is selected, the program will continue and back to step 1, or else the program will be ended. The user can save the results and the graphs to *.csv file or *.xlsx file and jpeg image formatted.

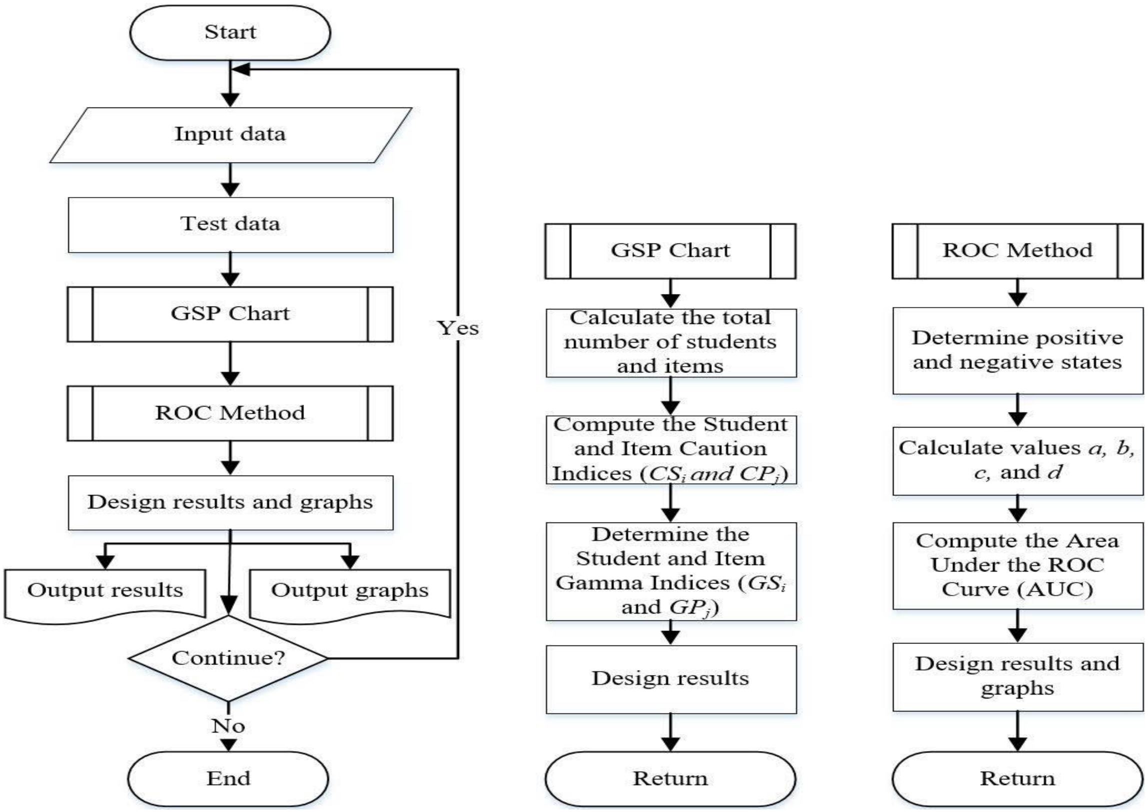

Figure 1. The flowchart for test item analysis via the GSP chart and ROC method

3. RESEARCH RESULTS

3.1 Test Structure Description

On December 17, 2024, the Ministry of Education and Training of Vietnam issued Official Dispatch No. 7991/BGDĐT-GDTrH regarding the implementation of testing and assessment for lower and upper secondary education. The Grade 8 Natural Science test in this study was administered in accordance with this Official Dispatch. The assessment structure employed a combination of objective test items and constructed-response components, accounting for 70% and 30% of the total score, respectively. Furthermore, the test items were stratified according to cognitive domains, with recognition, comprehension, and application levels allocated at ratios of 40%, 30%, and 30%, respectively. The specific components are as follows:

Part 1: Multiple-choice questions (3 points). This section consists of 12 questions, each providing four options with one correct answer. Each correct response is awarded 0.25 points.

Part 2: True/False questions (2 points). This subsection consists of 2 questions structured in a True/False format. Each question contains four sub-components (labeled a, b, c, and d); the examinee must evaluate the validity (True or False) of each sub-component. Each correctly identified sub-component is awarded 0.25 points.

