

THE DESIGN AND DEVELOPMENT OF A SEMI-AUTO COMPUTER GENERATED TESTING PAPER SYSTEM — A CASE STUDY IN THE SCHOOL OF CONTINUING EDUCATION AT CHINA UNIVERSITY OF GEOSCIENCES

Juan YANG,
Tsinghua University
juan-yang@tsinghua.edu.cn

Xibin HAN
Tsinghua University
hanxb@tsinghua.edu.cn

Qian ZHOU
Tsinghua University
zhouqian@tsinghua.edu.cn

ABSTRACT

With the rapidly increasing entrance examinations for the School of Continuing education of CUGB, there is an urgent need of the examination papers that have the parallel testing ability for the same academic program. Based on the analysis of parallel factors and the comparison of the manual generated mode and the computer generated mode, the most time-consuming and experience-consuming steps are found. By providing supports to those steps automatically, a semi-auto computer generated testing paper system was designed and developed for reducing the workload of the experienced faculties to generate testing papers. The papers were generated based on a gradually constructed item-bank. The system offered multiple templates of the testing paper structure plan for assisting semi-auto paper generated procedure and guaranteeing the parallel of the generated papers. In order to construct an item-bank in a short time, an easy-to-use authoring tool with powerful functions was designed and developed. The system supported the spring entrance examination of SCECUGB in 2010 successfully. Faculties gave positive feedback toward the system.

Keywords: semi-auto computer generated testing paper, item-bank, manual generated testing paper, test assembly

INTRODUCTION

As a pilot distance education school in China, the School of Continuing Education at China University of Geosciences in Beijing (SCECUGB) recruits students twice a year nationally. A batch of web-based distance education schools set up in recent years aggravated the recruit competition of distance education. SCECUGB has to expand the scope of enrollment. Gradually, part-time students instead of full-time ones become the main target recruit groups. The target recruit groups require to be enrolled at any time instead of the fixed time. SCECUGB has been forced to shorten the interval and increase the times of the entrance examinations. A national examination requires a set of parallel testing papers (Hwang, Chu and Yin 2008). The greatly increasing entrance examination requires a large amount of parallel testing papers for the same academic program. Paper generation is a time-consuming and resource-consuming work. It is impossible to generate those rapidly increasing papers manually as before because of the relative shortage of experienced faculty members.

As the result, computer generated paper system can solve the problem stated above for its convenient, safe and economic functions. However, those working systems have a common feature that a mature testing item-bank is associated. Unfortunately, there is not an item-bank of this kind in the case of SCECUGB. In this study, based on the analysis of parallel manual paper generation procedure and common computer paper generation procedure, the most time-consuming and experience-consuming steps are found. Consequently, by providing support to those steps, an altered semi-auto computer paper generation procedure is proposed. Based on this altered procedure, a semi-auto computer generated testing paper system is designed and developed, trying to provide supports assisting paper generation.

In order to provide sufficient supports for faculties alleviating their heavy work load, following issues should be solved in this paper.

(1) How to generate parallel papers?

Finding the most time-consuming steps and trying to finish it automatically.

(2) What is the most time-consuming step in the paper generation procedure?

Analysis in the manual paper generated procedure reveals that to construct Two-Way Charts indicting the paper structure plan is the most time-consuming and experience-consuming task.

(3) How to provide sufficient support for paper generation?

Based on embedded template paper structure plans and automatically generated paper structure plans, item selection in the computer generated mode and item compiling in the manual generated mode are simplified and greatly supported.

LITERATURE REVIEW

Computer generated testing paper system is a subsystem of a Computer Based Testing(CBT) System. Computerized tests are classified into two categories in the literature (e.g. Lilley,2004; López-Cuadrado,Pérez,Vadillo and Gutiérrez,2010): computer-based tests (CBTs) and computerized adaptive tests (CATs). The former ones are computerized fixed form tests which are similar to pen-and-pencil tests.

No matter based on CBTs or CATs, researches on the computer generated testing paper system usually focus on the algorithm optimization to achieve the fast and precise retrieval (Feng, 2010; Yu, 2008; Yuan, 2008; Hwang, Lin and Lin,2006;Chou,2000).

