A Model-Based Method for Content Validation of Automatically Generated Test Items

by

Xinxin Zhang

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Education

In
Measurement, Evaluation and Cognition

Department of Educational Psychology
University of Alberta

© Xinxin Zhang, 2016

Abstract

The purpose of this study is to describe a methodology to recover the model (cognitive and item models) from generated test items using a novel graph theory approach. Beginning with the generated test items and working backward to recover the original model (cognitive and item models) using a systematic process with graph theory serves as model-based method for validating automatically generated test items. The methodology is demonstrated using generated items from the medical education domain. The proposed methodology was found to be systematic and generalizable using three different datasets.

Table of Contents

List of Tables

List of Figures

Chapter 1: INTRODUCTION

The *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014) define “validity” as: “The degree to which accumulated evidence and theory support specific interpretation of test scores entailed by proposed uses of a test.” This potentially large-body of evidence can be categorized into content-related validity evidence, criterion-related validity evidence, and consequential-related validity evidence (Runder & Schafer, 2002). Test content is one of the sources of content–related validity evidence. The evidence is based on content relevance and representativeness of the items included in an instrument. This evidence is obtained from judgmental and logical analysis of the test items, which is usually done by subject matter experts (SMEs) who are experts in the domain of interest. “Validation” is defined as: “The process through which the validity of the proposed interpretation of test scores is investigated” (American Educational Research Association et al., 2014). Thus, content validation is a process for SMEs to analyze the relationship between a test’s content and what the test is intended to measure through an evaluation of the items on a test in relation to their relevance for the domain of interest and the representativeness of the relevant items.

Background to Problem

Automatic item generation (AIG) is an item development approach that uses both cognitive and psychometric theories to rapidly produce high-quality, content-specific test items, with the aid of computer technologies (Gierl &Haladyna, 2012). AIG relies on cognitive models designed by SMEs to produce new items through the use of algorithms that systematical organize and structure the item content. AIG is a three-step process consisting of cognitive model development, item model development, and item generation (Gierl & Lai, 2012). Typically, AIG content validation occurs after item generation. The purpose of content validation is to determine domain clarity, evaluate the items on a test in relation to their relevance and representativeness, and make sure no errors have occurred in the presentation of the items during the generation process. It currently relies on a one-item-at-a-time review of which SME evaluate the content of the generated items. Because AIG produces large numbers of items, SMEs usually sample generated items for review. If low relevance, low representativeness, or presentation flaws are observed, then the SMEs will review another sample of generated items. The feedback will be collected for revising the item model. Then the new generated items from the revised item model will be reviewed again to make sure the detected problems have been solved.

But the current approach to item review for content validation using generated items has three disadvantages. First, it is time consuming, especially when a large number of generated items are produced. Even though AIG is a breakthrough in the test development process because it satisfies the need for rapidly and efficiently producing large numbers of high-quality content-specific test items, its application of item review for content validation can still hinder the process. Suppose 2000 items, which is the minimum number of items for a 40-item computer adaptive test bank (Breithaupt, Ariel & Hare, 2009), are generated and ready for content validation. Estimated review time for each item by one SME, which includes collecting and recording the SME's judgments, is approximately 10 minutes. If three SMEs are involved in this process and they randomly review 70% of the items, then we can project that they would spend 42000 minutes (700 hours) alone just to review one sample of the generated items. If flaws are detected during the process, then more items need to be viewed which will require even more time.

Second, traditional item review has a high cost due to human capital. That is, the costs associated with the traditional item review method for content validation are severe. According to Statistics Canada, the average hourly wage for occupations in social science and education is $31.16. If we combine the above time estimation, then we can project that it would cost around $21812 alone just for content validation. Furthermore, this projection is made under an unreal assumption that SMEs are fully satisfied with the generated items and have no feedbacks about them. If flaws are detected, then human capital costs increase.

Third, the organization structure is redundant. From the perspective of organizational structure that delineates lines of communication, authority, and responsibilities and indicates how information flows, AIG‘s content validation has some redundancy (Ashkenas, 1995). Currently, AIG’s content validation process falls into a tall (vertical) structure which follows the layout of a pyramid. The SMEs who review individual item sit at the bottom of the pyramid. They are “subordinate” to the AIG model developer. When they find flaws in an item, their feedback about the item will be sent to the AIG model developer first, and then the model developer will revise it based on the feedback about individual items resulting in new items being generated from the revised model. For SMEs the unit of analysis they deal with is the item. However, for the model developer, the unit of analysis is the model. The transformation between item and model happens many times during the process which causes redundancy. This redundancy also causes the time and human capital cost to increase.

Purpose of Current Study

Because of these important disadvantages, an alternative method which can save time and cost as well as present a flat structure for content validation is needed. The purpose of the current research is to describe and illustrate an alternative method and demonstrate this method with three practical applications. The alternative method that will be described and demonstrated in the current study is validating AIG items through a recovery process that requires tracing the model (cognitive and item models) from the generated test items using graph theory. We call the new method an *AIG model-based review*. To-date, no one to our knowledge has used graph theory to validate model for AIG, specifically, or in the test development a review process, more generally. Hence, the purpose of our study is to describe and illustrate this new method.

Chapter 2: LITERATUREREVIEW

Item Development

Item development is one of the twelve essential, interrelated components required to create a test. The testing process starts with delineating an overall plan and concludes with producing test documentation to support its technical adequacy and validity (Lane, Raymond, Haladyna,& Downing, 2016). Item development involves activities like item writing, item content validation, item tryouts, and item banking, following the applicable standards to accumulate validity evidence to support and sometimes refute the intended interpretations and uses of test results.

The traditional item development approach begins by recruiting and training subject-matter experts (SMEs) to write items. They are responsible for locating related materials and creating items. Item writing is based on the judgement, experiences, and expertise of the SMEs. Once the items are developed, item content validation is conducted preferably by experts who were not involved in developing the items. The reviewers evaluate the items on a test in relation to their relevance for the domain of interest and the representativeness of the relevant items. They also evaluate the printing, font size, appropriateness of language. Depending on the outcomes from these reviews, some items are edited and reviewed again. Once items pass the content validation step, they are administered to a sample of examinees to evaluate the statistical properties. The items are typically evaluated for their difficulty and the extent to which they discriminate among examinees, which helps SMEs to decide which items will be retained for testing and which need delete or revise. The qualified items are then securely stored in a database for use on operational exams.

Automatic Item Generation (AIG)

Automatic item generation (AIG) is a recently developed and efficient way to generate items with the use of computer algorithms (Gierl & Lai, 2013). The role of the SMEs is not to locate materials and write individual items but to organize the resources and create meaningful item models for generating items. Gierl and Lai (2012) described a three-step AIG method. It includes developing cognitive model, creating item model, and generating items with the aid of computer technology. The first step in the AIG process is to develop cognitive models which highlight both the examinees' knowledge and skills required to solve the item as well as specify the content features in the items. To create the cognitive models, the SMEs are asked to identify and describe the key information that would be used to solve a parent item. Parent items are the representative items which highlight the underlying structure of the model. The representation is then documented as a cognitive model with problem and associated scenarios, sources of information, and elements and constrains (Gierl, Lai, & Turner, 2012). This cognitive model is used to guide the detailed rendering needed for item generation.

The second step is to create item models which contain the components in an assessment task, including the stem, the options, and the auxiliary information based on the cognitive model. The stem contains content and the question the examinee is required to answer. The option includes a set of alternative answers with one correct option and more incorrect options. Auxiliary information includes any additional content, in either the stem or option, required to generate an item, including text, images, tables, graphs, diagrams, audio, and/or video. The specific variables in an item model that are manipulated to produce new test items and the content used for these variables are identified in this step. Figure 1 shows an item model in the medical education domain. The upper box (stem-box) presents a stem with five variables

[HISTORY], [BP], [HR], [PHYSICAL_EXAM] and [FOLEY_OUTPUT]. The middle box (element box) shows the corresponded content for these variables. The bottom box (option box) lists all the options including keys and distractors.

Stem	A 25-year-old male is involved in a **[HISTORY].** Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him. When he arrives his blood presuure is **[BP]** and his heart rate is **[HR].** He has a a Glasgow Coma Scale score of 14. On examination, he has **[PHYSICAL_EXAM].** A foley catheter emirts **[FOLEY_OUTPUT]** urine. What is the best next step in the management of this patient?
Elements	HISTORY (Text): 1. a highway speed MVC 2. a highway speed MVC and was ejected from the vehicle 3. a motorcycle accident at highway speeds where his abdomen impacted the handlebars BP (Number): 1. 140/90 2. 135/78 3. 120/70 4.89/65 5. 80/50 6. 75/35 HR (Number): 1. 140 2.135 3. 128 4. 90 5. 87 6.75 PHYSICAL_EXAM (Text): 1.good air entry, a minimally distended abdomen with no guarding 2. good air entry, a large distended abdomen with guarding 3. decreased air entry to bases, a distended, peritonitis abdomen FOLEY_OUTPUT (Text): 1.200cc 2. 600cc 3. no 4. 100cc bloody
Options	1. Chest tube 2. Antibiotics 3. Laparotomy 4. Fluid resuscitation 5. Full-body CT scan

Figure 1. An item model with a stem and five options.

