



US010388177B2

(12) **United States Patent**
King et al.

(10) **Patent No.:** **US 10,388,177 B2**
(45) **Date of Patent:** **Aug. 20, 2019**

(54) **CLUSTER ANALYSIS OF PARTICIPANT RESPONSES FOR TEST GENERATION OR TEACHING**

(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE**, Cambridge, MA (US)

(72) Inventors: **Gary King**, Brookline, MA (US); **Brian Lukoff**, Boston, MA (US); **Eric Mazur**, Concord, MA (US)

(73) Assignee: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE**, Cambridge, MA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1265 days.

(21) Appl. No.: **13/871,627**

(22) Filed: **Apr. 26, 2013**

(65) **Prior Publication Data**

US 2013/0302775 A1 Nov. 14, 2013

Related U.S. Application Data

(60) Provisional application No. 61/639,260, filed on Apr. 27, 2012.

(51) **Int. Cl.**
G09B 7/00 (2006.01)

(52) **U.S. Cl.**
CPC **G09B 7/00** (2013.01)

(58) **Field of Classification Search**
CPC G09B 7/00; G06K 9/6218; G06K 9/6253; G06N 99/005
USPC 434/362
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,041,311 A *	3/2000	Chislenko	G06F 17/30699
			705/26.7
7,099,819 B2 *	8/2006	Sakai	G06F 17/2755
			704/1
7,716,217 B2 *	5/2010	Marston	G06Q 10/107
			707/728
7,840,556 B1 *	11/2010	Dayal	G06F 17/30442
			707/721
8,131,722 B2 *	3/2012	Sundaresan	G06F 17/3071
			707/736
2002/0123865 A1	9/2002	Whitney et al.	
2008/0126319 A1 *	5/2008	Bukai	G06F 17/30864
2009/0202969 A1 *	8/2009	Beauchamp	G09B 7/00
			434/335

(Continued)

FOREIGN PATENT DOCUMENTS

CN	1185226 A	6/1998
CN	1971603 A	5/2007
CN	104620300 A	5/2015

OTHER PUBLICATIONS

Schapiro et al., A Boosting-based System for Text Categorization, Machine Learning, 39(2/3):135-168, 2000.*

(Continued)

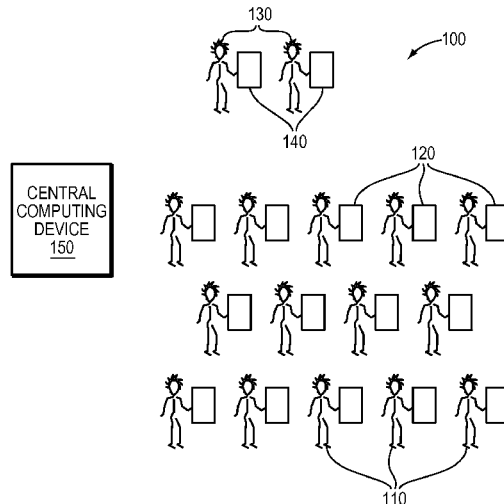
Primary Examiner — Eddy Saint-Vil

(74) Attorney, Agent, or Firm — Quarles and Brady LLP

(57) **ABSTRACT**

Textual responses to open-ended (i.e., free-response) items provided by participants (e.g., by means of mobile wireless devices) are automatically classified, enabling an instructor to assess the responses in a convenient, organized fashion and adjust instruction accordingly.

20 Claims, 4 Drawing Sheets



(56)

References Cited

U.S. PATENT DOCUMENTS

2010/0005089 A1* 1/2010 Yacobi G06Q 10/04
707/E17.014
2010/0235854 A1* 9/2010 Badgett G09B 7/06
725/24
2011/0093796 A1 4/2011 Plummer et al.
2012/0246161 A1* 9/2012 Murakami H04N 21/252
707/737
2013/0254216 A1* 9/2013 Yamamoto G06F 17/30663
707/749
2014/0059582 A1* 2/2014 Knowles G06Q 10/10
725/24

OTHER PUBLICATIONS

Hotho et al., A Brief Survey of Text Mining, *Journal for Computational Linguistics and Language Technology*, vol. 20 (2005), pp. 19-62.*
International Preliminary Report on Patentability dated Oct. 28, 2014 for International Patent Application PCT/US2013/038376 filed Apr. 26, 2013, 8 pages.
International Search Report and Written Opinion dated Jul. 10, 2013 for International Application No. PCT/US2013/038376 (6 pages).
Demark et al "Using Statistical Natural Language Processing for Understanding Complex 1-15 Responses to Free-Response Tasks", *International Journal of Testing*, vol. 4, Issue 4, 2004 retrieved from the internet. Retrieved on [Jun. 24, 2013] <URL:<http://www.education.umd.edu/EDMS/mislevy/CLI/NaturalLanguageProcessing.pdf>>entire document.

Attali, Yigal & Burstein, Jill, "Automated Essay Scoring With E-rater v.2," *Journal of Technology, Learning, and Assessment*, 4(3), 1-31 (2006).
Frey, Brendan J. and Dueck, Delbert, "Clustering by Passing Messages Between Data Points," *Science*, 315(5814):972-976 (2007).
Gan, Guojan, Ma, Chaoqun and Wu, Jianhong, "Data Clustering: Theory, Algorithms, and Applications," *Society for Industrial and Applied Mathematics*, 488 pages (2007).
Grimmer, Justin & King, Gary, "General Purpose Computer-Assisted Clustering Methodology: Supplemental Notes," Department of Political Science, Stanford University, Dec. 2, 2010, 21 pages.
Grimmer, Justin & King, Gary, "General Purpose Computer-Assisted Clustering and Conceptualization," *Proceedings of the National Academy of Sciences*, 108(7): 2643-2650 (2011).
Grimmer, Justin & King, Gary, "General purpose computer-assisted clustering and conceptualization," *Proceedings of the National Academy of Sciences*, Dec. 22, 2010, <http://www.pnas.org/content/108/7/2643> [Retrieved from the internet on Dec. 16, 2015 at 3:22:24 PM] 3 pages.
Haladyna, Thomas M., Downing, Steven M., & Rodriguez, Michael C., "A review of multiple-choice item-writing guidelines for classroom assessment," *Applied Measurement in Education*, 15(3), 309-334 (2002).
Kaufman, Leonard and Rousseeuw, Peter, "Finding Groups in Data: An Introduction to Cluster Analysis," New York: Wiley, 355 pages (1990).
Notification of First Office Action for Chinese Patent Application No. 201380033521.0, 11 pages. English Abstract.
Bennett, R.E., Morley, M., & Quardt, D., "Three Response Types for Broadening the Conception of Mathematical Problem Solving in Computerized Tests," *Applied Psychological Measurement*, 2000, 24(4): 294-309.