When a system is based on a gradually constructed item-bank, the algorithm optimization is not the key point. The main considerations to construct such systems should be the parallel generation of the testing papers in different entrance examinations for the same academic program, the procedure of paper generating, and the assistance for faculties to generate the papers efficiently.

Parallel Generation

The widespread use of computerized testing is blocked by following factors: (1) limited technological capacity (e.g. accessing computer and internet) to support computerized testing in some testing centers sometimes (Thurlow, Lazarus, Albus and Hodgson,2010), (2)candidates short of computer skills (3) inconvenience in answering items of some types (e.g. calculation questions in Mathematics) in computer. Accordingly, computerized testing and pen-and-pencil testing are carried out in parallel. Therefore, generated papers are all fixed testing papers, the same as pen-and-pencil trials. Thence, papers should not only be parallel in content but also should be parallel in form.

Parallel in form. The form of the testing papers in different entrance examinations for the same academic program should be paralleled in order to avoid the interference of themselves to the test results (Ma and Ding, 2006). The parallel form means the same item quantity, fixed testing time and the same structure: consisting of a set of ordered sections grouped by item-type.

Parallel in content. The content of the testing papers should also be parallel. According to the knowledge structure of the exam requirement (Fig 1), the proportion and the difficulty of testing items related to each part of the knowledge are set equally in different testing papers. For example, Fig 1 is a knowledge structure for the exam of Advanced Mathematics, which includes three layers. In layer 1, the proportions of knowledge components are described. In layer 2, the acknowledge components are divided into smaller pieces. In layer 3, all the knowledge pieces are shown as basic testing units with testing objectives. Zhao, Wang and Zhao (2010) described the standard of the content parallel in different papers. In layer 1, the testing items related to the knowledge components should be equal proportion. In layer 2, the testing items should cover 80% of the knowledge pieces. In layer 3, the testing items related to the basic testing units should be selected randomly, and the difficulty level should be matched with the testing objectives.

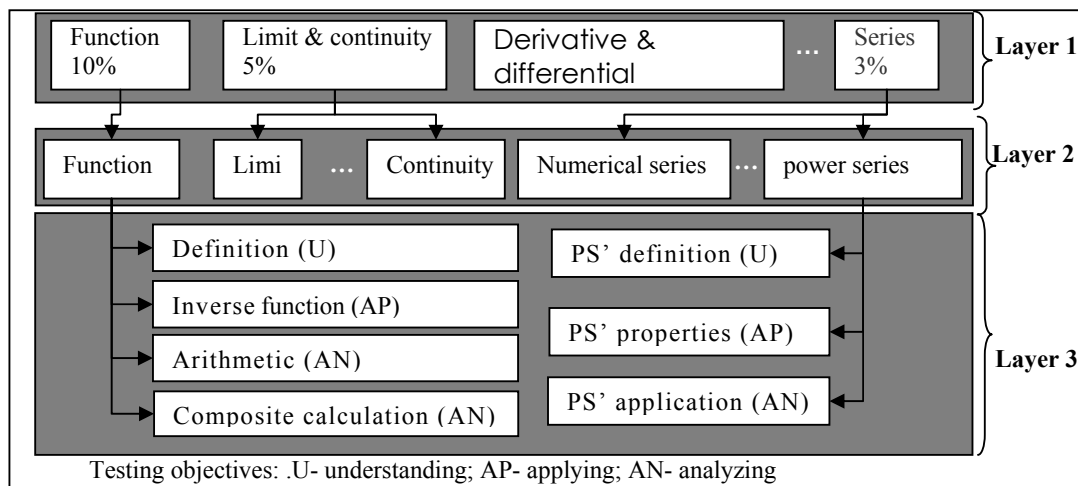


Figure 1. Knowledge Structure of the Advanced Mathematics

Manual Generated vs. Computer Generated

Manual Generated Procedure. Liu (2006) described the manual generated procedure (Fig 2). Faculties of SCECUGB confirmed the procedure. Besides, they specified the steps in the procedure (Fig 3). Firstly, analyzing the exam requirement and determining the purpose of testing. And then secondly, specifying it into a group of Two-Way Charts indicating the paper structure plan (Table 1,2). Thirdly, compiling the items and filling them into each type section according to the Two-Way Charts. Finally, checking the paper and its parallelism with other papers used in the same program, and jumping back to step 3 to revise the paper. The entire procedure is very complicated and time consuming, because the factors of item-type distribution, knowledge-component distribution, and difficulty distribution should be considered synthetically. Step 2 is recognized as the most tough and time-consuming step by faculties. Even faculties with the richest paper-generation experiences have to revise the paper structure plan constantly in order to achieve the final balance. After the paper structure is determined, the compilation and filling of the items become relatively simple.