The third step is to generate items using computer software. All possible combinations of the variable content and options are assembled subject to the constraints articulated in the cognitive model, which ensures the generated items are sensible and useful. Two generated items from the item model above are presented below (Figure 2).

1. A 25-year-old male is involved in a highway speed motor vehicle collision where he was ejected from the vehicle. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary centre. When he arrives his blood pressure is 75/35 and his heart rate is 140. He has a Glasgow Coma Scale score of 14. On examination, he has good air entry and a large distended abdomen with guarding. A foley catheter emits 100cc of bloody urine. What is the best next step in the management of this patient?

2. A 25-year-old male is involved in a motorcycle accident at highway speeds where his abdomen was impacted with the handlebars. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary centre. When he arrives his blood pressure is 89/65 and his heart rate is 128. He has a Glasgow

Coma Scale score of 14. On examination, he has decreased air entry to bases and a distended peritonitic adbomen. A foley catheter emits 200cc of urine.What is the best next step in the management of this patient?

Figure 2 Two generated items using the item model in Figure 1.

Currently, these generated items follow the same content validation process as with traditional items. That is, item review relies on a one-item-at-a-time evaluation where SMEs scrutinize the content of the generated items. However, this item review approach is time consuming and costly, particularly when large numbers of new items must be reviewed. The AIG model-based review method put forward in this research is designed to overcome these challenges.

Graph Theory

The AIG model-based review is based on a graph theory analysis of the generated test items. Graph theory (GT) is the study of mathematical structures used to model pairwise relations between objects. It is commonly used in mathematics and computer science.

The application of graph theory in developing an alternate content review method brings about some important advantages. First, graph can clearly and practically present the cognitive model and item model, which are the objects for the content view in our methodology. Second, the application of graph theory helps us deal with complicated problems that emerge during the content review process due to the abundance of algorithms. Third, graph theory's computerization characteristic makes the content review process more efficient and it expedites the overall application of AIG because the review process can occur more quickly and efficiently.

The first characteristic of GT is it is practical applicable. GT has been applied to different areas for analyzing concrete, real-world problems. In various areas like chemistry, psychology, linguistics, management science, communication science, and computer technology, many problems can be formulated and solved in graph theory. Cartwright, Harary, and Norman (1965) initiated the application of graph theory to social psychology during the 1960s. They used signed graphs with sign + or − attached to each of its arcs to represent social relationship between people within a particular group. Each vertex of a graph presents each person, and the arcs connecting the vertexes stand for the relation. If two persons share the same social traits, like "friendship" "same religious belief", then a positive sign would be assigned to the arc meaning that these two people are "related". If the two people are unrelated/opposite in terms of the social trait, then a negative sign will be attached. The signed graph's balance property plays a significant role in the research about Social System's Balance (Turner, 1991). A graph's balanced signed property is defined as a graph in which the vertex set can be partitioned into two subsets, so that any arc in each subset is positive, while any arc between subsets is negative. This property is used to mimic social system's balance. A social system is called balanced when any two of its people have a positive relation between them, or when it is possible to divide the group into two subgroups so that any two persons in the same subgroup have a positive relation between them while two persons of different subgroups have a negative relation between them (Balakrishnan & Ranganathan, 2012).

Graph theory has proven to be useful in linguistics, especially in computational linguistics, which is an interdisciplinary field concerned with the statistical or rule-based modeling of natural language from a computational perspective. In computational linguistics, language is characterized as a set of words and set of rules for forming sentences. Based on this perspective, the vertices of graphs are used to represent words and words strings. The arcs of the graphs represent certain syntactical relationships between them which are rules (Deo, 2004). Ross and Harary (1959) systematically applied graph theory to management sciences. They solved many management problems by building the correspondence between organizational concepts and graph theory. The organizational concepts, like redundancies, liaison persons, strengthening and weakening members of a group, are analogous to ideas from graph theory. Examples are directed path which passes through the same vertex more than once are used to present redundancies. A liaison person in an organization is analogous to a cut point of a connected graph. In the above examples, graphs are successfully used to model pairwise relations between real objects. The objects in our new methodology for content view will also be presented in graph from, which increases the readability.

The second characteristic of GT is it has an abundant number of algorithms to deal with complicated graph problems. Algorithms of graph theory, which direct the step-by-step operations like calculation, data processing, and automated reasoning, are readily available. The most basic algorithms like graph exploration algorithms are used to check the connectivity of graph through exploring all of graph's vertices. Other algorithms like Dijkstra's algorithm (Skiena, 1990), which is usually used in transportation area, aims at finding the shortest path. Prim's algorithm and Kruskal's algorithm (Ahuja, 1993) can be used for finding and generating minimum spanning tree. One example is that Oxford University's researchers used the construction of minimum spanning tree to mimic global foreign exchange market dynamics (McDonald, Suleman, Williams, & Howison, 2005). The availability of these graph theory algorithms and along with their practical application indicate the prospect of applying graph theory to deal with the complicated problems that may emerge during the content review process in item development.

The third characteristic is the computerization of storage methods which makes graph theory efficient. In order to store and retrieve GT's objects, graphs which represent the data structure have been developed. The two data structures commonly used for storing graphs are list structures and matrix structures. List structures represent the graph's information as an ordered sequence. Matrix structures represent it as a matrix (Black, 2004). In this study, we will use matrix structures to represent the graph of the model for content review which expedites the item review process.

Graphs are the core elements for graph theory research and analysis. Vertices (nodes) and edges (arcs) are the fundamental and indivisible units of a graph. A vertex *v* is expressed by a point or a circle. An edge is a link between two nodes. The edges may be directed or undirected. Directed edges connect ordered pairs of vertices where an arrow extending from one vertex to another vertex will be observed. Undirected edges connect unordered pairs of vertices by line. A directed graph is a graph whose edges are all directed. An undirected graph is a graph whose edges are all undirected. A graph with both directed and undirected edges is called a mixed graph (West, 2001). All three graph types are shown in Figure 3.

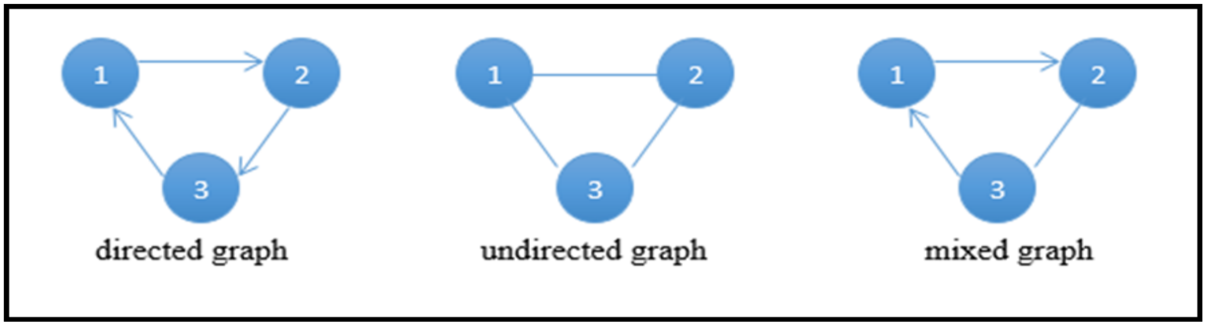
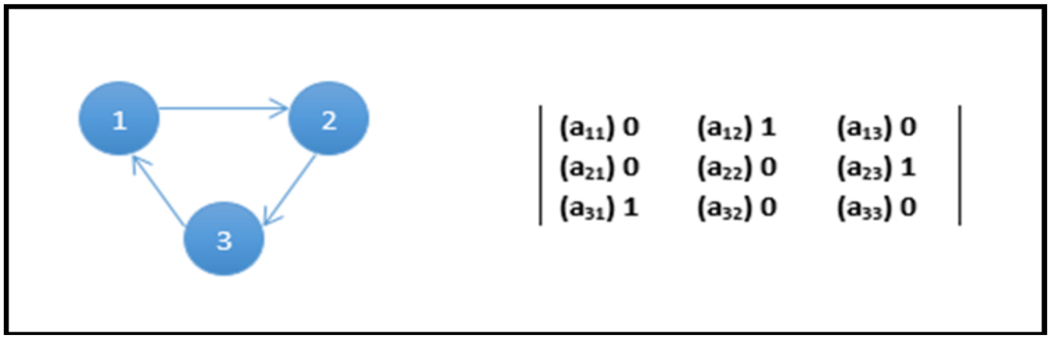


Figure 3. Examples of a directed graph, undirected graph, and mixes graph.