* cited by examiner

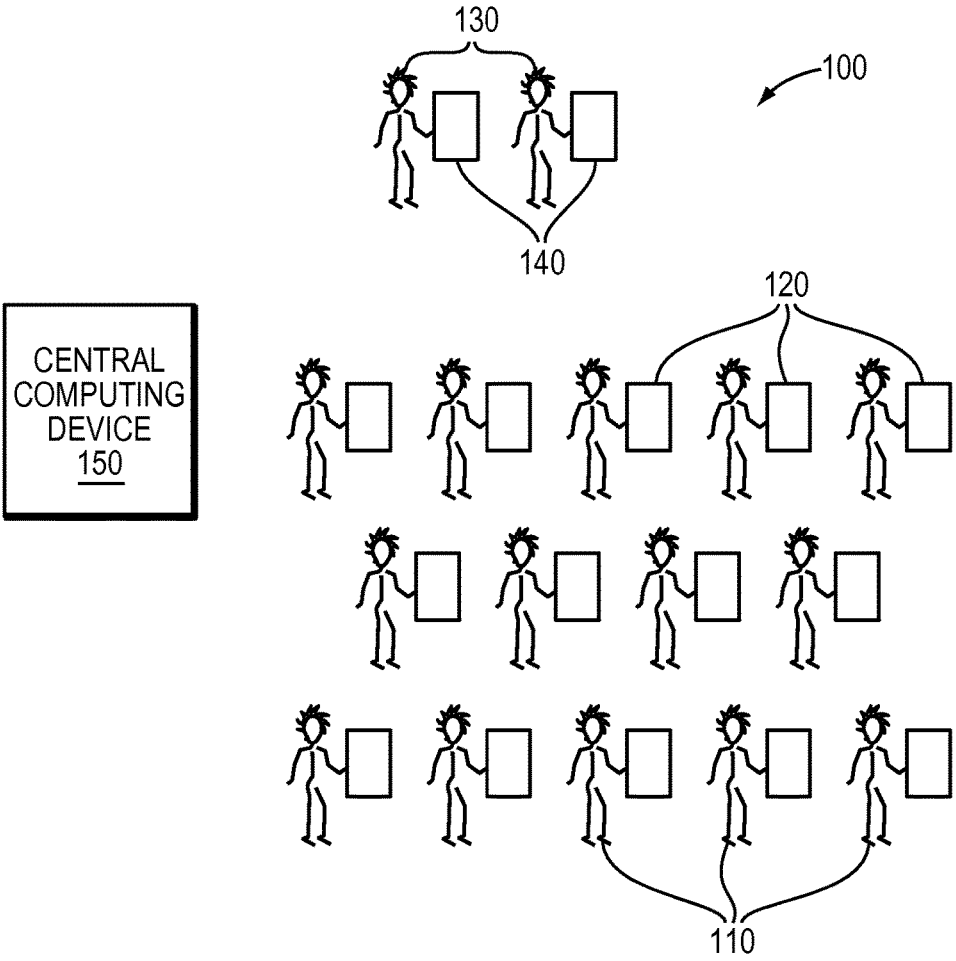


FIG. 1A

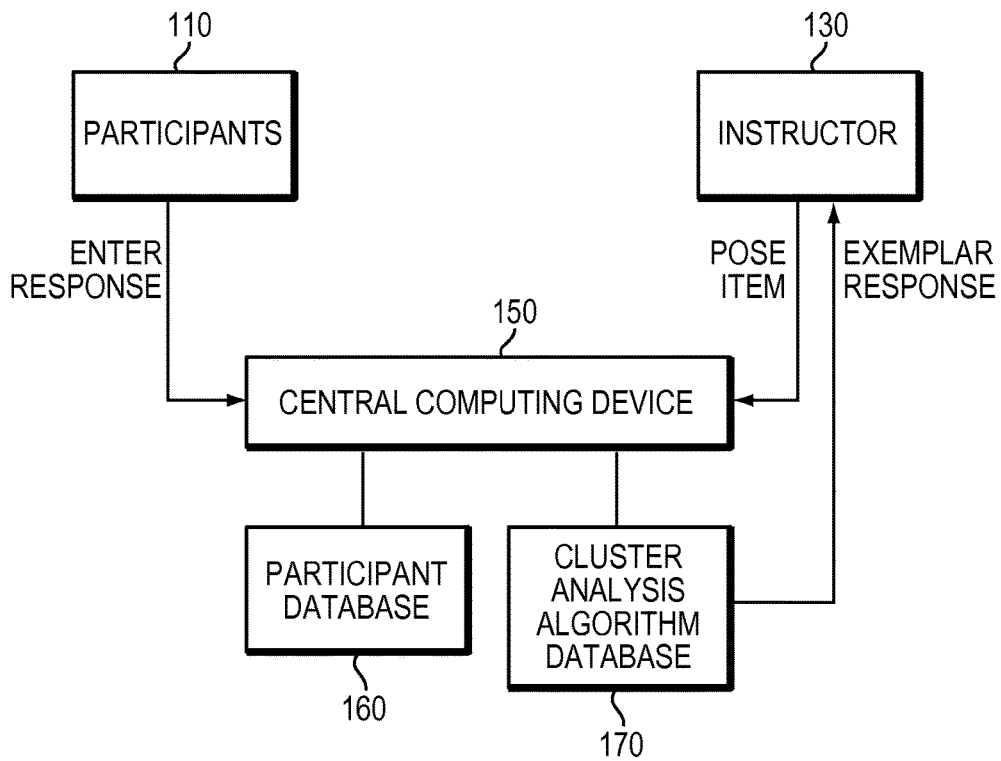


FIG. 1B

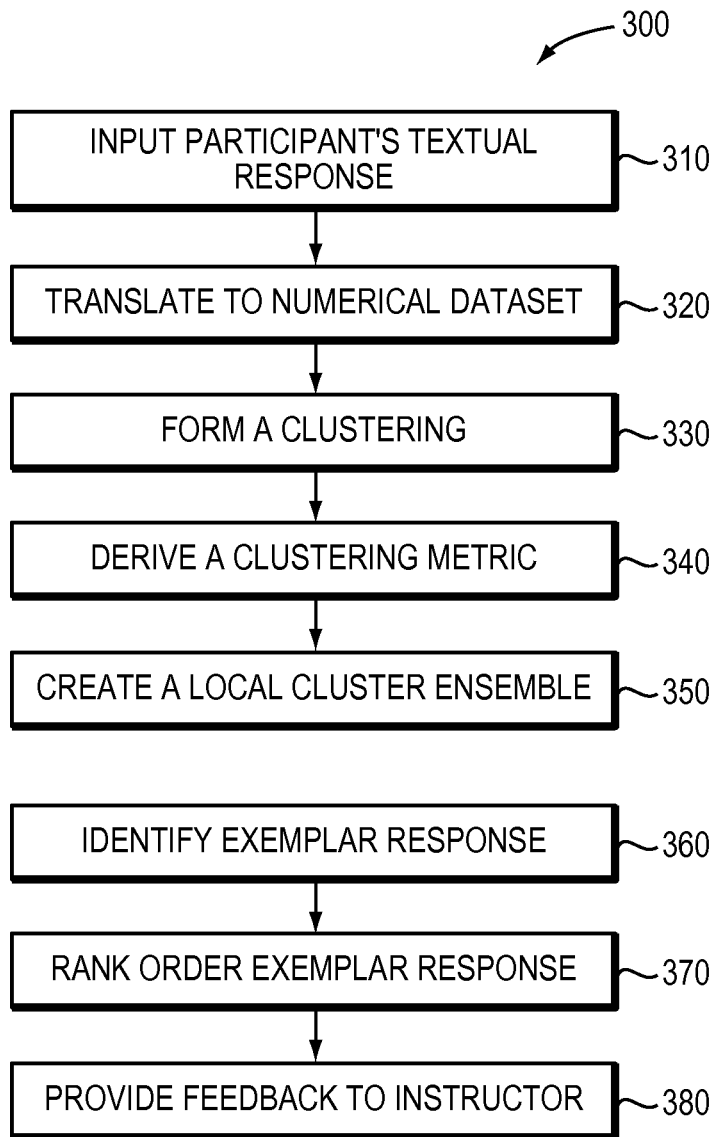


FIG. 2

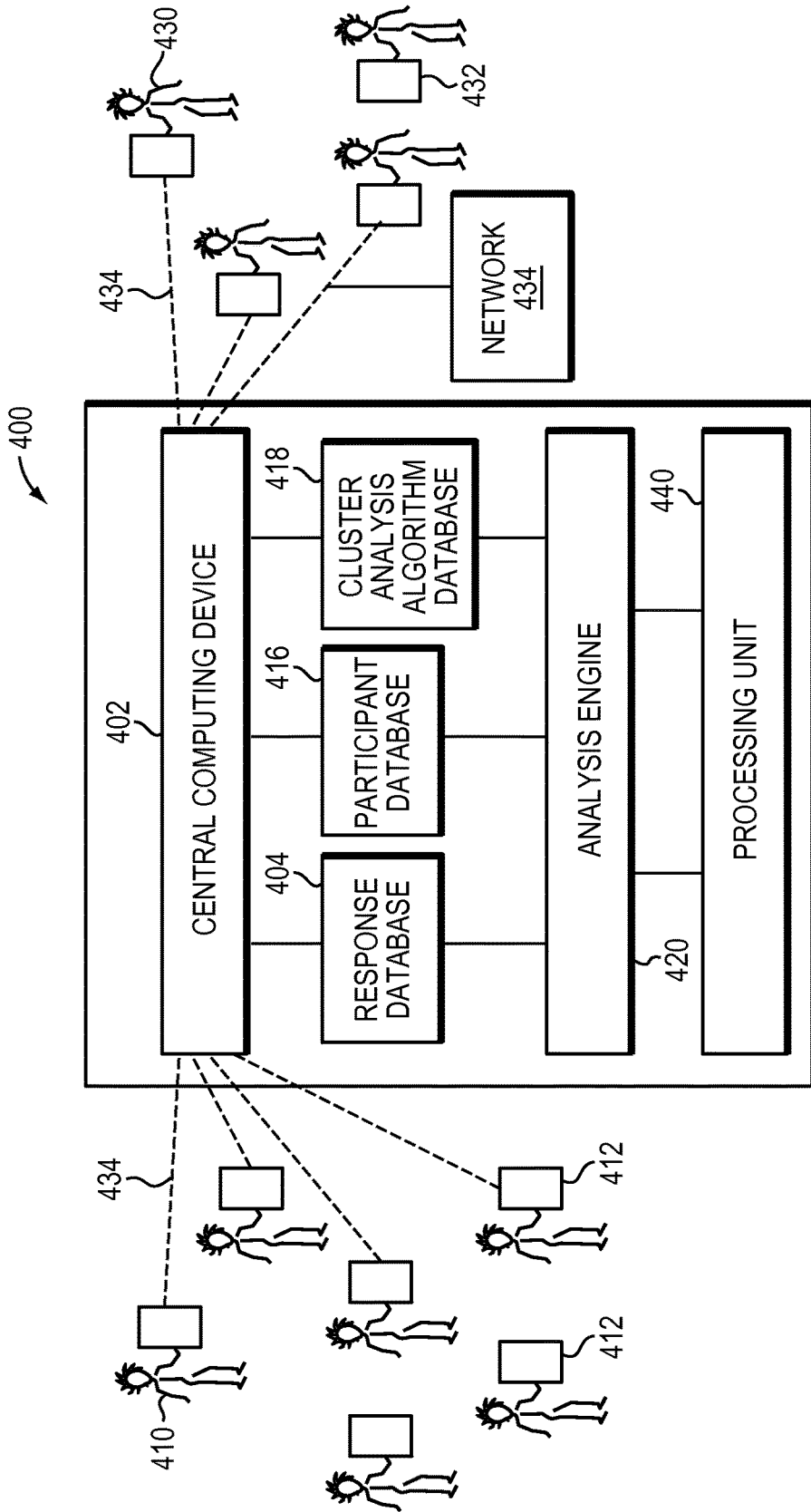


FIG. 3

CLUSTER ANALYSIS OF PARTICIPANT RESPONSES FOR TEST GENERATION OR TEACHING

CROSS-REFERENCE TO RELATED APPLICATION

This application claims priority to and the benefit of, and incorporates herein by reference in its entirety, U.S. Provisional Patent Application No. 61/639,260, which was filed on Apr. 27, 2012.