Computer Generated Mode. Fig 4 shows a Two-Way Chart based computer paper generated procedure, which is similar to manual paper generated procedure and all steps are adopted from “basic steps in classroom testing” proposed by Gronlund and Linn(1990). However, paper generated procedures in most systems do not support Two-Way Chart(Wang,2008). Therefore, the procedure is revised and the computer paper generated mode is showed in Fig 5.

Contrast to the manual mode, in computer generated mode (Fig 5), the procedure is composed by only 3 steps. Step 1 is the same as that in manual mode. While step 2 is completely different, the paper structure plan is specified automatically based on the parameters input in manual. But the plan is simpler than that in manual mode. In the selection process (step 3), item searching and filling in, paper structure plan revising and paper checking & revising are intertwined. Therefore, in the computer mode, partial paper structure planning process (the parts marked by * in Fig 3) and the whole checking and revising process are shifted into the item selection process.

In computer generated mode, selection algorithm optimization is emphasized. The purpose of the algorithm optimization is to increase the precision and the speed of item selection with some initial conditions given. If there are sufficient amount of testing items in the item-bank, under the aid of appropriate selection algorithm, papers fitting the requirement can be generated fast and easily.

This mode highlights following advantages. 1) simple input 2) fast speed 3) precision

Step 1: Determining the purpose of testing
 Step 2: Constructing the paper structure planning (Two-Way Chart)
 Step 3: Compiling and filling items according Two-Way Chart
 Step 4: Checking and Revising

Figure 2. The General Manual Generated Procedure (Liu,2006)

Step 1: Determining the purpose of testing
 (1) Specifying knowledge structure
 (2) Setting the proportion of knowledge components in layer 1
 (3) Defining testing objective of each basic unit in layer 3
 Step 2: Constructing the paper structure planning (Two-Way Charts)(**manual**)
 (1) Specifying type plan: scores, testing time and item's quantity of each type (see Table 1)
 (2) Specifying item type assignment according to difficulty level(see Table 1)
 (3) Specifying item type assignment according to knowledge(concept) components (see Table 2)
 Step 3: Compiling and filling items according Two-Way Charts(**manual**)
 Step 4: Checking and Revising(**manual**)
 (1) Checking papers
 (2) Checking parallel of the papers with the others
 (3) Returning the above the steps

Note: this procedure is extracted from interview with faculties of SCECUGB.

Figure 3. The Procedure of Manual Generated Mode(extracted from the interview with faculties of CUGB)

Table 1: Two-Way Chart Indicating Type Plan

Item type	Multiple Choice	Fill-in Blank	True/False	Calculation	Integrated Question	Sum
Item quantity	15	20	5	4	2	46
score	1(per item) 15(total)	2(per item) 20(total)	
Testing time (minutes)	15					
Difficulty	A	8				
	B	7				
	C					
	D					
Testing Objectives	I	8AI 7BI	5AI			
	II		1AII 7BII 1CII			
	III		4CIII			
	IV		2DIV			
sum						

Note: Difficulty : A. Easy, B. Medium, C. Difficult, D. Very Difficult

Testing Objectives: Bloom's taxonomy, I Remembering, II Understanding III Applying IV Analyzing (Anderson et al. , 2001)

Table 2: Two-Way Chart indicating Item Type Assignment according to Knowledge Components

Knowledge Item type		Bloom's taxonomy ^a			sum score
		Understanding (score)	Applying (score)	Analyzing (score)	
Concept 1	Multiple Choice	3			3
Concept 2		3			3
.....				
Concept 1	Fill-in Blank		4		4
Concept s				4	4
.....				
Concept m	True/False		4	4	8
Concept n			4	4	8
.....					
Concept 1			4		8
Concept a			4		8
.....			4		8
.....					
Sum score		9	34	28	100