On the edge of each graph we can assign a number as a weight to express more information. Suppose the vertices represent buildings on campus and we have an edge between two vertices meaning a path between them exists. The weight for this edge can be the path's length or the time that is required to walk from one building to another. As mentioned before, matrix structure is one of the data structures that can be used to represent a graph. Adjacency matrices specify the nodes' with adjacent relations. The adjacency matrix of a directed graph on n vertices is a $n \times n$ matrix where the diagonal entries a_{ij} are 0 and the non-diagonal entry a_{ij} can be 1, when there is an ordered edge from vertex i to vertex j. For a directed graph on 3 vertices, it has 3×3 adjacency matrix. The diagonal entries $a_{11,}a_{22}$ and a_{33} are zeroes and the non-diagonal entries can be 0 or 1, depending on if there is a directed edge. Figure 4 shows a directed graph which has three vertices 1, 2, and 3 as well as three directed edges $< 1,2 >,<$ $2,3 >$, and $< 3,1 >$and its 3×3 adjacency matrix, for which the entries of a_{12},a_{23} , and a_{31} are 1 and the remaining entries are 0.



$$\begin{vmatrix} (a_{11})\ 0 & (a_{12})\ 1 & (a_{13})\ 0 \\ (a_{21})\ 0 & (a_{22})\ 0 & (a_{23})\ 1 \\ (a_{31})\ 1 & (a_{32})\ 0 & (a_{33})\ 0 \end{vmatrix}$$

Figure 4. A directed graph and its adjacency matrix.

Graph Theory and the Model-Based Method

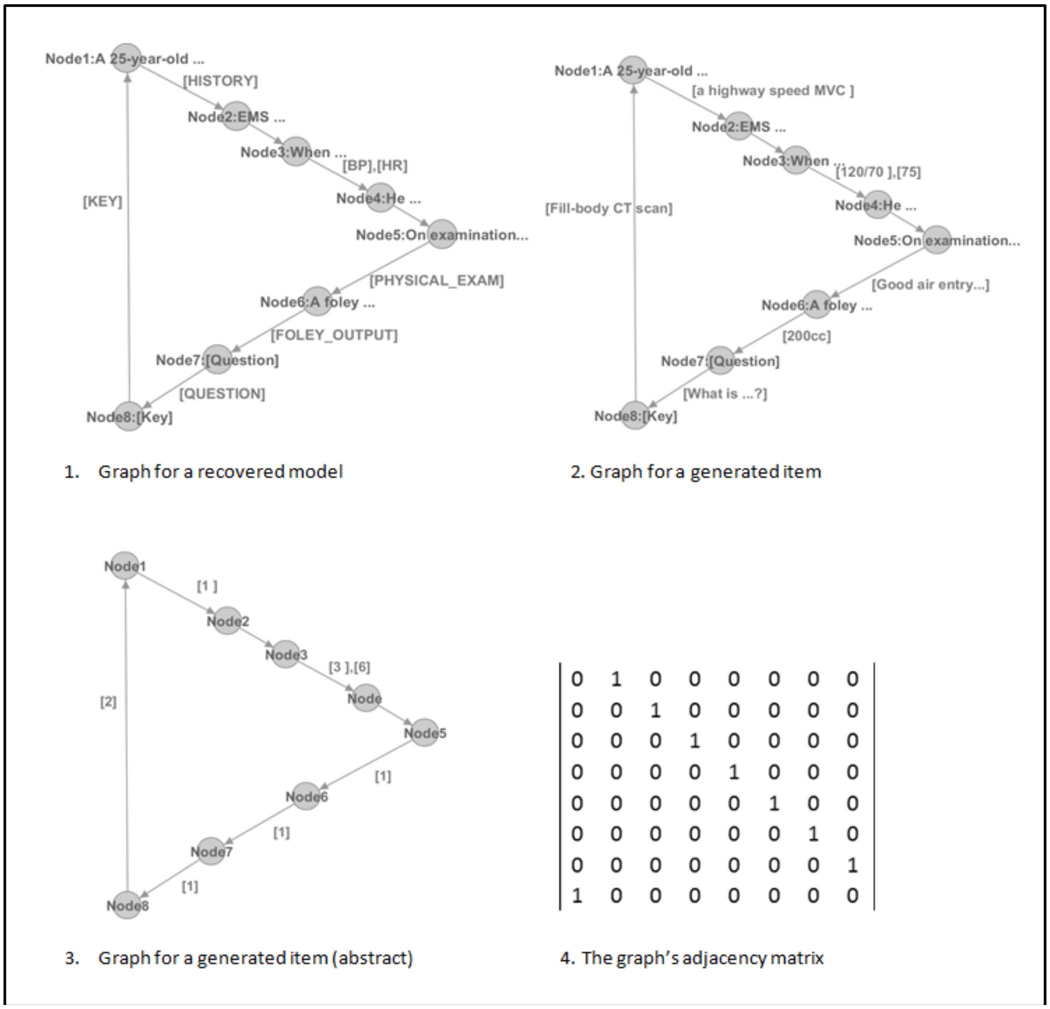
This graphical structure and its adjacency matrix are applied to present the recovered model from the generated items for review. The method for recovery will be discussed in more detail in the method section. In Figure 5, a recovered model is presented as an example. It is similar to the original item model, but with two more variables [QUESTION] and [KEY] in the stem-box. The corresponding values for these two variables are added in the element box and there is no option box.

Stem	A 25-year-old male is involved in a [HISTORY]. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary centre. When he arrives his blood pressure is [BP] and his heart rate is [HR]. He has a Glasgow Coma Scale score of 14. On examination, he has [PHYSICAL_EXAM].A foley catheter emits [FOLEY_OUTPUT] urine. [QUESTION]. [KEY].
Elements	HISTORY (Text): 1. a highway speed MVC 2. a highway speed MVC and was ejected from the vehicle 3. a motorcycle accident at highway speeds where his abdomen impacted the handlebars BP (): 1. 140/90 2. 135/78 3. 120/70 4.89/65 5. 80/50 6. 75/35 HR(): 1. 140 2.135 3. 128 4. 90 5. 87 6.75 PHYSICAL_EXAM (Text): 1. Good air entry, a minimally distended abdomen with no guarding 2. good air entry, a large distended abdomen with guarding 3.decreased air entry to bases, a distended, peritonitis abdomen FOLEY_OUTPUT (Text): 1. 200cc 2. 600cc 3. no 4. 100cc bloody QUESTION (Text): 1. What is the best next step in the management of this patient? KEY (Text): 1. Laparotomy 2. Full body CT

Figure 5. Example of a recovered model.

To present this model, we use a directed graph with eight nodes. Each of these nodes represents one sentence in the stem. As the first panel in Figure 6 shows, the first node is the first sentence of the stem “A 25-year-old male is involved in a [HISTORY]”. The second node is the second sentence “Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary centre.” The third node is the third sentence “When he arrives his blood pressure is [BP] and his heart rate is [HR].” The fourth node is the fourth sentence “He has a Glasgow Coma Scale score of 14.” The fifth node is the fifth sentence “On examination, he has [PHYSICAL_EXAM].” The sixth sentence is the sixth node “A Foley catheter emits [FOLEY_OUTPUT] urine.” The seventh node is the seventh sentence “[QUESTION].” The last node is the last sentence “[KEY]”. The weight of each edge presents the variables contained in the edge’s initial node’s sentence. This graphical structure can also be applied to the generated items when we replace the variables with content (see panel 2, Figure 6).

In order to simplify the graph, we can further replace the content of variables with their corresponded number listed in the element box. In our example, the content for the [HISTORY] “a highway speed MVC” corresponds to the number sequence “1”. We can use the number sequence “[1]” as the first edge’s weight. The content for the [BP] and [HR] are “120/70” and “75”, separately corresponding to the number sequence “3”,"6”. We can use the number sequence “[3], [6]” as the third edge’s weight. The same procedure can be applied to the remaining variable contents resulting in the generated item shown in panel 3 in Figure 6. Panel 4 in Figure 6 is the adjacency matrix of this graph.



1. Graph for a recovered model

2. Graph for a generated item

3. Graph for a generated item (abstract)

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

4. The graph’s adjacency matrix

Figure 4. Examples of using graph to represent model and items.

Statement of the Research Problem

Item development is the central component in the creation of a test because it provides the content which, in turn, produces the validity evidence that can be used to support or refute the intended interpretations and uses of test score results. Item development includes activities such as item writing, item content validation, item tryouts, and item banking. AIG is an alternative way to "write" large numbers of items. But the process of reviewing these items raises important new challenges. Graph theory, as described in this literature review, will serve as a method to operationalize a model-based content validation method that can help overcome some of the item review challenges.

Chapter 3:METHOD

An eight-step methodology using a graph theory approach was implemented to recover the model from the generated items. The generated items are multiple-choice items with a single stem and four options. The stem contains content (non-question component) and the question. The options include a set of alternative answers with one correct and three incorrect options.

