Abstract—In spite of the appealing characteristics of multiple-choice questions (MCQ) - assessment of knowledge using traditional MCQ is not objective and effective. Multiple choice questions are rarely considered a suitable substitute for traditional assessment of deep knowledge. Multiple choice questions are frequently used but there is no way to investigate non-functioning distractors in developed tests. This paper mentions advantages and disadvantages of traditional multiple choice questions and introduces a new way to investigate non-functioning distractors in developed tests based on information retrieval model. Unlike traditional statistical based analyzes above-mentioned approach allows test developers to analyze test quality before knowledge assessment process. This paper will not cover technical details for suggested model.

Index Terms—Knowledge assessment, Multiple choice questions, Information retrieval system, Quality analyze.

I. INTRODUCTION

Multiple-choice questions are now a strongly preferred testing instrument across many higher education disciplines and business students in particular. The concept of multiple-choice questions (MCQs) is not new [1][2] and has been adopted in a variety of disciplines including physical education, economics [2], mathematics [3], psychology and fine arts [4] and the physical sciences. Ultimately, there is no one ideal assessment process and so too, MCQ testing has its advantages and disadvantages.

A traditional multiple choice question (or MCQ) is one in which a student chooses one answer from a number of choices supplied. A multiple choice question consists of:

• Stem - the text of the question
• Options - the choices provided after the stem
• Key - the correct answer in the list of options
• Distracters - the incorrect answers in the list of options. [5]

There are several advantages to multiple choice tests:

• Compared with traditional assessment questions, MCQs offer faster marking, less necessity for cross-marking.
• MCQs are useful to educational institutions under pressure to reduce unit costs.
• MCQs can be administered into on-line assessments, such online assessments can be very effective, and can prompt correct answers directly after completion with clarification and reasoning.[6]

There are a number of commonly raised objections to the use of MCQs. In particular, many of them claim that:

• There is no tools or approaches how to investigate non-functioning distractors in developed tests beside statistical analyzes.
• With MCQs there is a possibility of guessing the correct answer, there are numerous methods to penalize students from guessing such as negative marking (not recommended as sometimes it produces negative effects to students who know the answers), more options to answers, adopting mathematical strategies to normalize marks, giving partial marks to an answer very near to the correct answer.
• MCQs cannot test oral or written skills; it can test only the theories.
• They assess only trivial recognition of facts, rather than high-level thinking, such as exercising judgment and synthesis
• They offer choices for answers, rather than ask the candidate to construct the answer. [6]

II. TEST QUALITY

Distractors play a vital role for the process of multiple-choice testing, in that good quality distractors ensure that the outcome of the tests provides more credible and objective picture of the knowledge of the testees involved. On the other hand, poor distractors would not contribute much to the accuracy of the assessment as obvious or too easy distractors will pose no challenge to the students and as a result, will not be able to distinguish high performing from low performing learners.

Despite the existing body of research that evaluates the optimal number of distractors in multiple-choice items, substantially less research has focused on examining non-functioning distractors in MCQs in general [7] and no recent studies have specifically examined the frequency of non-functioning distractors in teacher-generated items. Items
generated automatically are another painful topic (usually they contain more non-functioning distractors than items generated by teachers), however for us there is no differences how they are generated.

Examining distractor quality in generated tests is of our interest because the majority of tests which students take are teacher-generated and teachers spend a large amount of time developing test items. If this time can be reduced, it will have great practical significance to teaching faculty. Additionally, there is a need for more researches on the distractor quality in multiple-choice tests from different perspectives, including observational, item analytic and item similarity perspectives [8].