GOVERNMENT RIGHTS

This invention was made with Government support under National Science Foundation award IIS-0835338. The Government has certain rights in the invention.

FIELD OF THE INVENTION

In various embodiments, the present invention relates generally to cluster analysis of participant responses for test generation or teaching.

BACKGROUND

Multiple-choice items are easy to administer from a technical standpoint, but are often difficult and time-consuming to write because they require a comprehensive understanding of all possible responses that a participant might give. For questions with a single correct answer, this means anticipating all likely wrong answers based on a deep understanding of all the misconceptions and misunderstandings that might lead a participant to respond incorrectly. Even for questions without a single correct answer (e.g., attitudinal questions), it can be difficult to predict all possible opinions that a participant might have when faced with the item prompt. Developing response options for multiple-choice items is challenging enough to merit numerous sets of published guidelines and rules; see, e.g., Haladyna, Downing, & Rodriguez, "A review of multiple-choice item-writing guidelines for classroom assessment," *Applied Measurement in Education*, 15(3), 309-333 (2002).

In contrast, asking participants to respond to free-response items instead of multiple-choice items solves this particular problem but presents its own set of difficulties. Contrasted with supervised learning analysis, which may involve either an algorithmic definition or a training set of classified entries, the underlying clustering of the unsupervised learning for the participant responses is unknown. In addition, some rare types of free-response items can be scored automatically (see, e.g., Attali & Burstein, "Automated essay scoring with e-rater v.2," *Journal of Technology, Learning, and Assessment*, 4(3), 1-31 (2006); Bennett, Morley & Quardt, "Three Response Types for Broadening the Conception of Mathematical Problem Solving in Computerized Tests," *Applied Psychological Measurement*, 24(4), 294-309 (2000)), however, the time and monetary costs to set up and use such a system can make them infeasible. Furthermore, without a way of automatically scoring or classifying item responses, an instructor cannot process dozens or hundreds of participant responses in real time to dynamically adjust instruction, as is the goal with any formative assessment technique.

Consequently, there is a need for an approach using unsupervised learning that automatically classifies or scores

participant responses to open-ended items and provides real-time feedback to the instructor.

SUMMARY

The present invention addresses this problem by facilitating automatic classification or scoring of participant textual responses to open-ended (i.e., free-response) items, enabling an instructor to assess the responses in a convenient, organized fashion and adjust instruction accordingly. In various embodiments, students enter responses to the open-ended items posed by the instructor using a handheld device, e.g., a laptop computer or cell phone with web access. The term "item" as used herein refers to a question that an instructor poses to participants and that participants respond to, e.g., using handheld devices. Items may attempt to assess factual information, tease out conceptual difficulties, or measure participants' opinions. Although the present discussion focuses on students in a college classroom environment, it should be understood that the approach described herein is applicable to participants in any group pedagogical or intellectual endeavor, and the terms "student" and "participant" are used herein interchangeably.

Key to the present invention is the application of clustering techniques to identify "exemplar" responses from a collection of student answers. Whereas today instructors sift through long lists of answers and haphazardly read a few in an effort to detect, for example, the various ways in which students may stray from proper reasoning, the present invention utilizes cluster analysis to find the different types of answers, mistakes, and ideas students give, and based on this analysis, picks one or a small number of exemplar answers for the instructor to much more quickly read.

Accordingly, in a first aspect, method of analyzing and sorting at least one participant's textual responses to a free-response item. In various embodiments, the method comprises the steps of providing the free-response items to a plurality of participants; receiving textual responses from at least some of the participants; computationally processing the textual responses into clusters of response types; identifying an exemplar that typifies each of the response types; and reporting the textual responses corresponding to the exemplars. The method may further comprise generating a multiple-choice version of the free-response item using a plurality of the most highly-ranked exemplar responses.

In some embodiments, the textual responses are processed into clusters represented as numerical vectors, and the method may further comprise the steps of computationally clustering the numerical vectors according to a plurality of clustering algorithms, where each algorithm produces an ensemble of clusters in a vector space and the resulting plurality of ensembles overlaps no more than partially in the space; within at least some of the clusters, identifying an exemplar vector that is spatially closest to the largest number of other vectors; ranking each of the identified exemplar vectors in terms of its distance from the other exemplar vectors; and reporting the ranked list.

The method may further comprise pruning cluster ensembles having a similarity to other cluster ensembles that exceeds a threshold. In some embodiments, identifying the exemplar comprises measuring the distance between any two responses based on similarity of a number of common stemmed n-grams in each response. The free-response items are wirelessly provided to handheld devices associated with the participants in a participant database.

In some embodiments, the method further comprises the step of deriving a metric specifying a distance between

clusters in the ensembles, where exemplar vectors are identified only for clusters whose distance metric from all other clusters exceeds a threshold.

The item may be a question and the method may further comprise the step of using the exemplars to predict prototypical responses to the question. The method may comprise the step of computationally generating choices in a multiple-choice version of the question.

In another aspect, the invention pertains to a system for analyzing and sorting at least one participant's textual responses to a free-response item. In various embodiments, the system comprises a memory for storing a participant database comprising records each specifying a participant, a mobile device associated with the participant, and textual responses to free-response items provided by the participant via a mobile device; and a processor in operative communication with the memory. The processor is configured to (i) computationally process the textual responses into clusters of response types; (ii) identify an exemplar vector that typifies each of the response types; and (iii) cause the textual responses corresponding to the exemplar vectors to be displayed.

In various embodiments, the processor is further configured to (i) computationally process the textual responses into numerical vectors; (ii) computationally cluster the numerical vectors according to a plurality of clustering algorithms, where each algorithm produces an ensemble of clusters in a vector space and the resulting plurality of ensembles overlaps no more than partially in the space; within at least some of the clusters, identify an exemplar vector that is spatially closest to the largest number of other vectors; (iv) rank each of the identified exemplar vectors in terms of its distance from the other exemplar vectors; and (v) cause the textual responses corresponding to the ranked exemplar vectors to be reported in a ranked list to the instructor. The processor may be further configured to prune cluster ensembles having a similarity to other cluster ensembles that exceeds a threshold and/or to identify the exemplar vector by measuring the distance between any two responses based on similarity of a number of common stemmed n-grams in each response.