Step 1. Categorize the items

The purpose of this step is to categorize the items based on the number of sentences. The items which have the same number of sentences are placed in the same category or "bin" which can be a sheet of an Excel file for further processing. The assumption is the items with the same number of sentences might be generated from the same item model. Conversely, the items with a different number of sentences may be generated from a different item model. Based on this assumption, categorizing the items improves the efficiency of tracing the model. The outcome is different Excel sheets with similar items inside each sheet

Step 2. Parse the items

The purpose of this step is to parse non-question component of the items in all Excel sheets. The Stanford Parser was used. The Stanford Parser is a program developed by the Natural Language Processing Group at Stanford University (2016). The parser identifies the grammatical structure of the sentences, for instance, which groups of words go together as "phrases" and which words are the subject or object of a verb. The example below presents a parsed sentence from a medical item. "A/DT 25-year-old/JJ male/NN is/VBZ involved/VBN in/IN a/DT highway/NN speed/NN MVC/NN." Each word in the sentence is followed by a slash for separation and some capital letters which are the part-of-speech tags. Parts-of-speeches are the basic types of words in the English language that includes nouns, verbs, adjectives, adverbs, pronouns, conjunctions, prepositions, and interjections. The Stanford Parser sets up its own part-of-speech tags for a single word according to its role in the sentence. In this example, DT stands for determiner, JJ stands for adjective, NN stands for noun, VBZ stands for verb's 3rd single singular present, VBN stands for verb's past tense and IN stands for preposition. The outcome of this step is parsed sentences with identified grammatical structures within each Excel sheet.

Step 3. Restate the items

The purpose of this step is to restate each parsed sentence based on the grammatical structure. A grammatical link was used to realize this goal. The grammatical link consists of part-of-speech tags and space. Take this parsed sentence as an example: *"*A/DT 25-year-old/JJ male/NN is/VBZ involved/VBN in/IN a/DT highway/NN speed/NN MVC/NN." The basic grammatical structure of this sentence is: subjective-verb "male/NN is/ VBZ involved/VBN". "Male" is the subjective, and "is involved" is the verb. The phrases "A 25-year-old" and "in a highway speed MVC" separately modifies the subjective "male" and the verb "is involved". The grammatical link keeps the part-of-speech tags of the basic grammatical structure "male/NN is/ VBZ involved/VBN" and uses the space to replace the modification components "a 25-year-old" and "in a highway speed MVC". Thus the grammatical link of this sentence becomes "() NN VBZ VBN ()". There are two reasons for using a grammatical link. First, it improves the efficiency of recovering the model. The sentences which have the same grammatical link are more likely generated from the same item model. Second, the grammatical link is programming friendly because it flags the locations for matching. The importance of this point will be discussed later in the methods section. The outcome of this step is a grammatical link for each parsed sentence.

Step 4. Get the abstracted pattern

The purpose of this step is to get the abstracted pattern for each sentence. The abstracted pattern highlights what the sentence looks like, and where the specific variables are located in the sentence. In order to realize this goal, the sentences with the same grammatical link within an Excel sheet are gathered together for tracing the abstracted pattern through matching. To demonstrate the logic of this step, two sentences with the same grammatical link within one Excel sheet are used to illustrate this concept. The two sentences are: *"*A 25-year-old male is involved in a highway speed MVC.*"* and *"*A 25-year-old male is involved in a motorcycle accident at highway speeds where his abdomen impacted the handlebars." Their common grammatical link is "() NN VBZ VBN ()". The space of this grammatical link identifies and isolates where to compare and where to match. In the example, the first space directs a comparison between the modification components "a 25- year-old" and "a 25-year-old". The second space directs a comparison between the modification components "in a highway speed MVC" and "in a motorcycle accident at highway speeds where his abdomen impacted the handlebars". Then the corresponding words for the part-of speech tag in two sentences are compared separately. The first part-of-speech "NN" directs a comparison between "male" and "male". The second part-of speech tag "VBZ VBN" directs a comparison between "is involved" and "is involved". By keeping the same words and replacing the different word/phrases with brackets, an abstracted pattern "A 25-year-old male is involved in [ ]" is produced for these two sentences. The bracket indicates a variable. Then, the different phrases "a highway speed MVC" and "a motorcycle accident at highway speeds where his abdomen impacted the handlebars" are recorded for further processing. The outcome of this step is the abstracted pattern for each sentence.

Step 5. Develop the structure table

The purpose of this step is to develop the structure table. In order to develop this table, two sub-steps are required. First, the abstracted patterns in different Excel sheets are combined and listed in this table. Second, two abstracted patterns are added together to create a list that includes all of the information in a test item. In other words, the added abstracted patterns are for the question and the key, because every item has the question and key. Table 1 illustrated an outcome of this step. This structure table has nine abstracted patterns. The eighth is for question and the ninth is for the key.

Table 1

A Structure Table.

No	Abstracted Pattern
1	A 25-year-old male is involved in a [1].
2	Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center.
3	When he arrives his blood pressure is [2] and his heart rate is [3].
4	He has a Glasgow Coma Scale score of 14.
5	He is complaining of lower-rib pain on his [4].
6	On examination, he has [5].
7	A Foley catheter emits [6] urine.
8	[QUESTION].
9	[KEY].

Step 6. Develop the content table

The purpose of this step is to develop the content table, which lists the content for the variables in the structure table in step 5. The recorded word/phrases in each variable during the matching presented in step 4 are listed in this table as the content. Continuing with the previous example, for the first abstracted pattern in the structure table “A 25-year-old male is involved in [1].” the recorded phrases “a highway speed MVC” and “a motorcycle accident at highway speeds where his abdomen impacted the handlebars” in step 4 are listed in the content table as variable [1]’s content. Table 2 exemplifies an outcome of this step - a content table using the recorded information for each variable that is presented in Table1.

Table 2

A Content Table.

Variable	Conent
[1]	1. a highway speed MVC 2. a highway speed MVC and was ejected from the vehicle 3. a motorcycle accident at highway speeds where his abdomen impacted the handlebars
[2]	1. 140/90 2. 135/78 3. 120/70 4.89/65 5. 80/50 6. 75/35
[3]	1. 140 2.135 3. 128 4. 90 5. 87 6.75
[4]	1. Good air entry, a minimally distended abdomen with no guarding 2. good air entry, a large distended abdomen with guarding 3.decreased air entry to bases, a distended, peritonitis abdomen
[5]	1. right side 2. left side
[6]	1. 200cc 2. 600cc 3. no 4. 100cc bloody
[Question]	1. What is the best next step in the management of this patient? 2.What is the most likely diagnosis?
[Key]	1. Laparotomy 2. Full body CT 3. Splenic rupture

Step 7. Generate sequences

The purpose of this step is to list the structure and variable content for the items using sequences. Two sub-steps are required in step 7. The first sub-step is to get the structure sequence by matching the structure table in step 5 to the items’ abstracted patterns in step 4. The second sub-step is to get the variable content sequence by matching the content table in step 6 to the items. A generated item with its abstracted patterns is given in Figure 5 to demonstrate this concept. The outcome of sub-step 1 is a sequence “1.2.3.4.6.7.8.9”. This sequence represents the item’s structure. Each number in this sequence corresponds to an abstracted pattern listed in the structure table in Table1. For example, the 5th number “6” corresponds to the 6th abstracted pattern “On examination, he has [ ].” The outcome of sub-step 2 is “[1], [], [6; 1], [], [2], [4], [1], [1]”. This sequence represents the variable content for the item. One bracket in this outcome is for one abstracted pattern of the item and the numbers inside correspond to the content listed in the content table. For example, the second bracket [ ] is for the second abstracted pattern “Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center.” This bracket contains no number because the corresponding abstracted pattern doesn’t have any variables. The third bracket [6;1] is for the third abstracted pattern of this item which is “When he arrives his blood pressure is [2] and his heart rate is [3].” The numbers in the bracket separated by semicolon [6;1] correspond to the 6th value in variable [2] and the 1st value in variable [3] listed in the content table in Table 2, which are “75/35” and “140”. The outcome of this step is the structure and variable content sequences, which describe the recovered model.

Generated Item

A 25-year-old male is involved in a highway speed motor vehicle collision where he was ejected from the vehicle. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center. When he arrives his blood pressure is 75/35 and his heart rate is 140. He has a Glasgow Coma Scale score of 14. On examination, he has good air entry and a large distended abdomen with guarding. A Foley catheter emits 100cc of bloody urine. What is the most likely diagnosis?