Part 3: Short-answer questions (2 points). This section comprises 4 short-answer questions. Students are required to write or input their results as specified by the prompt. Each correct answer is awarded 0.5 points.

Part 4: Constructed-response questions (3 points). Examinees present their responses on paper or input content via a computer interface. This section consists of 3 questions, each containing two sub-parts (labeled a and b). Each correct sub-part is awarded 0.5 points.

3.2 Data Coding and Conversion

The following table outlines the conventions used to convert assessment results for item analysis and evaluation purposes.

Table 2. Item Coding and Data Transformation for Item Analysis and Evaluation

Test Structure (10-point scale)	Test Item Coding Convention	Data Transformation for Item Analysis and Evaluation
Part 1: Multiple-choice (3 points)	Consists of 12 questions (coded from 1 to 12)	For Parts 1, 2, and 3: Each correct response by the examinee is coded as 1; an incorrect response is coded as 0.
Part 2: True/False (2 points)	Consists of 2 questions (each comprising 4 response items, coded 13 to 20)	
Part 3: Short-answer (2 points)	Consists of 4 questions (coded from 21 to 24)	
Part 4: Constructed-response (3 points)	Consists of 3 questions (each comprising 2 response items, coded 25 to 30)	For Part 4: If the examinee's response meets the requirements, it is coded as 1; if it fails to meet requirements, it is coded as 0.

3.3 Empirical Data Analysis

The data in this study consist of the response results from 116 Grade 8 students on the Natural Science test. Prior to conducting the item analysis and selection, the researcher examined the reliability of the data using Cronbach's Alpha coefficient, which yielded a value of 0.876 (Cronbach, 1951). Furthermore, the data reliability was also evaluated using the Kuder–Richardson Formula 20 (KR20) and Formula 21 (KR21) coefficients (Foster, 2021). The calculated KR20 and KR21 values were 0.876 and 0.864, respectively. These results indicate that the data possess high reliability and are highly suitable for proceeding with the analysis and evaluation of the test items. A portion of the empirical data utilized in this study is presented in Table 3.

Table 3. The results of answering 30 items of 116 students (A portion of the empirical data)

S\P	Multiple-Choice												True/False								Short Answer				Constructed Response					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0
3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0
4	1	1	1	1	0	1	1	0	1	1	0	0	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	0	1	0
5	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	1	0
6	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	0	1
7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	0	1	1	1	0	0	1	0	0	0	1
9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	0	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1
12	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	0	0	1	1
13	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	0	0	1	1	1	1	1	0	1	0
14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	0	0	1	1	1	1	1
---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
102	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	0	0	0
103	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0
104	1	0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0
105	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	0	0	1	0	0	0	0	0	1
106	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1	0	1	1	1	0
107	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0
108	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	0
109	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0
110	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	1	1	0	1
111	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	0	1	0	1	0	1	1	1	1	1	0
112	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	0	0	1	0	1	0	1	1	0	1	0	1
113	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	0	1
114	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
115	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0
116	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0

Table 4. The analytical results of 30 items via the GSP chart and ROC method

Item ID	Total	Ratio	*CP*	*GP*	*AUC*	Results	Item ID	Total	Ratio	*CP*	*GP*	*AUC*	Results
1	109	0.94	0.09	0.91	0.92	*Qualified*	**16**	100	0.86	0.42	0.62	0.71	*Qualified*
2	108	0.93	0.11	0.87	0.93	*Qualified*	**17**	98	0.84	0.29	0.57	0.77	*Qualified*
3	111	0.96	0.00	1.00	1.00	*Qualified*	**18**	94	0.81	0.72	0.48	0.55	*Unqualified*
4	108	0.93	0.52	0.87	0.66	*Unqualified*	**19**	95	0.82	0.41	0.50	0.74	*Qualified*
5	106	0.91	0.51	0.80	0.62	*Unqualified*	**20**	92	0.79	0.30	0.43	0.76	*Qualified*
6	107	0.92	0.20	0.84	0.76	*Qualified*	**21**	98	0.84	0.66	0.57	0.64	*Unqualified*
7	108	0.93	0.63	0.87	0.66	*Unqualified*	**22**	87	0.75	0.31	0.33	0.72	*Qualified*
8	106	0.91	0.34	0.80	0.73	*Qualified*	**23**	86	0.74	0.50	0.31	0.71	*Qualified*
9	106	0.91	0.39	0.80	0.73	*Qualified*	**24**	78	0.67	0.30	0.16	0.75	*Qualified*
10	108	0.93	0.43	0.87	0.66	*Unqualified*	**25**	82	0.71	0.41	0.23	0.71	*Qualified*
11	105	0.91	0.51	0.77	0.65	*Unqualified*	**26**	85	0.73	0.56	0.29	0.58	*Unqualified*
12	103	0.89	0.24	0.71	0.83	*Qualified*	**27**	86	0.74	0.20	0.31	0.82	*Qualified*
13	100	0.86	0.25	0.62	0.71	*Qualified*	**28**	82	0.71	0.59	0.23	0.67	*Unqualified*
14	100	0.86	0.30	0.62	0.82	*Qualified*	**29**	75	0.65	0.41	0.11	0.64	*Unqualified*
15	99	0.85	0.24	0.60	0.76	*Qualified*	**30**	68	0.59	0.57	0.00	0.63	*Unqualified*