In some embodiments, the processor is further configured to generate a multiple-choice version of the free-response item using a plurality of the most highly-ranked exemplar responses. The system may wirelessly communicate the free-response items to handheld devices associated with the participants in a participant database.

Reference throughout this specification to "one example," "an example," "one embodiment," or "an embodiment" means that a particular feature, structure, or characteristic described in connection with the example is included in at least one example of the present technology. Thus, the occurrences of the phrases "in one example," "in an example," "one embodiment," or "an embodiment" in various places throughout this specification are not necessarily all referring to the same example. Furthermore, the particular features, structures, routines, steps, or characteristics may be combined in any suitable manner in one or more examples of the technology. The headings provided herein are for convenience only and are not intended to limit or interpret the scope or meaning of the claimed technology.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings, like reference characters generally refer to the same parts throughout the different views. Also, the drawings are not necessarily to scale, with an emphasis instead generally being placed upon illustrating the prin-

ciples of the invention. In the following description, various embodiments of the present invention are described with reference to the following drawings, in which:

FIG. 1A schematically illustrates an activity involving communication between participants and instructors;

FIG. 1B is a block diagram illustrating deployment of a representative system in accordance with the current invention;

FIG. 2 depicts a method for cluster analysis and feedback generation in accordance with embodiments of the current invention; and

FIG. 3 illustrates a system for analyzing participant response in accordance with embodiments of the current invention.

DETAILED DESCRIPTION

FIG. 1A depicts an exemplary event or activity **100** where a gathering of participants or students **110**, each having a handheld device **120**, respond to one or more open-ended items posed by one or more instructors or organizers **130**, who also may each have a handheld device **140**. Again, the nature of the activity is not material to the invention, nor are the specific roles of instructor and participant; instructor(s) **130** may, for example, be chosen from the participants **110** or may be a different group of people. What is important is for the activity to involve the transmission of the posed open-ended items and the ability to gather responses from the participants.

The participants **110** and the instructor(s) **130** may directly communicate via the handheld devices **120**, **140** or indirectly via a central computing device **150**. The handheld devices **120**, **140** may be, for example, computers, cell phones, tablets or other electronic devices that transmit participants' responses, attitudes, opinions, knowledge, characteristics, and/or understanding of relevant subject matter or items to one another, to the instructor(s) **130** and/or to the central computing device **150** via a network. The transmitted subject matter may be in response to a question or a discussion topic that the instructor(s) **130** pose(s) to participants **110**. In one representative implementation, the participants **110** are students, the event **100** is a class and the instructor **130** is a professor or teacher who wishes to gauge student understanding of pedagogic content. The students may be co-located or geographically dispersed (e.g., participating in distance learning classes). In another embodiment, the activity is a corporate training event attended by employees or a workshop attended by workshop participants where the session leader wishes to gauge the opinions or responses of the participants. A person of ordinary skill in the art will appreciate that the invention described herein may be applicable to any group of participants pursuing pedagogical, intellectual or business endeavors.

In accordance with the invention, an algorithm adapted to the problem of automatically selecting a small set of exemplar responses is applied, and the instructor may use the resulting exemplar responses as the basis for whole-class discussion in real time, or as response options for a new multiple-choice question that can be administered in a future event. Exemplars may be selected via text analysis of open-ended text responses automatically as described in the ensuing discussion.

Referring to FIG. 1B, at the beginning of the event, the instructors **130** may wirelessly or otherwise (e.g., by means of a wired connection) transmit their open-ended items to the participants (collectively indicated at **110**) directly or to a central computing device **150**, which communicates the

items to the participants **110** via a network; the participants **110** then may wirelessly or otherwise (e.g., by a wired connection) transmit their responses to the items to the central computing device **150** via a network, where their responses are stored in a participant database **160**. The central computing device **150** may be implemented in the instructor's handheld device or in a separate, independent system in wireless (or wired) communication therewith. The instructor **130** may select one or more cluster-analysis algorithms for use in connection with the items that will be provided to the participants **110**, and transmits the selected option(s) to (or enters it into) the central computing device **150**; cluster analysis algorithms are stored in an algorithm database **170**. The central computing device **150** performs clustering based on characteristics of the participants and the selected cluster analysis algorithm. The results of the analysis typically include a computed exemplar response of the clustering (and its rank order), which is automatically communicated to the the instructors **130** via their handheld devices.

The approach of the invention is to cluster and analyze student responses so that conceptually similar responses are grouped together. Embodiments of the current invention identify an "exemplar response" (i.e., a response that exemplifies whatever pattern is apparent among the responses in that cluster) from each meaningful cluster and then present to the instructor one or more of these exemplar responses that are found to be most different from each other.

In one embodiment, it is recognized that different algorithms and different parameters for those algorithms will cluster responses together in different ways. Rather than trying to focus on the "best" way of clustering all of the responses, the present invention recognizes that there are many meaningful ways to do so. Accordingly, in this embodiment, multiple clustering algorithms are preferably applied, resulting in overlapping clusters because different algorithms produce different partitions of the responses. If there are 20 responses, for example, there are about a million different possible clusters, but perhaps the algorithms employed will together identify only a small subset of those as meaningful (i.e., grouping a set of responses that are all similar in some way). A representative approach adapts the computer-assisted method set forth in Grimmer & King, "General Purpose Computer-Assisted Clustering and Conceptualization," Proceedings of the National Academy of Sciences (2011) (hereafter "G&K," the entire disclosure of which is hereby incorporated by reference).

In other embodiments, one or more traditional clustering algorithms are employed. Any clustering algorithm that permits identification of exemplar responses may be used to advantage. Suitable techniques are described, for example, in Guojan Gan, Chaoqun Ma, and Jianhong Wu, *Data Clustering: Theory, Algorithms, and Applications*, Siam, 2007; Leonard Kaufman and Peter Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: Wiley (1990), and in Frey, B. J. and Dueck, D., "Clustering by Passing Messages Between Data Points," *Science* 315(5814):972 (2007). The entire disclosures of these reference are hereby incorporated by reference.

In the following discussion, the term "n-gram" refers to a sequence of n words from a document (in this case free-text student responses). So, in the previous sentence, each word is a 1-gram, the phrases "a sequence" and "sequence of" are 2-grams, etc. "Stemming" is the process of taking a word and reducing it to its basic root. For example, in the previous sentence the stems would be "stem" for "stemming," "reduce" for "reducing", etc. In this way, the terms "reduce"

and "reducing" are recognized as the same for purposes of analysis. A "stemmed n-gram" is an n-gram where each word is reduced to its stem. Algorithms for stemming are conventional in the art.