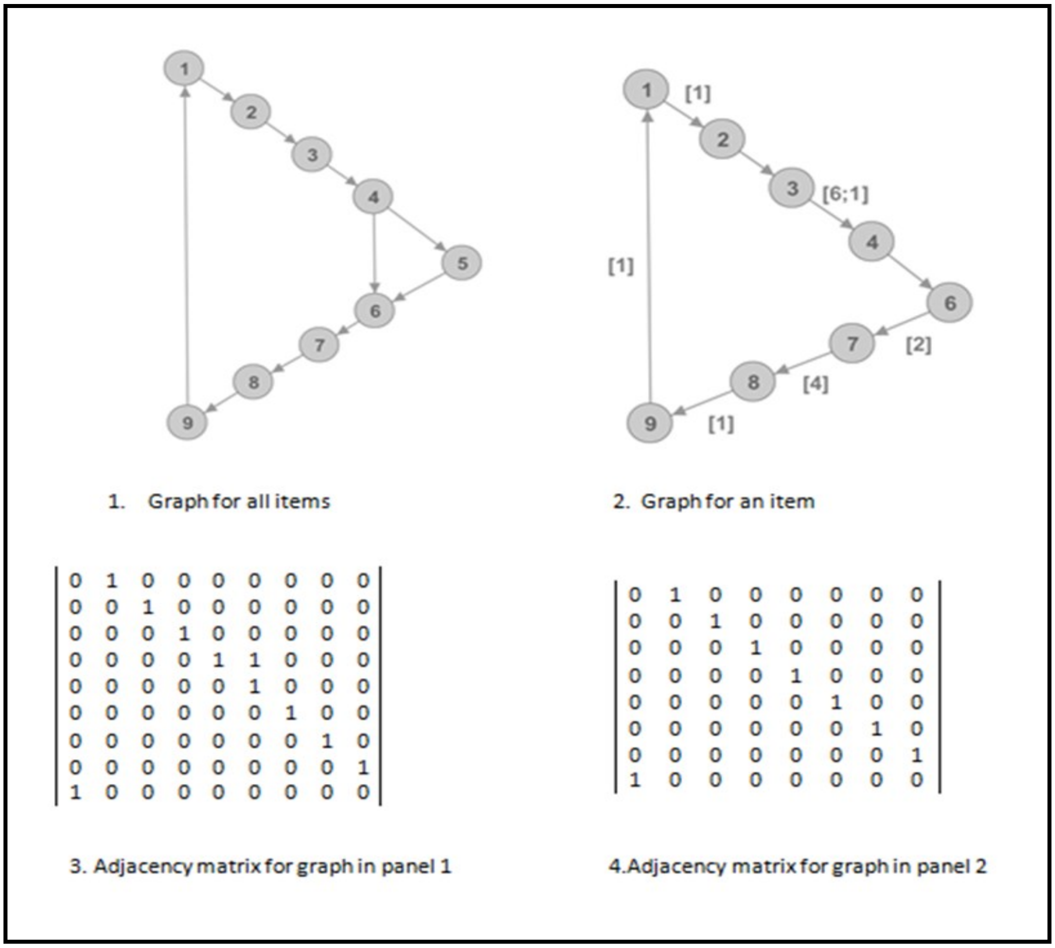
Item's abstracted patterns

A 25-year-old male is involved in a [1].Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center.When he arrives his blood pressure is [2] and his heart rate is [3].He has a Glasgow Coma Scale score of 14.On examination, he has [5].A Foley catheter emits [6] urine.[QUESTION].[KEY].

Figure 5. A generated item and its abstracted patterns.

Step 8. Apply graph theory

The purpose of this step is applying graph theory to present the recovered model. Two sub-steps are taken. First, the graph is used to describe the recovered model. Then, the adjacency matrix is used to describe the graph. In sub-step 1, the nodes of the graph are used to express the structure sequence and the weights are used to express the variable content sequences developed from step 7. Panel 1 in Figure 8 presents a graph with nine nodes. The two paths of this graph present two structure sequences “1.2.3.4.6.7.8.9” and “1.2.3.4.5.6.7.8.9“. This graph indicates the structure for all generated items. In other words, all the generated items are from the model with two paths “1.2.3.4.6.7.8.9” and “1.2.3.4.5.6.7.8.9 “. Panel 2 presents a graph with eight nodes for a typical item with the structure sequence “1.2.3.4.6.7.8.9” and the variable content sequence “[1], [], [6; 1], [], [2], [4], [1], [1]”. The weight of the edge is the number contained in each bracket if it is not []. In sub-step 2, the graph's adjacency matrix is produced as panel 3 and 4 shows. The outcome of this is the graph and adjacency matrix.



1. Graph for all items

2. Graph for an item

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

3. Adjacency matrix for graph in panel 1

4.Adjacency matrix for graph in panel 2

Figure 6. An example about step 8’s outcome.

Chapter 4:RESULTS

Three different sets of items from the medical science domain were generated and the model was recovered to demonstrate our method. All the generated items were created by medical SMEs using the three-step AIG process that included developing a cognitive model, the creating item model, and generating items (Gierl, Lai, &Turner, 2012). A total of 11035 items were generated, including 938 items related to abdominal trauma, 8109 items related to post-operative fever, and 1988 items related to hernia. The items in each dataset were generated based on individual cognitive models and the derived item models.

After applying the 8-step methodology for recovery, the results for the model structure table, content table, graph, and graph matrix were produced and are presented in the following section. The structure table identifies what components (i.e., content, question, and key) are in the model, how the components are structured, and where the specific variables are located. The content table further specifies the content of the specific variables. The graph structures for generated items are also presented. The graph is expressed and also displayed as a matrix using graph theory.

Results from Abdominal Trauma Dataset

Table 3 is the structure table developed in step 5 of recovering the model in abdominal trauma dataset. It lists nine abstracted patterns separately in rows. This table identifies the three components in the model. The first to seventh abstracted patterns (row 1 to row 7) are the content (non-question) component. The eighth abstracted pattern (row 8) is the question component and the ninth abstracted pattern (row 9) is the key component. The brackets indicate the variables in the model. The model in abdominal trauma dataset has eight variables in total.

Variable [1] to variable [6] are located in the content component. They include variable [1] presented in the first abstracted pattern, variable [2] and [3] presented in the third abstracted pattern, variable [4] presented in the fifth abstracted pattern, variable [5] presented in the sixth abstracted pattern, and variable [6] presented in the seventh abstracted pattern. The seventh variable [question] is located in the question component (the eighth abstracted pattern) and the eighth variable [key] is located in the key component (the ninth abstracted pattern).

Table 3

Structure Table for Abdominal Trauma Dataset.

No	Abstracted Pattern
1	A 25-year-old male is involved in a [1].
2	Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center.
3	When he arrives his blood pressure is [2] and his heart rate is [3].
4	He has a Glasgow Coma Scale score of 14.
5	He is complaining of lower-rib pain on his [4]
6	On examination, he has [5].
7	A Foley catheter emits [6] urine.
8	[QUESTION].
9	[KEY].

Table 4 is the content table developed in the step 6. It lists the content for the variables in Table 3. For example, variable [2] in the third abstracted pattern of Table 3 has six values varying from “1. 140/90” to “6. 75/35”. Variable [key] in the ninth abstracted pattern of Table 3 has three values varying from “1. Laparotomy” to “ 3. Splenic rupture”.

Table 4

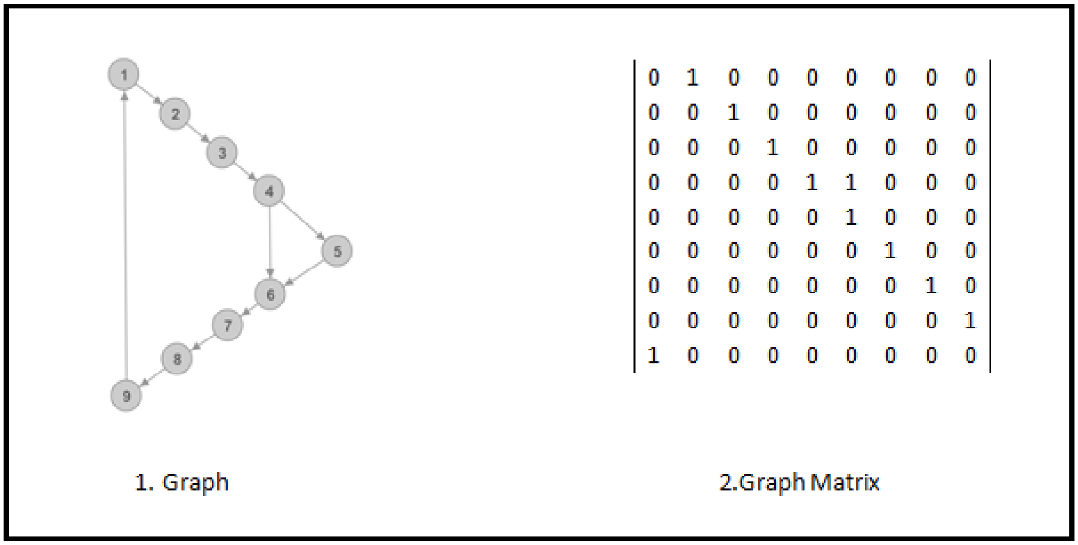
Content Table for Abdominal Trauma Dataset.

