Big Data with small cases: A method for discovering students centered contexts for Physics courses

M. S. Bülbül
msahinbulbul@gmail.com

Department of Primary Education, Kafkas University, Turkey

Abstract. This article proposes a methodology that could assist teachers in understanding their students' primary needs or interests to decide on the kind of examples or contexts to be used in the classroom. The methodology was tested on 100 volunteers from university (N=50) and high school (N=50) in Ankara, Turkey. The participants were asked to write down the first word they thought of when they were presented with a single letter from the Turkish alphabet, which contains 29 letters. Then all the collected words (29x100=2900) were analyzed with the online word cloud creator, Wordle. According to results, the most cited words from the high-school classes were similar to each other, while the data from university participants showed more diversity. The most chosen word by the participants may give some clues in relation to the context that the teacher can utilize in planning a course. This study shows how to use a big-data-visualization-tool-based methodology to analyze the data gleaned from the participants' life-long experiences.

Keywords: Big Data, Physics courses, word cloud

Introduction

Learning occurs more effectively when students are interested in and attracted to the target subject matter (Blumenfeld et al., 1991). Therefore, it is important that educators utilize real-life examples that have a direct connection with the students' lives. When teaching physics, teachers regularly face questions from the students about the practical use of the knowledge that is being imparted. For instance, during the teaching of Ohm's law, students can react negatively (Bennett, Hogarth & Lubben, 2003) asking questions such as "outside the physics laboratory, when and where will we use this knowledge?" To avoid this situation, educators and curriculum designers have created new ways such as inquiry-based learning (Edelson, Gordin & Pea, 1999) or integrating student questions (Alimisis, 2013) in which to present the content. One of these methods is context-based education, which provides students with an environment to learn and retain the necessary knowledge by connecting topics to real life and transferring them from one context to another (Gilbert, 2006). Therefore, most of the developed countries that were at the top of educational rankings adapted their science education systems according to context-based education, such as the Salters Twenty First Century Science GCSE in the UK (Bennett & Lubben, 2006) and the ChiK (chemistry in context) project in Germany (Parchmann et al., 2006).

Context-based education is an approach, which influences all parts of the education system. For instance, PISA is an international assessment organization, which transformed the assessment of the scientific skills procedure through a context-based structure. Moreover, this approach is not only useful for face-to-face education but also distance learning utilizing context-based videos. Choi and Johnson (2005) commented that context-based movies are more memorable than the traditional text-based instruction. Context-based education is one of the important dimensions of the current research into science education. However, there is a need to determine how to decide on the necessary contexts for science learners, for example, about the country they live in, the school they attend and their age. The aim of this article is to present a method that can be implemented with science students to assist teachers and science curriculum designers in the choice of a context for instruction.

Thinking (Philosophy), Classical Science and Big Data Science

Science began with individuals or small groups of human exploring what existed in nature, for instance gaining experience of fire and developing an understanding of the phenomenon. Classical science answered questions by designing experiments and using tools to collect data. Initially, this data was only shared with a few, and access was limited to scholars. In today's scientific world, a vast amount of articles are published, and through the advent of technology, the collected data can be accessed not only by scholars and scientists but also by the public in general. Figure 1 offers a representation of the interpretation of the phenomenon of fire by taking the big data path or the classical science route to investigate thousands of data sources concerning the fire. However, in the end, for both big data science and classical science, there is a need to think about the conclusions, and this requires the most complex of tools, the human brain.

Although the history of science, the approach from a philosophical aspect, and linear scientific methods to complex and non-linear science can be analyzed from the perspectives of technology, science has always depended on the brain, which has its limitations and continues to be evaluated. Despite its limitations, the brain is not a simple organ; it stores all the collected data using a complex procedure of connecting neurons. In the human brain storage system, all the focused and non-focused data collected by the senses is analyzed across the whole network when required. This is achieved by activating the neurons.

Searching words or web

In the human brain, all words are classified as semantic networks, and learning new words strengthens or weakens other structures, this dynamic network influences speech, behavior and a person's choice of interest (Arbesman, Strogatz & Vitevitch, 2010; Borge-Holthoefer, Arenas, 2010; Chan & Vitevitch, 2010). In the literature, Griffiths, Steyvers and Firl (2007) used the association of words in the human mind to compare hyperlinks in web pages. They proposed that in both human retrieval from memory and an internet search involve finding the items associated with that search.