In some embodiments, the procedure represents each clustering as a point in a two-dimensional space, such that clusterings (points) close together in the space are almost the same, and those farther apart may warrant a closer look because they differ in some important ways. In effect, this visualization translates the uninterpretable chaos of huge numbers of possible clusterings into a simple framework that instructor can comprehend and use to efficiently select one or small number of clusterings that conveys the most useful information.

A representative method **300** for analyzing participants' responses to an open-ended item in accordance with embodiments utilizing the G&K approach is shown in FIG. **2**, and the representative steps for creating a vector space of clusterings are as follows.

First, a set of textual documents **310** is translated to a numerical dataset in the step of **320**. This step is necessary only when the items to be clustered are text documents or in general not already numerical. For each set of text documents of variable length, a common framework is used to represent them quantitatively. This may involve transforming the text to numbers by removing punctuation, stemming the words, computing 1, 2, 3, 4, and 5-grams, removing stemmed n-grams that appear in fewer than 1% or more than 99% of the responses, and finally representing each response as a vector summarizing the presence or absence (or for some purposes a count) of each stemmed n-gram. For English documents, about 3,500 unique word stems usually remain in the entire corpus. Each document is then coded with a set of (about 3,500) variables, each corresponding to the number of times a word stem is used in that document. Since most human language is highly repetitive, this representation is usually more than adequate. This general procedure can also be applied to multiple representations of the same documents, e.g., term-weighting representations such as tf-idf (term frequency/inverse document frequency), part-of-speech tagging, tokenization rules such as replacing "do" and "not" with "do_not", etc. Likewise, variants of kernel methods—procedures to produce a similarity metric between documents without explicitly representing the words in a matrix—can also be utilized as the basis for the numerical representation.

In step **330**, a plurality of clustering algorithms is applied, one at a time, to the numerical dataset. Each approach may represent different substantive assumptions that are difficult to express before their application, but desirable effects result from the fact of their differences rather than the nature of those differences. Virtually any computationally efficient clustering methods can be employed in combination in this step, so long as the result is that each document assigned either to a single cluster or to different clusters with weights that sum to 1. Representative clustering techniques are listed in the SI Appendix to G&K, available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041127/> (incorporated herein by reference).

Ideally, the clusters generated by the different clustering algorithms are separable, i.e., they produce clusters that do not overlap significantly if at all. One approach to generating such clusters is to use as many different clustering techniques as necessary to produce, from the numerical dataset, three or four clusters that do not overlap. Alternatively, the approach of G&K may be used and a large number of clustering algorithms applied to the dataset, thereby forming

a space (e.g., a vector space) of clusterings; then, three or four points equidistant from each other in the space are selected and those clusters employed for the remainder of the analysis as described below. Still another approach is to apply several existing cluster analysis methods, such as the well-known k-means algorithm, and use the outputs of these directly—specifically, the output of each of these algorithms is a clustering of responses, which is added to the clusterings found by other algorithms. But it may be that, for example, k-means produces a clustering that is not very different from the clustering produced by other algorithms (e.g., if there were 26 responses A-Z and one algorithm produces two groups, A-K and L-Z, and k-means produces two groups consisting of A-L and M-Z). Accordingly, some of these clusterings may be discarded if they are too similar to the clusterings already found using other methods, since they do not enhance the separability exploited by the current invention.

In step 340, a distance metric is derived to measure the similarity between any pair of clusters. Each clustering represents a partitioning of the participant responses to an item into mutually exclusive and exhaustive categories—that is, each clustering represents an ensemble of clusters. Typically, the clusterings are relatively dissimilar. The similarity metric should satisfy several criteria. First, the distance should be a function of the number of pairs of documents not placed together (i.e., not found in the same cluster) in both clusterings. Second, the distance should be invariant to the number of documents, given any fixed number of clusters in each clustering. Third, a scale is set by fixing the minimum distance to zero and the maximum distance to $\log(k)$, where k is the number of clusters.

In step 340, the metric established in step 330 is used to create a metric space of clusterings, along with a lower dimensional Euclidean representation useful for visualization. The matrix of distances between each pair in the set of J clusterings can be represented in a J -dimensional metric space, which is projected down to two Euclidean dimensions for visualization purposes. Because projection entails the loss of information, a multidimensional scaling method that retains the most crucial information is employed. In particular, it is the small distances that should be preserved most accurately because these reflect the clusterings to be combined (in the next step) into local ensembles. As the distance between two clusterings increases, a higher level of distortion will affect the results less. The well-known Sammon multidimensional scaling algorithm satisfies these criteria and may be used to advantage.

Optionally, in step 350, a “local cluster ensemble” method is used as a way to summarize any point in the space. This allows for the fast exploration of the space, ensuring that users of the system are able to quickly identify partitions useful for item on which the participants have responded. A “cluster ensemble” is a technique used to produce a single clustering by averaging in a specific way across many individual clusterings. To preserve local diversity and avoid the infinite regress resulting from clustering a set of clusterings, the method generates local cluster ensembles, which are defined as a new clustering created at a point in the space of clusterings from a weighted average of nearby existing clusterings. The local cluster ensemble approach will approximate more possible clusterings as additional clustering methods are employed in step 330, and will never be worse, and usually will be considerably better, in approximating a new clustering than the closest existing observed

point. Cluster ensembles having a degree of similarity to other cluster ensembles in excess of a threshold may be pruned.

In step 360, for each cluster or cluster ensemble within each clustering, an exemplar response is identified—that is, given a group of responses that a clustering algorithm has identified as all being similar, one of these responses that exemplifies this similarity is selected. One way to do this is to measure the distance between any two responses based on similarity of the number of common stemmed n -grams in each response, so that the number of stemmed n -grams present in both is used as the measure of similarity between any two responses. For example, the responses “increased voltage to battery” and “increasing volts” both have the stemmed 2-gram “increase volt” in common. The single response in each cluster with the smallest (spatial) distance to the largest number of other responses is identified as the exemplar.