Variable	Conent
[1]	1. a highway speed MVC 2. a highway speed MVC and was ejected from the vehicle 3. a motorcycle accident at highway speeds where his abdomen impacted the handlebars
[2]	1. 140/90 2. 135/78 3. 120/70 4.89/65 5. 80/50 6. 75/35
[3]	1. 140 2.135 3. 128 4. 90 5. 87 6.75
[4]	1. Good air entry, a minimally distended abdomen with no guarding 2. good air entry, a large distended abdomen with guarding 3.decreased air entry to bases, a distended, peritonitis abdomen
[5]	1. right side 2 left side
[6]	1. 200cc 2. 600cc 3. no 4. 100cc bloody
[QUESTION]	1. What is the best next step in the management of this patient? 2.What is the most likely diagnosis?
[KEY]	1. Laparotomy 2. Full body CT 3. Splenic rupture

Panel 1 in Figure 9 is the graph developed in the step 8. It structures the generated items from the abdominal trauma dataset. In other words, it presents the recovered model in abdominal trauma dataset. This graph has two paths which are "1.2.3.4.6.7.8.9" and "1.2.3.4.5.6.7.8.9". It indicates that all the generated items in abdominal trauma dataset are from the model with the structure sequences "1.2.3.4.6.7.8.9" and "1.2.3.4.5.6.7.8.9". Based on Table 3, we know the model has two paths. The first path is "A 25-year-old male is involved in a [1]. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center. When he arrives his blood pressure is [2] and his heart rate is [3]. He has a Glasgow Coma Scale score of 14. On examination, he has [5]. A Foley catheter emits [6] urine. [QUESTION]. [KEY]." The second path is "A 25-year-old male is involved in a [1]. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center. When he arrives his blood pressure is [2] and his heart rate is [3]. He has a Glasgow Coma Scale score of 14. He is complaining of lower-rib pain on his [4]. On examination, he has

[5]. A Foley catheter emits [6] urine. [QUESTION]. [KEY]”.

Panel 2 in Figure 9 is the adjacency matrix for the graph in panel 1, developed in step 8 of the recovering process. This $9 * 9$ matrix specifies the adjacent relations of the nine vertexes in the graph. The non-diagonal entries $a_{12,}a_{23}$, a_{34}, a_{45}, a_{46}, a_{56}, a_{67}, $a_{78,}a_{89,}a_{91}$ with 1 indicate 10 ordered edges from vertex 1 to vertex 2, vertex 2 to vertex 3, vertex 3 to vertex 4, vertex 4 to vertex 5, vertex 4 to vertex 6, vertex 5 to vertex 6, vertex 6 to vertex 7, vertex 7 to vertex 8, vertex 8 to vertex 9, and vertex 9 to vertex 1.



$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}$$

1. Graph

2.Graph Matrix

*Figure 7.*A graph and a graph matrix for abdominal trauma dataset.

Results from Post-Operative Fever Dataset

Table 5 is the structure table developed in step 5 of recovering the model in post-operative fever dataset. It lists eight abstracted patterns separately in rows. This table identifies the three components in the model. The first to the six abstracted patterns (row 1 to row 6) are the content (non-question) component. The seventh abstracted pattern (row 7) is the question component and the eighth abstracted pattern (row 8) is the key component. The brackets indicate the variables in the model. The model in post-operative fever dataset has eight variables in total.

Variable [1] to variable [6] are located in the content component. They include variable [1] presented in the first and fourth abstracted pattern, variable [2] presented in the first and fourth abstracted pattern, variable [3] presented in the first and fifth abstracted pattern, variable [4] presented in the second abstracted pattern, variable [5] presented in the second and sixth abstracted pattern, and variable [6] presented in the fifth abstracted pattern. The seventh variable [question] is located in the question component (the seventh abstracted pattern) and the eighth variable [key] is located in the key component (the eighth abstracted pattern).

Table 5

Structure Table for Post-Operative Fever Dataset.

No	Abstracted Pattern
1	A [1] [2] has a [3].
2	On post-operative day [4] he has a temperature of [5] C.
3	Physical examination reveals tenderness in the abdominal region with guarding and rebound.
4	A [1]-year-old [2] was readmitted to the hospital for pain in the abdominal area.
5	[6] was on post-operative day 3 recovering from a [3].
6	The patient has a temperature of [5].
7	[QUESTION]
8	[KEY]

Table 6 is the content table developed in the step 6. It lists the content for the variables in Table 5. For example, variable [1] in the first and fourth abstracted pattern of Table 5 have three values varying from “1.40-year-old” to “3. 70-year-old”.

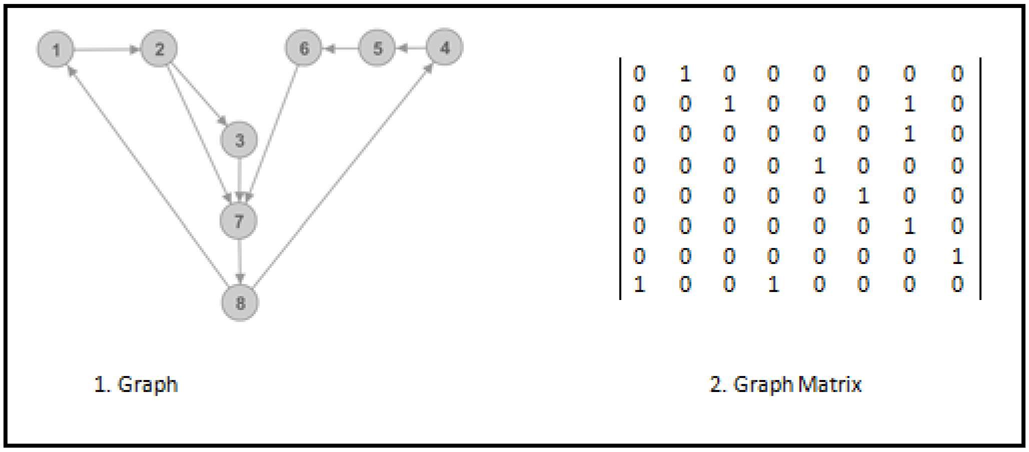
Table 6

Content Table for Post-Operative Fever Dataset.

Variable	Content
[1]	1.40-year-old; 2. 55-year-old; 3. 70-year-old
[2]	1.woman;2.man
[3]	1.appendectomy; 2.gastrectomy; 3.right hemicholectomy; 4.left hemicholectomy; 5.laparoscopic cholecystectomy
[4]	1. 3; 2.4; 3.6; 4.2
[5]	1.38; 2.38.5
[6]	1.she; 2.he
[QUESTION]	1. Which one of the following is the most likely diagnosis? 2. Which one of the following is the best next step for this patient?
[KEY]	1.Anitbiotics; 2.Mobilize;3.Reopen wound; 4.Antibiotics;5.Anti-coagulation;6.Drainage;7.Urinary tract infection; 8.Actelectasis; 9.Wound infection; 10.Pneumonia; 11.Deep vein thrombosis; 12.Deep space infection

Panel 1 in Figure 10 is the graph developed in the step 8. It presents the recovered model in post-operative fever dataset. This graph has three paths which are "1.2.7.8.", "1.2.3.7.8" and "4.5.6.7.8". It indicates that all the generated items in post-operative fever dataset are from the model with the structure sequences "1.2.7.8", "1.2.3.7.8" and "4.5.6.7.8". Based on Table 5, three paths can be identified. The first path is "A [1] [2] has a [3]. On post-operative day [4] he has a temperature of [5] C. [QUESTION]. [KEY]." The second path is "A [1] [2] has a [3]. On post-operative day [4] he has a temperature of [5] C. Physical examination reveals tenderness in the abdominal region with guarding and rebound. [QUESTION]. [KEY]." The third path is "A [1]-year-old [2] was readmitted to the hospital for pain in the abdominal area. [6] was on post-operative day 3 recovering from a [3]. The patient has a temperature of [5]. [QUESTION]. [KEY]."

Panel 2 in Figure 10 is the adjacency matrix for the graph in panel 1, developed in the step 8 of the recovering process. This $8 * 8$ matrix specifies the adjacent relations of the eight vertexes in the graph. The non-diagonal entries $a_{12}, a_{23}, a_{27,}a_{37,}a_{45}, a_{56}, a_{67,}a_{78}, a_{81,}a_{84}$, with 1 indicate 10 ordered edges from vertex 1 to vertex 2, vertex 2 to vertex 3, vertex 2 to vertex 7, vertex 3 to vertex 7, vertex 4 to vertex 5, and vertex 5 to vertex 6, vertex 6 to vertex 7, vertex 7 to vertex 8, vertex 8 to vertex 1, and vertex 8 to vertex 4.



$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{vmatrix}$$

1. Graph

2. Graph Matrix

Figure 8. A graph and a graph matrix for post-operative fever dataset.

Results from Hernia Dataset

Table 7 is the structure table developed in step 5 of recovering the model in hernia dataset. It lists twelve abstracted patterns separately in rows. This table identifies the three components in the model. The first to the tenth abstracted patterns (row 1 to row 10) are the content (non-question) component. The eleventh abstracted pattern (row 11) is the question component and the twelfth abstracted pattern (row 12) is the key component. The brackets indicate the variables in the model. The model in hernia dataset has nine variables in total. Variable [1] to variable [7] are located in the content component. They include variable [1] presented in the first, seventh and ninth abstracted pattern, variable [2] presented in the first, second, sixth, seventh and tenth abstracted pattern, variable [3] presented in the third, fourth, fifth and sixth abstracted pattern, variable [4] presented in the third, fourth, fifth and sixth abstracted pattern, variable [5] presented in the seventh and ninth abstracted pattern, variable [6] presented in the seventh abstracted pattern, and variable [7] presented in the first and eighth abstracted pattern. The eighth variable [QUESTION] is located in the question component (the eleventh abstracted pattern) and the ninth variable [KEY] is located in the key component (the twelfth abstracted pattern).