Note: ^a Bloom's taxonomy of cognitive domain (Anderson et al., 2001)

-
- Step 1: Determining the purpose of testing
 Step 2: Constructing the two-way chart
 Step 3: Selecting appropriate items according to the Two-Way Chart
 Step 4: Preparing relevant items
 Step 5: Assembling the test
-

Figure 4. The Computer Generated Procedure (Gronlund et al.,1990)

-
- Step 1: Determining the purpose of testing
 (1) Specifying knowledge structure
 (2) Setting the proportion of knowledge components in layer 1
 (3) Defining testing objective of each basic unit in layer 3
-

-
- Step 2: Constructing the paper structure planning (**semi-automatic**)
- (1) Specifying the general type plan: scores, testing time and item's quantity of each type (**manual**)
 - (2) Constructing the paper structure planning (see Table 1).(**automatic**)
- Step 3: Selecting appropriate items according to the criteria inputted in step 2(**automatic**)
- (1) Searching and filling in the items
 - (2) Revising the paper structure plan
 - . Type and item assignment according to knowledge components*
 - . Type and item assignment according to difficulty level *
 - (3) Checking and revising
- Note: * task belonging to the paper structure planning process in the manual generation mode
-

Figure 5. The Procedure of Computer Generated Mode

THE DESIGN OF A SEMI-AUTO COMPUTER GENERATED TESTING PAPER SYSTEM

The Semi-Auto Computer Generated Mode

In the case of SCECUGB, there is no mature item-bank. In order to continue the automatic generation procedure when short of items, a new mode, other than manual and automatic computer generated ones, is designed and developed in this study. It is called Semi-Auto Computer Generated mode. In this mode, the procedure is more similar to manual mode, while most of the steps are processed automatically just like those in computer generated mode.

In this mode (Fig 6), step 1 is the same as that in manual mode and computer generated mode. In step 2, the first executed three tasks of the paper structure planning are the same as those in the manual mode, and the fourth task, assigning items of different types to knowledge components (see Table 3), is newly added in the paper structure plan. All the four tasks are computer generated automatically based on the exam requirement analysis. In item selection step(step 3), if short of the items, semi-auto mode starts to work and the procedure jumps to step 4. In step 4, the system requires the faculties to compile and fill in the items according to the specific criteria manually. Otherwise, the item compiling and filling in are executed automatically, which are the same as that in the computer mode.

This mode highlights two advantages obviously. 1) With the separation of the item selection and paper structure planning, the process of paper generation could continue when short of items. 2) Computer finished most of the time consuming tasks(constructing the paper structure planning) in the manual mode.

-
- Step 1: Determining the purpose of testing
- (1) Specifying knowledge structure
 - (2) Setting the proportion of knowledge components in layer 1
 - (3) Defining testing objective of each basic unit in layer 3
- Step 2: Constructing the paper structure planning (**automatic**)
- (1) Specifying Type plan: scores, testing time and item's quantity of each type(see Table1)
 - (2) Specifying Item type assignment according to difficulty level. (see Table1)
 - (3) Specifying Item type assignment according to knowledge components(see Table2)
 - (4) Specifying Assigning items of different types to knowledge components *
- Step 3: Selecting appropriate items according to the Two-Way Chart(**automatic**)
- Step 4. Items compiling & filling in (**manual**) **
- Step 5. Checking & revising (**automatic**)

Note: * task added in the mode
 ** step can be omitted if item selection is successful

Figure 6. The Procedure of Semi-Auto Computer Generated Mode

Table 3: Assignment of Items to Knowledge Components in the Paper Structure Plan

Item type	Item	Knowledge components
Multiple choice	Item Mc ₁	Component M ₁ .
	Item Mc ₂	Component M ₂

	Item Mc ₂₀	Component M ₂₀
Fill-in Blank	Item Ms ₁	Component S ₁ .

	Item Ms ₁₀	...
True/False	Item M _{t1}	Component M ₁

Calculation	Item C ₁	Component C ₁ , Component C ₂ ...