In step 370, the exemplars from all the clusters within all the clusterings are rank-ordered from most to least distinctive. For this purpose, “distinctive” is defined as the distance from (i.e., less similar to) the set of other exemplars, so that the farther an exemplar is from the other exemplars, the higher will be its rank. With reference to FIGS. 1A and 2, in step 380 these exemplars are presented in ranked order to the instructor(s) 130, who will read them to gain a very quick view of the understandings and misunderstandings of students in the class. Alternatively, the exemplars can be used to predict prototypical responses to questions in order to provide choices in a multiple-choice version of the question.

A representative system 400 for analyzing participants’ responses in accordance with embodiments of the current invention is shown in FIG. 3. The system 400 is based on the framework introduced in FIG. 1A, and comprises a central computing device 402 that includes (or is in communication with) a database 404 in which responses are stored. An ungrouped set of participants 410 each has a handheld device 412; as explained above in connection with FIG. 1A, the participants may be co-located or may be geographically dispersed and communicate via a network 434. Items are provided to, and responses are received from participants via the handheld devices 412.

The central computing device 402 includes (or is in communication with) a participant database 416, which contains records for the participants 410 as well as for other individuals who may be used in a training set; the database 416 may contain records relating each participant to a handheld device, for example. A database 418 contains cluster algorithms utilized as described above in connection with analysis of the received responses. The computing device 402 executes an analysis engine 420 that performs the steps shown in FIG. 2, accessing responses stored in the database 404 and performing the clustering analysis described above utilizing the entries in the databases 404, 418. Based on the analysis, the central computing device 402 communicates the exemplars in ranked order to the handheld devices 430 of the instructors 432. Typically, communication with the handheld devices 412 and 432 occurs wirelessly, e.g., by SMS, an automated telephone call, e-mail or other suitable form of communication.

The analysis engine 420 may be implemented by computer-executable instructions, such as program modules, that are executed by a conventional computer. Generally, program modules include routines, programs, objects, components, data structures, etc. that performs particular tasks or implement particular abstract data types. Those skilled in the art will appreciate that the invention may be practiced with

various computer system configurations, including multi-processor systems, microprocessor-based or programmable consumer electronics, minicomputers, mainframe computers, and the like. The invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote computer-storage media including memory storage devices.

The central computing device **402** may comprise or consist of a general-purpose computing device in the form of a computer including a processing unit **440**, a system memory, and a system bus that couples various system components including the system memory to the processing unit. During operation, the system memory contains the instructions implementing the analysis engine **420**. Computers typically include a variety of computer-readable media that can form part of the system memory and be read by the processing unit. By way of example, and not limitation, computer readable media may comprise computer storage media and communication media. The system memory may include computer storage media in the form of volatile and/or nonvolatile memory such as read only memory (ROM) and random access memory (RAM). A basic input/output system (BIOS), containing the basic routines that help to transfer information between elements, such as during start-up, is typically stored in ROM. RAM typically contains data and/or program modules that are immediately accessible to and/or presently being operated on by processing unit. The data or program modules may include an operating system, application programs, other program modules, and program data. The operating system may be or include a variety of operating systems such as Microsoft WINDOWS operating system, the Unix operating system, the Linux operating system, the Xenix operating system, the IBM AIX operating system, the Hewlett Packard UX operating system, the Novell NETWARE operating system, the Sun Microsystems SOLARIS operating system, the OS/2 operating system, the BeOS operating system, the MACINTOSH operating system, the APACHE operating system, an OPENSTEP operating system or another operating system of platform.

Any suitable programming language may be used to implement without undue experimentation the analytical functions described above. Illustratively, the programming language used may include assembly language, Ada, APL, Basic, C, C++, C*, COBOL, dBase, Forth, FORTRAN, Java, Modula-2, Pascal, Prolog, Python, REXX, and/or JavaScript for example. Further, it is not necessary that a single type of instruction or programming language be utilized in conjunction with the operation of the system and method of the invention. Rather, any number of different programming languages may be utilized as is necessary or desirable.

The computing environment may also include other removable/nonremovable, volatile/nonvolatile computer storage media. For example, a hard disk drive may read or write to nonremovable, nonvolatile magnetic media. A magnetic disk drive may read from or writes to a removable, nonvolatile magnetic disk, and an optical disk drive may read from or write to a removable, nonvolatile optical disk such as a CD-ROM or other optical media. Other removable/nonremovable, volatile/nonvolatile computer storage media that can be used in the exemplary operating environment include, but are not limited to, magnetic tape cassettes, flash memory cards, digital versatile disks, digital video tape,

solid state RAM, solid state ROM, and the like. The storage media are typically connected to the system bus through a removable or nonremovable memory interface.

The processing unit **440** that executes commands and instructions may be a general-purpose processor, but may utilize any of a wide variety of other technologies including special-purpose hardware, a microcomputer, mini-computer, mainframe computer, programmed microprocessor, microcontroller, peripheral integrated circuit element, a CSIC (customer-specific integrated circuit), ASIC (application-specific integrated circuit), a logic circuit, a digital signal processor, a programmable logic device such as an FPGA (field-programmable gate array), PLD (programmable logic device), PLA (programmable logic array), or any other device or arrangement of devices that is capable of implementing the steps of the processes of the invention.

The network **434** may be a wired or wireless local area network (LAN) and a wide area network (WAN), wireless personal area network (PAN) and/or other types of networks. When used in a LAN networking environment, the central computing device **402** is connected to the LAN through a network interface or adapter. When used in a WAN networking environment, the central computing device **402** may include a modem or other network interface, e.g., a standard telecommunications platform for cellular communication. Mobile devices **412**, **432** may be connected over the Internet, an Intranet, Extranet, Ethernet, the wireless telecommunications infrastructure, or any other system that provides communications. Some suitable communications protocols may include TCP/IP, UDP, or OSI for example. Furthermore, components of the system may communicate through a combination of wired or wireless paths.

The terms and expressions employed herein are used as terms and expressions of description and not of limitation, and there is no intention, in the use of such terms and expressions, of excluding any equivalents of the features shown and described or portions thereof. In addition, having described certain embodiments of the invention, it will be apparent to those of ordinary skill in the art that other embodiments incorporating the concepts disclosed herein may be used without departing from the spirit and scope of the invention. Accordingly, the described embodiments are to be considered in all respects as only illustrative and not restrictive.