From the results displayed on the Graphical User Interface (GUI) of the MATLAB toolbox for test item analysis based on the GSP chart and ROC method, users can observe the computed *CP, GP*, and *AUC* values of the items, along with the visual graphs of the *CP* and *GP* values and the item ROC curves during the analysis process. Based on the item analysis results utilizing the GSP chart and ROC method (Figure 2 and Table 4), if the Epsilon threshold is set to 0.7, 11 items are not accepted in this study (Items 4, 5, 7, 10, 11, 18, 21, 26, 28, 29, and 30).

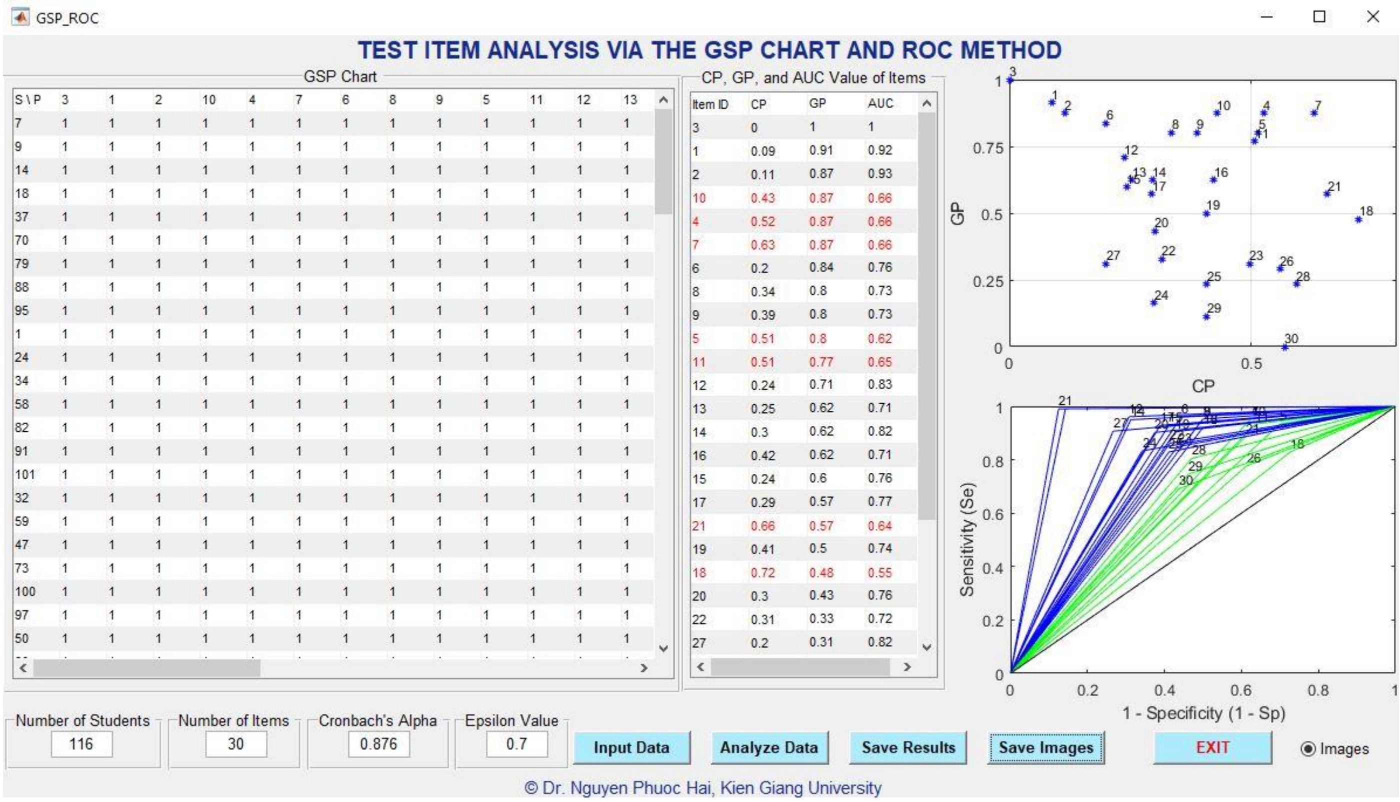

Figure 2. Graphical user interface for analyzing 30 test items in 8th grade Natural Science

4. DISCUSSION

The proposed application of the GSP chart and ROC method for analyzing and selecting Grade 8 Natural Science test items, thereby constructing an item bank to assess student learning outcomes, is demonstrated to be both suitable and reliable. Consequently, it can be effectively applied to develop test item banks for other subjects in accordance with Official Dispatch No. 7991/BGDĐT-GDTrH, issued by the Ministry of Education and Training of Vietnam. Based on the results presented in Table 4, it is evident that a higher *CP* coefficient corresponds to a lower *AUC* value, and vice versa. A high *CP* coefficient coupled with a low *AUC* value indicates that the corresponding test item possesses low discrimination power. This phenomenon occurs when low-ability students manage to answer difficult items correctly. Therefore, this research method facilitates the identification and elimination of test items with high *CP* coefficients and low *AUC* values. Additionally, the MATLAB toolbox utilized in this study proves to be highly useful and convenient, assisting in visually locating test items based on the Item Caution Index

(*CP*) and the *GP* value (Figure 2). This functionality significantly aids in the removal of test items with a *CP* coefficient greater than 0.5. Furthermore, this method is also capable of computing the Student Caution Index (*CS*), *GS* value, and *AUC* value for individual students. Based on these metrics, the academic proficiency of each student can be accurately determined. Moreover, it provides teachers with essential diagnostic information for a more precise assessment of student learning outcomes, enabling them to promptly propose measures to adjust teaching activities. This ultimately fulfills pedagogical objectives and meets the demands of current educational innovation.