	Item C ₅	...
Integrated question	Item I ₁	Component I ₁ , Component I ₂ ...
	Item I ₂	Component I _m , Component I _n ...

The Construction of the Item-Bank

The construction of an item-bank includes three databases: knowledge structure, paper structure plan, and testing items.

(1) The database of the knowledge structure. The knowledge structure (see Fig 1) described based on the exam requirement is recorded in the database. The description of the proportion of the knowledge components in layer 1 is recorded as well. Also, the description of the testing objective of each basic unit in layer 3 is recorded.

(2) The database of paper structure plan. In the database, paper structure plans (see Table 1,2) are recorded, the fields of which include testing time, difficulty level, item type proportion, knowledge component proportion of each item type, difficulty level of each item type, as well as assignment of items of different types to knowledge components (Table 3). The templates of the plans are defaulted based on the previous manually generated papers.

(3) The database of the testing items. The metadata and the entities of the testing items are recorded in the database. The metadata is defined by China E-Learning Technology Standardization Committee [CELTC] (2003) in the Technical Specification for Educational Resource Construction Information Model (CELTS-41.1 CD1.0, 2003). It includes manually input attributes (such as, item type, difficulty, discrimination, exam requirement, criterion, score, secrecy) and self-generated attributes (such as, modified difficulty, modified discrimination, used time, exposal date). Item attributes are used to aid the item selection. If an item covered two or more knowledge units, the ratio should be indicated. Therefore, items in the database are connected and organized based upon the knowledge structure quantitatively.

Item Authoring and Storage

Item-bank of SCECUGB is gradually built. An authoring tool is designed to support the faculties to do testing items editing including formula editing and text-graphics integration. The authoring tool is developed based on the comparison of three technical solutions (Table 4). Finally, the solution of plug-in based on MS Word is chosen, because it can provide powerful functions with user-familiar interface and requires less work load to develop. The authoring tool developed based on the solution is called TheoEditor.

Table 4: The Developing Solutions of the Authoring Tools

Solution		Advantages	Disadvantages
Authoring tool based on web browser		No plug-ins installation required	Hard to develop, poor performance in text-graphics integration
Client authoring tool	Self-developed plug-in	Text-graphics integration supported	Limited functions, installation required
	A plug-in based on MS Word	High performance in functions, user-familiar interface	Installation required

Figure 7 shows the process of the editing testing items. With the help of Browser-Helper-Object (BHO, browser Interactive interface opened for third-party programmers), TheoEditor will be started while web browser tries to access some specific URL. Next, TheoEditor invokes the client-installed MS Word to edit an item file. Then,

TheoEditor saves the edited file and exit MS Word. The file is saved into two copies, one as doc format used for further editing, and the other as html format used for viewing. Finally, TheoEditor uploaded the files to the server.

TheoEditor supported batch-input of items by tagging specific defined symbols on the items in the Word. All the items with title, question, answer and metadata are tagged separately. TheoEditor identifies and extracts the metadata and entities of the items from the Word file, and save them into the item-bank.

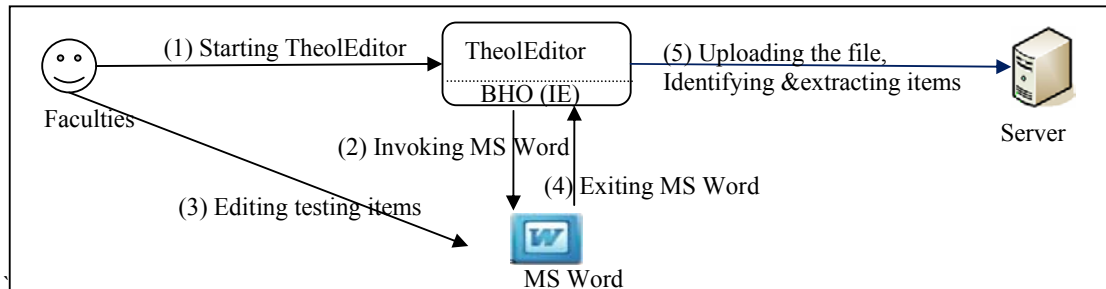


Figure 7. TheoEditor Editing Items Process (Guan, Han, Zhou and Shen, 2008)