Test Item quality analysis results can be used to identify and remove non-functioning distractors from tests. Items with more functioning distractors were more difficult and more discriminating. If properly constructed, MCQs are able to test higher levels of cognitive reasoning and can accurately discriminate between high- and low-achieving students [9,10]. It is widely accepted, however, that well-constructed MCQ items are time consuming and difficult to write [11]. Furthermore, there is more to writing good MCQs than writing good questions. One aspect where many MCQs fail is in having effective distractors. Teachers often spend a great deal of time constructing the stem and much less time on developing plausible options to the correct answer. High quality MCQs, however, also need the options to be well written [12]. In a classroom setting where test items are designed to measure educational outcomes, distractors must perform acceptably and each distractor should be based on a common misconception about the correct answer [13]. Non-functioning distractors are options that are selected infrequently (<5%) by examinees or otherwise do not perform as expected or options that are far from problem domain or options that obviously incorrect. As such, these options should be removed from the item [12] or be replaced with a more plausible option.

Previous studies have used various methods for evaluating distractor quality, including response frequency (non-functioning is usually defined as <5%) [13,14], poor distractor discrimination, expert judgment [14,15], and examination of option characteristic curves (trace lines) [13]. Trace lines graphically display the response patterns of the item options but typically require a large sample of examinees (200+) [16].

III. INFORMATION RETRIEVAL MODEL

Issues related to analysis and evaluation of non-functional distractors can be solved by the theory of artificial intelligence. However, this theory has incredible complexity and enormous number of objects to research. How will be the further development of the theory and practice of artificial intelligence applied to issues of computer-based evaluation and investigation is difficult to predict [18, 22, 24]. Therefore, we have only one approach: look for ways to improve the control and assessment of knowledge without substituting teachers’ functionalities by computer solutions, but using the computer solutions for resolving knowledge assessment problems.

The main reason for limiting the power of computer based solution in the knowledge assessment sphere is ambiguity (multi-variant) in the representation of test items (distractors). The reasons of ambiguity are synonyms, homonyms and polysemous vocabulary, idiomatic turns, as well as numerous relations between lexical units, which are characteristics of any natural language. Above-mentioned characteristics need to be considered even in cases of the simplest test items. A similar problem exists in the development of information retrieval systems, which have to solve essentially the same issues as in the case of investigating and analyzing distractors in test items. However, information retrieval systems have a longer road of development compared with the e-learning systems. So it makes sense to investigate systems of information retrieval in order to explore the possibility of reusing concepts of information retrieval model and use it as a basis for analyzing test items’ distractors. [17, 23]

IV. DOCUMENT IMAGE AND DISCRIPTOR DICTIONARIES

In the descriptor based Information Retrieval Systems (IRS) there is a concept of Document Image (DI) which is a list of keywords or descriptors, and other information: numbers, names, product names, and others, reflecting the basic semantic content of the document where the document refers to any unit of information: article, a technical description of a device, a monograph, patent, scientific report, etc. The terms "keyword" and "descriptor" have much in common, but, strictly speaking, they are not synonymous. Keyword - a term which refers to the words of the document, many of which reflects its core semantic content. The process of generating document image (DI) from document is called Indexing. We used well known algorithms and solutions for keyword extractions such as KEA (Keyphrase extraction algorithm), Maui (Multi-purpose automatic topic indexing) and etc. [27]