Table 7

Structure Table for Hernia Dataset.

No	Abstracted Pattern
1	A [1] was admitted with pain in the [2] from a few months ago [7].
2	Patient complaints of a mass in [2] which has been a problem since a few months ago.
3	On examination, the mass is [3] and lab work came back with [4].
4	Upon further examination, the patient had [4] and the mass is [3].
5	With [4] and [3] in the area, the patient is otherwise nominal.
6	There is [3] in [2] and the patient had [4].
7	A [5]-year-old [1] presented with a mass [6] in the [2].
8	It occurred a few months ago [7].
9	The patient is a [5]-year-old [1].
10	Patient presents with a mass in the [2] from a few months ago.
11	[QUESTION]
12	[KEY]

Table 8 is the content table developed in the step 6. It lists the content for the variables in Table 7. For example, variable [2] in the first, second, sixth, seventh and tenth abstracted pattern of Table 7 have four values varying from “1.pertruding but with no pain” to “4.tender and reducible”.

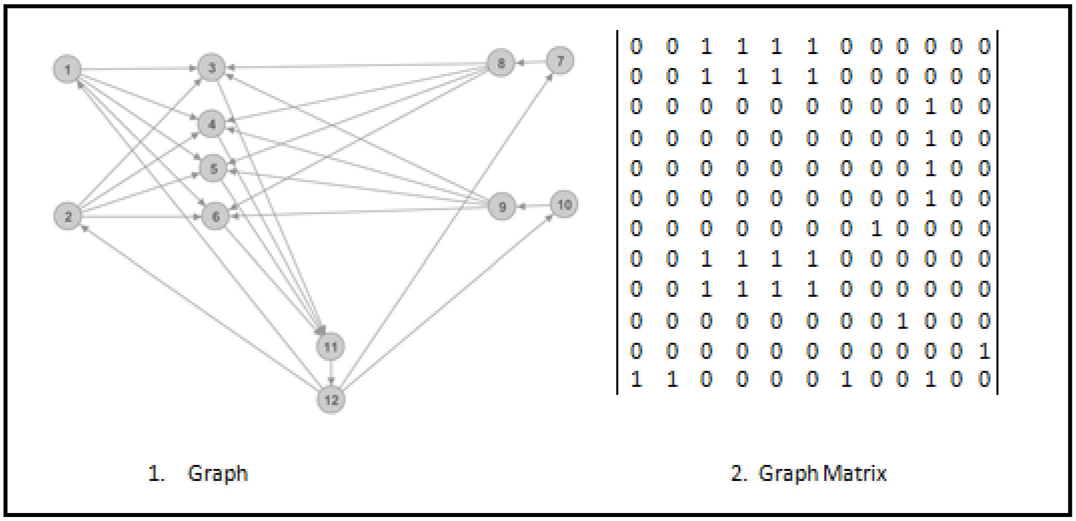
Table 7

Content Table for Hernia Dataset.

Variable	Content
[1]	1.man; 2.woman
[2]	1.area near a recent surgery; 2. right groin; 3. umbilicus; 4.left groin
[3]	1.pertruding but with no pain; 2.tenderness; 3.tender and exhibiting redness; 4.tender and reducible
[4]	1. elevated white blood cell count; 2. normal vitals
[5]	1.25;2.30;3.35;4.40;5.50;6.55;7.60
[6]	1.and intense pain; 2.and severe pain; 3.and mild pain;4.null
[7]	1.null; 2.after moving a piano
[QUESTION]	1. What is the best next step? 2. Which one of the following is the best prognosis? 3. Given this information, what is the best course of action?
[KEY]	1.ice applied to mass

Panel 1 in Figure 11 is the graph developed in the step 8. It presents the recovered item model in hernia dataset. This graph has sixteen paths which are:“1.3.11.12”, “1.4.11.12”,”1.5.11.12”,“1.6.11.12”,”2.3.11.12”,”2.4.11.12”,”2.5.11.12”,”2.6.11.12”,”7.8.3.11.12”,”7.8.4.11.12”,”7.8.5.11.12”,”7.8.6.11.12”,”10.9.3.11.12”, ”10.9.4.11.12”,”10.9.5.11.12”, and ”10.9.6.11.12”. It indicates that all the generated items in hernia dataset are from the model with the structure sequences listed above. Based on Table 7, we know the sixteen paths in the model. For example, sequence “1.3.11.12” corresponds to the path “A [1] was admitted with pain in the [2] from a few months ago [7]. On examination, the mass is [3], and lab work came back with [4]. [QUESTION]. [KEY]”.Sequence “10.9.3.11.12” corresponds to the path “Patient presents with a mass in the [2] from a few months ago. The patient is a [5]-year-old [1]. On examination, the mass is [3] and lab work came back with [4]. [QUESTION]. [KEY].”

Panel 2 in Figure 11 is the adjacency matrix for the graph in panel 1, developed in the step 8 of the recovering process. This 16*16 matrix specifies the adjacent relations of the sixteen vertexes in the graph. The non-diagonal entries $a_{13,}a_{14}, a_{15}, a_{16,}a_{23}, a_{24}, a_{25}, a_{26}, a_{78}$ $a_{83,}a_{84}, a_{85}, a_{86}, a_{10.9}, a_{93}, a_{94}, a_{95}, a_{96}, a_{3.11}, a_{4.11}, a_{5.11}, a_{6.11}a_{11.12,}a_{12.1}, a_{12.2}, a_{12.7,}a_{12.10}$ with 1 indicate 27 ordered edges from vertex 1 to vertex 3, vertex 1 to vertex 4, vertex 1 to vertex 5, vertex 1 to vertex 6, vertex 2 to vertex 3, vertex 2 to vertex 4, vertex 2 to vertex 5, vertex 2 to vertex 6, vertex 7 to vertex 8, vertex 8 to vertex 3, vertex 8 to vertex 4 ,vertex 8 to vertex 5,vertex 8 to vertex 6, vertex 10 to vertex 9, vertex 9 to vertex 4, vertex 9 to vertex 5, vertex 9 to vertex 6, vertex 9 to vertex 3, vertex 3 to vertex 11, vertex 4 to vertex 11, vertex 5 to vertex 11, vertex 6 to vertex 11, vertex 11 to vertex 12, vertex 12 to vertex 1, vertex 12 to vertex 2, vertex 12 to vertex 7, and vertex 12 to vertex 10.



1. Graph

$$\begin{vmatrix}
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0
\end{vmatrix}$$

2. Graph Matrix

Figure 9. A graph and a graph matrix for hernia dataset.

Comparison among the Three Models in the Three Datasets

The comparison among the structure tables, content tables, graphs and graph matrixes across the three datasets shows there are similarities and differences among these models used to generated the items in the three datasets.

There are four similarities. The first similarity is three datasets have the same components including a content (non-question) component, a question component, and a key component, as Table 3, 5, 7 shows. The second similarity is all three datasets have variables with varying values as Table 4, 6, 8 shows. The third similarity is all three datasets can be structured as a directed graph as panel 1 of Figure 9, 10, 11 shows. The fourth similarity is all three data sets' structures can be mathematically expressed as matrixes as panel 2 of Figure 9, 10, 11 shows.

These datasets also contain four important differences. The first difference is each dataset has its own abstracted patterns in the components. Take the content component for example, in which, as Table 3 shows, abdominal trauma dataset has seven abstracted patterns with a main concept in abdominal trauma and information resources from physical examination, like blood pressure and heart rate. As Table 5 shows post-operative fever dataset has six abstracted patterns with a scenario on post-operative day and information resources from physical examination, like temperature in the content component. As Table 7 shows hernia dataset has ten abstracted patterns with a main concept in mass and information resources from physical examination and lab work, in the content component. The second difference is each dataset has different variables and corresponded values. As Table 4, 6, 8 separately show, abdominal trauma dataset has eight variables, post-operative fever dataset has eight variables and hernia dataset has nine variables. The values for these variables differ. The third difference is each dataset has a differently structured graph. As panel 1 in Figure 9, Figure 10, and Figure 11 separately show, abdominal trauma dataset is structured as a nine-node graph with two paths "1.2.3.4.6.7.8.9" and "1.2.3.4.5.6.7.8.9", post-operative fever dataset is structured as an eight-node graph three paths "1.2.7.8", "1.2.3.7.8" and "4.5.6.7.8", and hernia dataset is structured as a twelve-node graph with sixteen paths. The fourth difference is each dataset is expressed differently in matrixes. As panel 2 in Figure 9, Figure 10 and Figure 11 separately show, abdominal trauma dataset is mathematically expressed as a 9*9 matrix with ten non-diagonal entries with 1; post-operative fever dataset is mathematically expressed as a 8*8 matrix with six non-diagonal entries with 1; hernia dataset is mathematically expressed as an 12*12 matrix with twelve non-diagonal entries with 1.