Generation and Utilization of the Paper Structure Plan

(1) Generation of the Template Paper Structure Plan

Some paper structure plans are recorded as templates in the item-bank. These template plans are reverse-generated automatically from the testing papers which either are compiled manually, or computer generated in the past entrance examinations. Every paper structure plan is composed by two multidimensional Two-Way Charts (Table 1,2). Figure 8 shows the automatic reverse-generation procedure: the corresponding relationships between items of different types and knowledge components (Table 3) are extracted from the papers, and then the elements of a paper structure plan (Table 1,2) are generated based on the relationships, which includes testing time, difficulty level, item type proportion, knowledge component proportion of each item type and difficulty level of each item type. The parallel in form and content of the templates are checked automatically based on the exam requirement analysis.

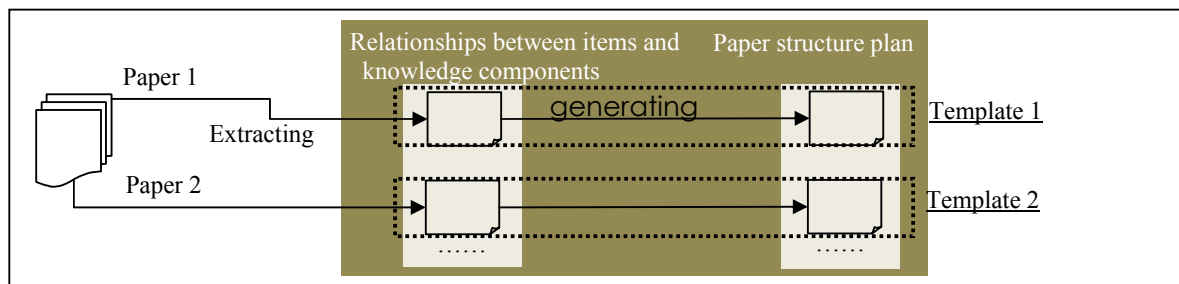


Figure 8. Reverse-generation Procedure of the Templates for Paper Structure Plan

(2) Generation of the Paper Structure Plan

Besides of those template paper structure plans, the system also supports automatic paper structure plan generation.

The procedure is in the following sequence:

Step 1. Specifying the general type plan: scores, testing time and item's quantity of each type (in manual).

Step 2. Specifying the item type assignment according to knowledge (concept) components (Table 2), conforming to the knowledge allocation defined in the examination requirements. Specifying the detailed item-knowledge assignment.

This step is composed by following sub-steps:

Step2.1 Sorting the item type according to the average knowledge component amount from highest to lowest.

Step2.2 Picking the highest agreement item type, (e.g. *Integrated Question*, in Advanced Mathematics).

Step2.3 Picking a set of knowledge components of an integrated question within the range in the database (the knowledge components of some item type are usually fixed) at random. The choice is recorded as the knowledge components of the first Integrated Question.

Step2.4 Picking another set of knowledge components.

Step2.5 Checking whether this choice would overlap the knowledge components chosen in the former steps too much, that is, if the majority of knowledge components would be overlap with chosen knowledge components previously. If it is so, discarding this set and going back to Step 2.4 to pick another set of knowledge components, but ensuring that all the discarding knowledge components would not be selected again.

Step2.6 Assigning the set of knowledge components chosen above to the next testing item.

Step2.7 Continuing Step 2.4-2.6 until knowledge components assignment for all testing items of the select item type is finished.

Step2.8 Pick the next item type.

Step2.9 If every item of this type usually consists two or more knowledge components, continuing Step 2.4-2.7.

Step2.10 Pick all the item types, items of which covering only one knowledge unit.

Step2.11 Assigning knowledge unit to rest items with according to the examination requirement analysis.

In this way, the detailed item-knowledge assignment and item type assignment according to knowledge components are set.

Step 3. Specifying the item type assignment according to difficulty level and testing objectives (see Table 1)

(3) Utilization of the paper structure plan

The paper structure plans work as the blueprint of testing paper, greatly assisting the paper generation procedure. In the computer generated mode, the selection criteria transferred from those detailed knowledge components of every item accelerates item selection.