Suppose that we want to make a document image of a paper (denoting it by the letter A), which refers to the java. First of all, this document image should contain the keyword "java" Then we are adding the words that characterize the content of the article from other positions. Indexed document goes to the general information retrieval array. Then we are searching documents from general information retrieval array for keyword "java". Search is performed by a simple algorithm. It takes the image of the first document, and checks whether it contain the word "java" If it contains, then the document is considered to be responding to query, and subject to extradition. If it does not contain - remains in the array. This operation is performed for each document image in array. As a result of searching all articles/documents whose document images contain the word "java" will be found. Among them will be the article A. If you perform a search on the phrase "platform-independent object-oriented programming
language" or "byte code", Article A will not be found, although these combinations is largely synonymous with the word "java". To avoid this situation we need to add keywords "platform-independent object-oriented programming language" and "byte code" to the document image of article A, considering them synonymous with the keyword "java". The same rule applies to other keywords: If they have synonyms, then all they need to be added in the document image. But there is another way. Its essence is that before creating information retrieval system (IRS), initially for a given subject preparing special dictionary of descriptors, in which synonymous terms are organized into groups and each group selects one term as a representative of this group, which is usually assigned to the code. Such representative is called "descriptor". "The descriptor (from Lat. Descrivo - describe) - lexical unit (word, phrase) of information retrieval language that serves to express the basic semantic content of documents" [20]. This definition implies that the descriptor does the same role as the keyword. For all keywords taken from the indexed text, with the help of Descriptor Dictionary corresponding descriptor codes are included to DI. The descriptor, as opposed to a keyword is chosen not from the indexed text, but from the special (regulatory) descriptor vocabulary. With the help of descriptor codes all keywords of document are assigned to an artificial semantic uniqueness. Denoting by \( K \) the code which corresponds to the synonyms of "java", "platform-independent object-oriented programming language", "byte code". Now when indexing articles on the java, document image will include only descriptor \( K \). We can easily exclude keywords "java", "platform-independent object-oriented programming language" and "byte code" from document image, because when searching for any synonyms associated with descriptor \( K \), first we are looking into the descriptor dictionary, finding the descriptor and then performing search by \( K \). Using Descriptor Dictionaries we are reducing the size of Document Image. [25]

V. MATCH CRITERION IN INFORMATION RETRIEVAL SYSTEM

In the simplest case, the search is performed on one keyword/attribute (as shown in the previous section). However, there are requests which consist of several keywords/attributes. By analogy with the document image in this case we speak about search query image (SQI). Having keywords and descriptors is not enough to perform search. We need to specify criterion, called the criterion of semantic matching or just matching criteria based on which the computer decides to provide the document as responding or not. Even if the search query image consists of a single descriptor \( K \), there are at least two criteria for issuance. The first one is formulated as follows: the document is issued, if document image contains a descriptor \( K \). Second: the document is issued, if its document image doesn’t contain \( K \). Of course, in practice the first case is more common, so if you perform search for one keyword/attribute you may avoid defining the matching criterion (as shown in the example of previous section).

However, if we have two or more keywords/attributes in the search, matching criteria must be specified always. For example, if the query formulation identified four keywords \( A_1, A_2, A_3, A_4 \), there may be criteria such as:

- The document is issued, if in his document image contains at least two keywords of the four specified;
- The document is issued, if the document image contains two keywords \( A_1 \) and \( A_2 \), or three keywords, \( A_2, A_3 \) and \( A_4 \), but there is no keywords \( A_1 \);
- The document is issued, if the document image contains all the given keywords, etc.

In more complex cases, verbal language match criterion may be too cumbersome and inconvenient to use. Therefore, we should use a mathematical model.

The simplest and most natural way is to define Boolean model. In construction of Boolean model, we introduce the following interpretation of the logical variables of Boolean functions. Let’s denote attributes/keywords as \( A_1, A_2, A_3, \ldots, A_n \), where \( n \) – is the number of attributes/keywords in request. They also will be viewed as logical arguments. In this case, we assume that if a word of search query image (SQI) \( A_i (i = 1, 2, 3, \ldots, n) \) is contained among the words of a document image (DI) of this document, then the logical argument \( A_i = 1 \). If the word \( A_i \) is absent among the words of DI, then \( A_i = 0 \).

The logical operations of conjunction, disjunction and inversion interpreted as follows. Suppose that search is performed on the basis of criteria specified by the conjunction of the form \( A_i A_j \). Then the document is considered responding, and subject to extradition if his DI contains both attributes/keywords \( A_i \) and \( A_j \) \( (i, j = 1, 2, 3, \ldots, n; i \neq j) \). When searching through the disjunction \( A_i + A_j \), the document is considered to be responsible for a given query if its document image contains at least one of the attributes/keywords of \( A_i \) and \( A_j \). In the case of search based on the inversion (negation) \( \neg A_i \) document is responding to the request if in his document image no attribute of \( A_i \). In all these cases it is assumed that the number of attributes/keywords that form an image of the document can be anything.