What is claimed is:

1. A method of determining a plurality of textual responses to generate a future test option to correspond with a response to a free response item, the method comprising the steps of:

determining, at a server computer, one or more participant devices and one or more instructor devices, wherein the one or more participant devices and the one or more instructor devices are geographically dispersed and communicate via a network, wherein each of the one or more participant devices is operated by a respective participant and each respective participant is associated with a respective characteristic;

receiving, at the server computer, the free response item from one of the one or more instructor devices;

receiving, at the server computer, the plurality of textual responses from the one or more participant devices;

removing punctuation from the plurality of textual responses;

translating into a numerical dataset the plurality of textual responses resulting from the removing punctuation step;

11

applying, at the server computer, a clustering algorithm to the numerical dataset to generate one or more clusters, wherein the clustering algorithm is based at least in part on the respective characteristic;

generating the one or more clusters from the clustering algorithm; 5

determining a single cluster as a weighted average of the one or more clusters, wherein the determined single cluster is used to generate the future test option based at least in part on the plurality of textual responses and the determined single cluster; and 10

wirelessly providing the future test option by the server computer to the one or more instructor devices.

2. The method of claim 1, further comprising:

after generating the one or more clusters from the clustering algorithm, determining a distance metric associated with the one or more clusters to identify similarities between the one or more clusters, wherein the one or more clusters include at least one incorrect answer to the free response item. 20

3. The method of claim 2, wherein the distance metric analyzes a number of the plurality of textual responses not found in the same cluster.

4. The method of claim 2, wherein the distance metric is associated with a number between zero and $\log(a \text{ number of the one or more clusters})$. 25

5. The method of claim 2, further comprising:

ranking the one or more clusters based in part on the distance metric; and

transmitting the ranked clusters to the one or more instructor devices. 30

6. The method of claim 1, wherein the plurality of textual responses are translated to the numerical dataset by:

stemming one or more words in the plurality of textual responses; 35

computing 1, 2, 3, 4, and 5-grams;

removing stemmed n-grams that appear in fewer than 1-percent or more than 99-percent of the plurality of textual responses; and

generating the stemmed n-grams as a vector. 40

7. The method of claim 1, further comprising:

removing a second cluster from the one or more clusters, wherein the distance metric associated with the second cluster exceeds a threshold.

8. The method of claim 1, further comprising: 45

storing, by the server computer, participant device information in a participant database.

9. The method of claim 1, wherein four clusters are generated by the clustering algorithm to the numerical dataset. 50

10. One or more non-transitory machine-readable medium having machine-executable instructions configured to perform the machine-implementable method for:

determining, at a server computer, one or more participant devices and one or more instructor devices, wherein the one or more participant devices and the one or more instructor devices are geographically dispersed and communicate via a network, wherein each of the one or more participant devices is operated by a respective participant and each respective participant is associated with a respective characteristic; 60

receiving, at the server computer, a free response item from one of the one or more instructor devices;

receiving, at the server computer, the plurality of textual responses from the one or more participant devices; 65

removing punctuation from the plurality of textual responses;

12

translating into a numerical dataset the plurality of textual responses resulting from the removing punctuation step;

applying, at the server computer, a clustering algorithm to the numerical dataset to generate one or more clusters, wherein the clustering algorithm is based at least in part on the respective characteristic;

generating the one or more clusters from the clustering algorithm;

determining a single cluster as a weighted average of the one or more clusters, wherein the determined single cluster is used to generate the future test option based at least in part on the plurality of textual responses and the determined single cluster; and

wirelessly providing the future test option by the server computer to the one or more instructor devices.

11. The one or more non-transitory machine-readable medium of claim 10, wherein the method further comprises: after generating the one or more clusters from the clustering algorithm, determining a distance metric associated with the one or more clusters to identify similarities between the one or more clusters, wherein the one or more clusters include at least one incorrect answer to the free response item.

12. The one or more non-transitory machine-readable medium of claim 11, wherein the distance metric analyzes a number of the plurality of textual responses not found in the same cluster.

13. The one or more non-transitory machine-readable medium of claim 11, wherein the distance metric is associated with a number between zero and $\log(a \text{ number of the one or more clusters})$.

14. The one or more non-transitory machine-readable medium of claim 11, wherein the method further comprises: ranking the one or more clusters based in part on the distance metric; and

transmitting the ranked clusters to the one or more instructor devices.

15. The one or more non-transitory machine-readable medium of claim 10, wherein the plurality of textual responses are translated to the numerical dataset by:

stemming one or more words in the plurality of textual responses; 35

computing 1, 2, 3, 4, and 5-grams;

removing stemmed n-grams that appear in fewer than 1-percent or more than 99-percent of the plurality of textual responses; and

generating the stemmed n-grams as a vector.

16. A computer system for determining a plurality of textual responses to generate a future test option to correspond with a response to a free response item, the system comprising:

a processor; and

one or memories coupled with said one or more processors, wherein the one or more processors and one or more memories are configured to;

determine, at a server computer, one or more participant devices and one or more instructor devices, wherein the one or more participant devices and the one or more instructor devices are geographically dispersed and communicate via a network, wherein each of the one or more participant devices is operated by a respective participant and each respective participant is associated with a respective characteristic;

receive, at the server computer, the free response item from one of the one or more instructor devices;

13

receive, at the server computer, the plurality of textual responses from the one or more participant devices;
 remove punctuation from the plurality of textual responses;
 translate into a numerical dataset the plurality of textual responses resulting from the remove punctuation step;
 apply, at the server computer, a clustering algorithm to the numerical dataset to generate one or more clusters;
 generate the one or more clusters from the clustering algorithm;
 determine a single cluster as a weighted average of the one or more clusters, wherein the determined single cluster is used to generate the future test option based at least in part on the plurality of textual responses and the determined single cluster; and
 wirelessly provide the future test option by the server computer to the one or more instructor devices.

14

17. The computer system of claim 16, wherein the one or more memories are further configured to:
 after generating the one or more clusters from the clustering algorithm, determining a distance metric associated with the one or more clusters to identify similarities between the one or more clusters, wherein the one or more clusters include at least one incorrect answer to the free response item.
 18. The computer system of claim 17, wherein the distance metric analyzes a number of the plurality of textual responses not found in the same cluster.
 19. The computer system of claim 17, wherein the distance metric is associated with a number between zero and log (a number of the one or more clusters).
 20. The computer system of claim 17, wherein the one or more memories are further configured to:
 ranking the one or more clusters based in part on the distance metric; and
 transmitting the ranked clusters to the one or more instructor devices.

* * * * *