In the manual generated mode when short of items, the system not only provides the detailed knowledge components as item compiling rule, but also provides sample testing items consisting similar knowledge components to faculties for reference.

THE APPLICATION IN THE SCHOOL OF CONTINUING EDUCATION AT CUGB

After the installation of the semi-auto computer generated paper system in SCECUGB, items were collected and imported into the database. The post-test review showed that faculties were satisfied with the authoring tool for its seamless integration with MS Word.

After a month, in the spring entrance examination in the distance examination center located in Luohe City, Henan province, the system was put into service. The item-bank with no more than 100 items successfully supported the paper generating procedure. Experienced faculties showed positive attitude toward using this system.

CONCLUSION

In order to alleviate faculties' heavy work load in paper generation, automatic computer generated system is proposed as the solution. However, the shortage of a mature item-bank makes the common generation systems unsuitable. Based on the analysis and comparison of manual paper generated mode and computer automatically paper generated mode, in the paper generation procedure a semi-auto system is accordingly designed and developed. By providing strong support to the most time-consuming step, constructing paper structure plan, the system supports the paper generation greatly. The system not only provides template paper structure plan and also supports the automatic generation of new paper structure plan.

This system worked well in the situation of the short of testing items in the item-bank. The system provides a powerful authoring tool for editing items assisting the item-bank construction. In real entrance exam condition, this system was tested and has obtained a positive result.

REFERENCES

- Anderson, L.W. and Krathwohl, D.R. (Eds.) (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- China E-Learning Technology Standardization Committee [CELTSC] (2003). Technical Specification for Educational Resource Construction: Information Model (CELTS-41.1 CD1.0). Retrieved from <http://www.celtsc.edu.cn/680751c665875e93/folder.2006-04-03.8417036039/celts-41/celts-41-1-cd1-0.pdf/view>
- Chou, C. (2000). Constructing a computer-assisted testing and evaluation system on the World Wide Web-the CATES experience. *IEEE Transactions on Education*, 43(3), 266–272.
- Feng, C. B. (2010). *A genetic algorithm based on intelligent grouping model and system design*, Unpublished M.S Thesis, North-east Normal University, P.R.China.

- Gronlund, N. E. and Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: MacMillan.
- Guan, E.J., Han, X.B, Zhou, Q. and Shen, W.H. (2008). The item bank management system for university common courses: design, software development and application. *China Education Inf.* 12, 30-33.
- Hwang, G. J., Chu, H. C., Yin, P.Y., Lin J.Y.(2008).An innovative parallel test sheet composition approach to meet multiple assessment criteria for national tests. *Computers and Education* 51,1058–1072
- Hwang, G. J., Lin, B. M. T. and Lin, T. L. (2006). An effective approach for test-sheet composition from large-scale item banks. *Computers and Education*, 46(2), 122–139.
- López-Cuadrado ,J., Pérez ,T. A., Vadillo ,J. Á., Gutiérrez, J.(2010). Calibration of an item bank for the assessment of Basque language knowledge. *Computers and Education*, 55, 1044–1055.
- Lilley,M., Barker,T., Britton,C.(2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers and Education* 43, 109–123
- Liu, B. (2006). Quality control mechanism of self-taught examination. *China Examinations.* 04, 14-17.
- Ma, Q. and Yu, J. (2006). Constructing paper sank in the self-taught examination. *China Examinations.* 07, 22-25.
- Wang, T.H., Wang, K.H., Huang, S.C.(2008). Designing a Web-based assessment environment for improving pre-service teacher assessment literacy. *Computers and Education* 51, 448–462.
- Thurlow, M., Lazarus, S. S., Albus, D. and Hodgson, J. (2010). *Computer-based testing: Practices and considerations* (Synthesis Report 78). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Yu, H. (2008). *Research of test paper based on GA and realization of item bank system*, Unpublished M.S Thesis, Chongqing University, P.R.China.
- Yuan, X.P. (2008). *Research on intelligent examination system based on adaptive genetic algorithms*, Unpublished M.S Thesis, Central South University, P.R.China.
- Zhao, H.Y., Wang, Q.Y. and Zhao X.M. (2010). The parallelism of higher education self-taught examination papers and its evaluation dimensions. *China Examinations.* 02, 14-20.