Boolean function, which will be called as a search/retrieval function, assigning to the retrieval of the document. We assume that the document responds to the request and subject to extradition if

\[
f(\neg A_1, \neg A_2, \neg A_3, \ldots, \neg A_n) = 1
\]

If the document doesn’t respond to a query then

\[
f(A_1, A_2, A_3, \ldots, A_n) = 0
\]

Consider an example. Suppose we want to search for: "Find all publications of the minimization of Boolean formulas with the help of diagrams".
For definiteness, here and henceforth we assume that all the documents are indexed and keywords are available i.e. we will not use descriptor dictionaries. This means that SQI should include not only the words of the query formulation, but also their synonyms, and other similar meaning words. We are selecting the following words from query:

A1 - minimization;
A2 - Boolean;
A3 - a formula;
A4 - diagram

The search/retrieval function is a conjunction of four arguments:

\[ f = \overline{A1} \overline{A2} A3 \overline{A4}. \]

Adding to these words respectively:

A5 - simplification;
A6 - logic;
A7 - function;
A8 - map

Taking into account the synonyms search/retrieval function becomes complicated:

\[ f = (\overline{A1} \overline{A5})(\overline{A2} + \overline{A6})(\overline{A3} + \overline{A7})(\overline{A4} + \overline{A8}). \]

Extending this function, by adding two new logical arguments A9 and A10:

\[ f = (\overline{A1} + \overline{A5})(\overline{A2} + \overline{A6})(\overline{A3} + \overline{A7})(\overline{A4} + \overline{A8}) + \\
(\overline{A4} + \overline{A8})(\overline{A9} + \overline{A10}) \]

Where A9 is Veitch and A10 is Karnaugh.

To the word "formula" and "function" should be added synonymous word "expression". Denoting this by symbol A11. Then we will have:

\[ f = (\overline{A1} + \overline{A5})(\overline{A2} + \overline{A6})(\overline{A3} + \overline{A7} + \overline{A11})(\overline{A4} + \overline{A8}) + \\
(\overline{A4} + \overline{A8})(\overline{A9} + \overline{A10}) \]

Dwell on this, though, more complex functions can be extended by different grammatical forms of keywords.

If we open all the brackets, i.e. represent an expression in disjunctive normal form (DNF), we will obtain a formula containing 28 conjunctions:

\[ f = \overline{A1} A2 \overline{A3} A4 + \overline{A1} A2 \overline{A3} \overline{A8} + \overline{A1} A2 \overline{A7} A4 + \\
+ \overline{A1} A2 \overline{A7} \overline{A8} + \overline{A1} A2 A11 A4 + \\
+ \overline{A1} A2 A11 \overline{A8} + + \overline{A4} A9 + \overline{A4} A10 + \\
+ \overline{A8} A9 + \overline{A8} A10 \]

Where each of the conjunctions describes a single sub-query.

Decoding the first four and last four conjunctions introducing sub-queries taking into account grammar rules:

A1 A2 A3 A4 - minimization of Boolean formulas by diagrams;
A1 A2 A7 A4 - minimization of Boolean functions by diagrams;
A1 A2 A7 A8 - the minimization of Boolean function by maps;
A4 A9 – Veitch diagram;
A4 A10 - Karnaugh diagram;
A8 A9 – Veitch map;
A8 A10 - Karnaugh map.

This shows that using Boolean functions gives opportunity to represent complex queries in relatively simple way.

Concluding this subsection with the following observation. Authors working in the field of mathematics, usually strictly distinguish between terms such as function and the formula explaining that any Boolean function can be represented in many ways: table, graph charts, etc. However, from the practical point of view, such strictness is not so important to distinguish between functions and formulas to emphasize in every phrase.