5. CONCLUSIONS

The research results indicate that the GSP chart and ROC method can be applied to analyze and select test items for subjects implemented in accordance with Official Dispatch No. 7991/BGDĐT-GDTrH, issued by the Ministry of Education and Training of Vietnam on December 17, 2024, regarding the implementation of testing and assessment for lower and upper secondary education. This study will be highly beneficial for educational administrators, lecturers, teachers, and all those interested in seeking a method to enhance the efficiency of constructing item banks to assess student learning outcomes, thereby meeting the requirements of the 2018 General Education Program. Furthermore, this research has expanded and optimized a MATLAB toolbox for test item analysis based on the GSP chart and ROC method, specifically accommodating test structures that integrate various objective and constructed-response formats. On the Graphical User Interface (GUI) of the MATLAB toolbox, the positions of test items can be visually located based on the Item Caution Index (*CP*) and the *GP* value, making the analysis and selection of items significantly easier. Ultimately, the findings demonstrate that educators can utilize this method to evaluate the quality of test items and effectively apply it to build item banks for assessing student learning outcomes.

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the experts who dedicated their time to share their expertise on the analysis, evaluation, and development of item banks for assessing student learning outcomes. Furthermore, we extend our deepest appreciation to the teachers and students of Dong Hung A Secondary School for providing the empirical data and offering valuable feedback on the analysis, evaluation, and construction of the item bank. Their insightful contributions have been instrumental in refining our ideas and enhancing the overall quality of this research.