The similarities indicate the recovering method is systematic. After applying the 8-step methodology for recovery, all the datasets were systematically presented as a structure table with three components, a content table containing variables and variable-content, a directed graph, and a graph matrix. The differences indicate the recovering method is generalizable. After applying the 8-step methodology for recovery, each dataset was presented as a structure table with its own abstracted patterns, which presents the main concept, associated scenarios, and the necessary information resources. The diversity of these abstracted patterns means the main concept, associated scenarios, and the necessary information varies in dataset. Each dataset has its own content table containing different variables and variable-content. Recall these variables are related to the main concept, associated scenarios and necessary information, thus the diverse variables again indicate the main concept, associated scenarios and the necessary information varies in dataset. Each dataset was structured as a specific directed graph, which describes the model in the dataset. The diverse structures of these graphs indicate each dataset has different models. Each dataset was mathematically expressed as a varied matrix, which is the representation of the graph in graph theory. The variability of the matrixes means there will be variability among the graphs, indicating again each dataset has varied models.

Chapter 5: DISCUSSION

Automatic item generation is a new approach for item development that satisfies a testing agencies' requirement to produce large numbers of high-quality items in a timely and cost-effective manner. With the aid of the computer, models are used to generate items. After the items are generated, the one-item-at-a-time validation method is used by SMEs to review the generated items in order to analyze the relationship between the content and what the item is intended to measure. But this validation method is time consuming and costly, particularly when large numbers of new items must be reviewed. In order to overcome these challenges, a model-based validation method was developed and demonstrated in this study.

The model-based validation method we described in this paper requires eight steps. First, items are categorized based on the number of sentences. Second, items are parsed using the grammatical structure of sentences. Third, the parsed items are restated. Fourth, the abstracted patterns for items are developed. Fifth, a structure table with listed abstracted patterns is created. Sixth, a content table which specifies the content of the variables in the structure table is developed. Seventh, the structure and variable content for the items are listed in sequences. Eighth, graph theory is applied to present the recovered models.

Using this eight-step process, large numbers of generated items can be validated by reviewing the structure table, content table, graph, adjacency matrix or/and variable content sequences. These outcomes provide the SME with important benefits during the review process. The structure table lists all the abstracted patterns which are used to evaluate the main concept, its associated scenarios, and the information resources within each abstracted pattern. The content table specifies the content of the variables in the structure tables which are used to evaluate the appropriateness of the content and the accuracy of the presentation. The graph structures the item which is used to evaluate the individual task structure. For example, panel 1 in Figure 9 has two paths“1.2.3.4.6.7.8.9” and “1.2.3.4.5.6.7.8.9”, indicating two task structures. The first task structure is “A 25-year-old male is involved in a [1]. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center. When he arrives his blood pressure is [2] and his heart rate is [3]. He has a Glasgow Coma Scale score of 14. On examination, he has [5]. A Foley catheter emits [6] urine. [QUESTION]. [KEY].” The second task structure is “A 25-year-old male is involved in a [1]. Emergency Medical Services (EMS) resuscitates him with 2L crystalloid and transports him to your tertiary center. When he arrives his blood pressure is [2] and his heart rate is [3]. He has a Glasgow Coma Scale score of 14. He is complaining of lower-rib pain on his [4]. On examination, he has [5]. A Foley catheter emits [6] urine. [QUESTION]. [KEY]”. The adjacency matrix, which is a representation of the graph, is used to evaluate the complexity of the dataset. The more non-diagonal entries inside the matrix, the more complex the model will be. Furthermore, the variable content sequences developed in the seventh step of our method can also be evaluated to ensure the combinations of variables are reasonably related to the key. Depending on practical needs, SMEs can review any combinations of the products for validation. For example, if they focus more on the structure of the dataset than the concrete content, then they can just review the graph. After the model-based validation, feedback will be provided to the original AIG model developer for improving their item model.

The model-based validation method is a recovery process whereas the cognitive modeling/item modeling steps in AIG is a development process. Cognitive modeling requires the development of a structure that specifies the knowledge and skills required to solve test items which leads to the creation of new items. By comparison, the validation method begins with the generated test items and works backward to recover the original model (item model/cognitive model) using a systematic process supported by graph theory analysis. The model-based validation method is a solution to the challenging problem of item review when large numbers of generated items are created. Using this method, the SMEs can avoid reviewing the content of every selected item. Instead, they review the summarized products extracted from the items, which saves time and effort. Furthermore, the information organizational structure is transformed from tall to flat, as the object for the SMEs and model developers are in the model level, which simplifies the information organizational structure and improves the efficiency of ccommunication between SMEs and model developers.

Limitation of the Study

The limitation of this study is the demonstration of the methodology focuses only on the generated items within a small number of content areas in the medical education domain. There are many more generated items in different domains and content areas, like mathematics and science, that could be recovered, especially as the types of cognitive models that can be used for AIG is expanding. Thus this research lacks evidence to support the use of model-based validation in all content areas. This limitation affects the generalizability of the methodology to some degree. Furthermore, this limitation confines the target readers to those who have knowledge in the medical education domain.

Recommended Directions for Future Research

Four areas of future research are recommended. The first recommendation is to overcome the limitation mentioned in the previous section by demonstrating the use of this methodology across diverse content areas. The type of research will help ensure the proposed method is generalizable across content areas. The second recommendation is related to the validation method. Although a model-based validation method for generated items has been developed and demonstrated in this research, an alternative validation method for traditional items hasn't been developed. Hence future study is required to develop an alternative validation method using graph theory for traditional items. The type of research will expand the existed validation system. The third recommendation is related to computer technology. So far, the eight steps are individually operatized using different tools and with different degrees of automation. For example, step 2 requires parsing the items using the Stanford Parser and step 3 requires restating the items using Excel. In order to systematically and efficiently recover the items, a computer program which can operationalize the eight steps step-by-step is needed. Hence future study is required to develop a more comprehensive computer program that can implement all eight steps. The fourth recommendation is related to the application of graph theory. Graph theory has many more applications in addition to structuring the dataset, as was illustrated in this research. It can also help us deal with complicated problems, like item generation, distractor generation, and difficulty analysis. The structure of the graph can be used to map and direct the item generation. The node analysis can be used to generate the reasonable distractors through ensuring the key and the distractors have shared nodes which represents the variables. The path analysis can be used to estimate and analyze the item difficulty through giving the estimated variable difficulty as a path weight. Hence future study to apply graph theory to solve more diverse problems in item development including item generation, distractor generation, and difficulty analysis is warranted.

REFERENCES

Ahuja, R. K. (1993). *Network flows* (Doctoral dissertation, Technische Universitat Darmstadt).

Ashkenas, R. (1995). *The Boundaryless Organization: Breaking the Chains of Organizational Structure. The Jossey-Bass Management Series*. Jossey-Bass, Inc., Publishers, 350 Sansome Street, San Francisco, CA 94104..

American educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014).*Standards for educational and psychological testing*. American Educational Research Association.

Balakrishnan, R., & Ranganathan, K. (2012). *A textbook of graph theory*. Springer Science & Business Media.

Black, P. E. (2004). *Dictionary of algorithms and data structures*. National Institute of Standards and Technology.

Breithaupt, K., Ariel, A. A., & Hare, D. R. (2009). Assembling an inventory of multistage adaptive testing systems. In *Elements of adaptive testing* (pp. 247-266). Springer New York.

Cartwright, D., Harary, F., & Norman, R. (1965). Structural models: An introduction to the theory of directed graphs. *New York*.

Deo, N. (2004). Graph theory with applications to engineering and computer science: PHI Learning Pvt. *Ltd India*.

Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education*, *46*(8), 757-765.

Gierl, M. J., & Haladyna, T. M. (Eds.). (2012). *Automatic item generation: Theory and practice*. Routledge.

Gierl, M. J., & Lai, H. (2012). The role of item models in automatic item generation. *International Journal of Testing*, *12*(3), 273-298.

Gierl, M. J., & Lai, H. (2013). Instructional Topics in Educational Measurement (ITEMS) Module: Using automated processes to generate test items. *Educational Measurement: Issues and Practice*, *32*(3), 36-50.

Lane, S., Raymond, M. R., & Haladyna, T. M. (Eds.). (2015). *Handbook of test development*. Routledge.

McDonald, M., Suleman, O., Williams, S., Howison, S., & Johnson, N. F. (2005). Detecting a currency's dominance or dependence using foreign exchange network trees. *Physical Review E*, *72*(4), 046106.

Ross, I. C., & Harary, F. (1959). A description of strengthening and weakening members of a group. *Sociometry*, *22*(2), 139-147.

Rudner, L. M., & Schafer, W. D. (2002). *What teachers need to know about assessment*. Washington, DC: National Education Association.

Skiena, S. (1990). Dijkstra's algorithm. *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica, Reading, MA: Addison-Wesley*, 225-227.

Turner, J. H. (1991). *The structure of sociological theory* (5th ed., pp. 540-572). Belmont, CA: Wadsworth.

The Stanford Parser: A statistical parser. (n.d.). Retrieved March 30, 2016, from http://nlp.stanford.edu/software/lex-parser.shtml.

West, D. B. (2001). *Introduction to graph theory* (Vol. 2). Upper Saddle River: Prentice hall.