VI. MATCH CRITERION IN KNOWLEDGE ASSESSMENT SYSTEMS

Fig. 1 and 2 shows the scheme of automatic search of documentation, and automatic matching of functioning distractors, respectively.

Consider Fig. 1. It shows the basic operations as a result of which the subscriber receives the information he/she needs.

![Fig. 1. Scheme of automated search of documents](image-url)
The search doesn’t performed on the original publications; it’s made on document images in the form of lists of keywords. At the stage of comparing the search query image with the document image the logical arguments of Boolean function are defining, which correspond to words of SQI. Found values of the logical arguments are substituted into the Boolean function which simulates match criterion. If the function returns true value, then the document is issued to the subscriber. [20, 27]

Scheme of automated matching of distractors with documents presented in Fig. 2, is a complete analog of the automated search of documents.

As we already talked distractors play a vital role for the process of multiple-choice testing, in that good quality distractors ensure that the outcome of the tests provides more credible and objective picture of the knowledge of the testees involved. So finding non-functioning distractors in test items will improve whole assessment process. In our case we will find non-functioning distractors by finding functioning distractors, all remaining distractors will be considered as a non-functioning. As the automated solution in this case isn’t 100% accurate at the final stage received results would be nice to be reviewed by test administrators.

VII. OTHER SOLUTIONS/SUGGESTIONS

Previous studies have used various methods for evaluating distractor quality, including response frequency (non-functioning is usually defined as <5%). This approach is statistical option we should wait some time collect some statistical data and eliminate non-functioning distractors. In case of question base changes the cycle should be repeated.

It’s possible to use any of the theories of artificial intelligence to get semantic analysis of the distractors and documents. As we have no major achievements in artificial intelligence; in computer-aided learning can be used only experimentations.[12, 13, 16, 17, 18, 22]

VIII. CONCLUSION

Solving the problem of computer recognition of functioning and non-functioning distractors is possible in three main ways:

1) The use of any of the theories of artificial intelligence to get semantic analysis of the distractors and documents;

2) The use of the statistical analyzes of tests and eliminate non-functioning distractors. This approach isn’t effective because test items should be used in real examinations to collect statistical data;

3) The construction of a Boolean model: the distractors are evaluated based on models of information retrieval systems. Each distractor’s query image is compared with document image in order to detect is it functioning distractor or not, and the evaluation criteria specified by Boolean function.

To them as a special case joins a fourth option: the distractors are evaluated on the basis of special algorithms for generating models and comparing them with the distractors.

The first of these options doesn’t have practical value yet, since at present the progress achieved in addressing the problem of artificial intelligence in computer-aided learning can be used only in the highly specific special cases and not more than at the level of experimentation.

The second option is based on statistical analyzes.

The third option is an information retrieval model, where the basic tenets of the theory apply to the IRS on the construction of systems for automated assessment of knowledge. This option offers the possibility of simulation of Boolean functions of any criterion for evaluating the distractors and does not require their semantic analysis.

The fourth option is fully applicable in practice, but in a limited area. Mostly this area forms mathematical formulas that due to the transformations can be represented in multiple ways. If required, for example, to insert the missing letters, numbers to decline a noun, conjugate a verb, to formulate a theorem and so forth; it is difficult to say whether the general algorithm exists or not that generates the model for all similar cases.

As the most promising at the present time we can consider only the information-retrieval option. Information retrieval systems have a longer road of development compared with artificial intelligence systems. So it makes sense to use above-mentioned models to improve knowledge assessment quality. We are going to cover all of its technical implementation in the next upcoming paper.

IX. FUTURE PLANS

As we are tried our solution in technical subjects like programming, information security and etc… we plan to roll
out our trials to cover more natural subjects. We also intend to establish the connection between psychological and computer adaptive testing (CAT) techniques. In order to improve assessment and learning process we are going to merge abovementioned solution with the hint generation methods.

ACKNOWLEDGMENT

This research was supported by Research Center, State Engineering University, Yerevan, Armenia.

REFERENCES