REFERENCES

1) Bui, A. K., & Bui, N. P. (2018). Using IATA to analyze, assess and improve the quality of the multiple-choice questions in chapter power functions, exponential functions and logarithmic functions. *Can Tho University Journal of Science*, *54*(9), 81–93.
2) Bui, N. Q. (2017). Assessment of the quality of multiple choice test bank for the module of Introduction to Anthropology by using the RASCH model and QUEST software. *Science of Technology Development*, *20*(X3), 42–54.
3) Cronbach, L. J. (1951). *Coefficient alpha and the internal structure of tests*. Psychometrika, 16(3), 297-334.
4) Doan, H. C., Le, A. V., & Pham, H. U. (2016). Applying 3-parameter logistic model in validating the level of difficulty, discrimination and guessing of items in a multiple choice test. *Ho Chi Minh city University of Education Journal of Science*, *7 (85)*, 174.
5) Foster, R. C. (2021). KR20 and KR21 for some nondichotomous data (it's not just cronbach's alpha). *Educational and Psychological Measurement*, *81*(6), 1172-1202.
6) Kumar, R., & Indrayan, A. (2011). Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatrics*, *48*, 277–287.
7) Nguyen, P. H. (2016). Using GSP chart and ROC method to analyze mutiple-choice test items and assess learning outcomes of students. *Journal of Education Science, Vietnam Institute of Education Science*, *134*(11), 32–37.
8) Nguyen, P. H. (2017). Using GSP chart and ROC method to analyze and select mutiple-choice test items. *Dong Thap University Journal of Science*, *24*(2), 11–17.
9) Nguyen, P. H., Nguyen, V. C., & Trinh, T. K. B. (2024). Assessing the Quality of Multiple-Choice Test Items Based on BILOG-MG, R (*ltm*), IATA, and GSP-ROC. *WVSU Research Journal*, *13*(1), 1-12.
10) Nguyen, P. H., & Du, T. N. (2015). The analysis and selection of mutiple-choice test items based on S-P chart, Grey Relational Analysis, and ROC curve. *Ho Chi Minh city University of Education Journal of Science*, *6 (72)*, 163.
11) Nguyen, P. H., Ta, P. H., & Trinh, T. K. B. (2018). Building the standard test score based on the combination of T-GM(m,n), GRA and ROC method. *Dong Thap University Journal of Science*, *35*, 10–17.
12) Nguyen, P. H., & Trinh, T. K. B. (2017). Assessment of students' learning outcomes a combination of GSP chart and ROC method. *AGU International Journal of Sciences*, *17*(5), 103–112.
13) Nguyen, P. H., & Trinh, T. K. B. (2022). Assessment of Students' Learning Outcomes in Higher Education. *International Journal of Uncertainty and Innovation Research*, *4*(1), 39–52.
14) Nguyen, P. H., & Trinh, T. K. B. (2023). Researching and Building A Computer-Based System for Testing and Assessing Learners' Ability (A Case Study in Preschool Teacher Training). *TNU Journal of Science and Technology, 229*(04), 124–132.
15) Nguyen, P. H., & Trinh, T. K. B. (2024). Developing A Computer-Based System for Testing and Assessing Learners' Ability: A Case Study in General Education. *Vinh University Journal of Science, 53*, 331–340.
16) Nguyen, P. H., Nguyen P. T., Ho C. P., Trinh, T. K. B., & Nagai M. (2017). The Prediction of the Admission Teacher's Number in Taiwan by using T-GM(1,n) and T-GM(2,n) Method. *Journal of Grey System*, *20*(3), 139-150.
17) Nguyen, P. H., Sheu, T. W., & Nagai, M. (2015). Predicting the Student Learning Outcomes Based on the Combination of Taylor Approximation Method and Grey Models. *VNU Journal of Science: Education Research*, *31*, 70-83.

Corresponding Author: Nguyen Phuoc Hai

18) Nguyen, V. C., & Nguyen, P. H. (2020). Analyzing and Selecting Multiple-Choice Test Items Based on Classical Test Theory and Item Response Theory. *Ho Chi Minh city University of Education Journal of Science*, *17*(10), 1804.
19) Pham, T. M., & Bui, Đ. N. (2019). The IATA software for analyzing, evaluation of multiple-choice questions at Ha Noi Metropolitan University. *Scientific Journal of Ha Noi Metropolitan University*, *20*(2019), 97–108.
20) Tavakol, M., & Dennick, R. (2012). Standard setting: The application of the receiver operating characteristic method. In *Int J Med Educ* (Vol. 3, pp. 198–200). IJME.

Corresponding Author: Nguyen Phuoc Hai