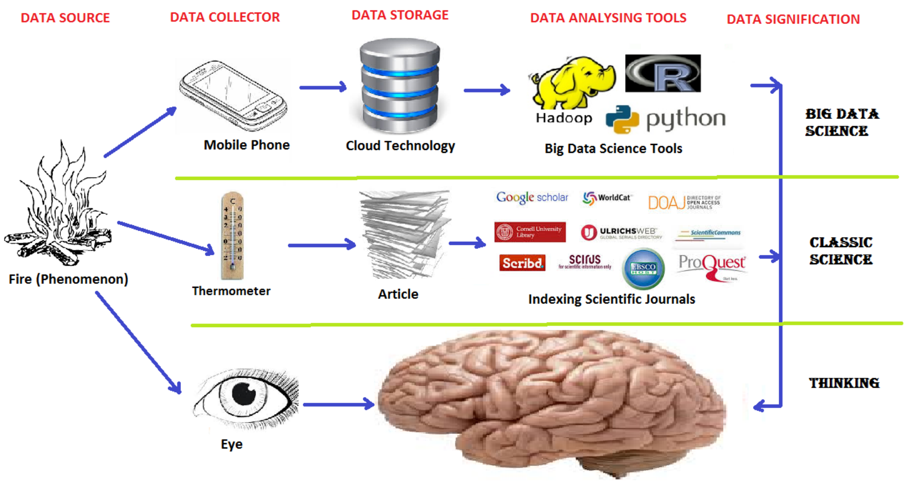

Figure 1. Comparative history and current status of science

Assume that you are listening a story, interaction of the words in the story and those in your memory create a deeper understanding by helping you visualize the content. The same idea can be applied to the tools used in big data science that help researchers visualize the collected data. Thus, to investigate the data from different storage units in the human brain, researchers and educators need tools and procedures. In the following sections the tool, word clouds, and the procedure, the First Letter Method, are presented.

Using word clouds in education

The content of big data differs according to how often the data is collected, whether it is structured, and the size of the stored data (Sagiroglu & Sinanc, 2013). The human brain gathers data from birth, and by making connections, different structures/networks are constructed by the individual, thus, when a concept question is asked to someone, her answer will be built from all meaningful structures that they have acquired through their life to date. From this perspective, thus, the teacher facing a class has to take into account the enormous amount of collected data held by the students.

Big data analysis has been used across a broad range of activities, from monitoring individual student performance to testing and evaluating curricula (Picciano, 2012). In education systems, big data analyzing methods let researchers work on real-time data and with real users' characteristics (West, 2012). It is important to use visualization to reveal the whole picture of big data, (Dadzie & Rowe, 2011). One method is to use a word cloud application. For example, Wordle is web page-based and allows the user to input data and produce word clouds in which repeated words are presented in increasing font size depending on the frequency of occurrence in the data. Through this visualization, the researcher/user can easily see the main elements of input data (Berlanga et al., 2012). Wordle has been used in various educational research studies (McNaught & Lam, 2010), one particular instance was the analysis of elementary pre-service teachers' views about science education (Asim, 2015).

An example of a word cloud from the current study in Figure 2 was created from the 10th-grade students' contexts lists about the Turkish high school physics curriculum. As it seen from the word cloud, there is a homogeneity of word distribution. This may be correct for contexts list, but, for the method proposed in this paper, we need some bigger and smaller letters to decide appropriate contexts for the researching cases.

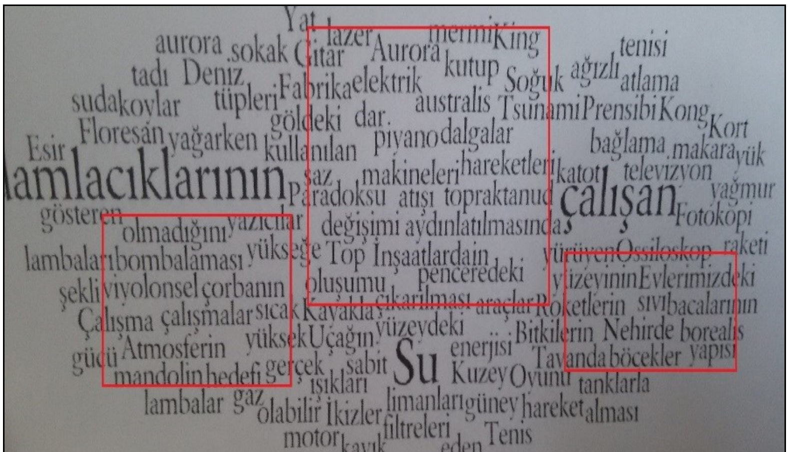

Figure 2. An example of a word cloud

The proposed method

There are different big data analyzing methods one of which is the Hashing method, which transforms collected data into shorter or indexed values (Chen, Mao & Liu, 2014). Another way of sorting data is to use the First Letter Method (FLM). Essentially, this involves presenting a person with a letter of the alphabet and asking them to speak or write a word starting with the target letter. According to this method, it is assumed that every relevant context and/or concept can be written in a word and that word starts with a letter. The word that the person offers is selected from thousands of words. It is also assumed that the chosen word usually reflects the structure of the human brain. If this were not true, then every person would write the same word in response to the letter and a word cloud created in this situation would present a homogeneity of word distribution. The range of words that the participants provided in this study confirms that different brains will write different words in response to the same letter of the Turkish alphabet.

Usually, mathematicians use axioms to refer to the self-evident facts. In order to understand the correlations between big data science and the theoretical framework of FLM, three axioms of FLM are given below.

1st axiom: All data (for example; concepts, facts, laws, and objects) has a name.

2nd axiom: Every name starts with a letter.

3rd axiom: Human brain is more engaged with selected names than unselected names when responding to the letter task (The first occurred concept when it was asked reflects the distribution of concept network's structure).

Possible problems in the use of FLM

When using FLM, there are two potential problems based on the use of Wordle and human brain functions. Wordle may label certain words such as 'the', 'on' and 'of' in English as meaningless words and this can appear in the output image. This occurs because Wordle works mechanically constructing the output image on the frequency of words repeated more than once (McNaught & Lam, 2010). However, it is possible with Wordle for the user to eliminate certain words by determining a threshold frequency value that will eliminate words under a given value. The other potential problem is related to brain functions. Commonly, FLM requires the participant to present the word they first considered referring to the most cited network; however, there may be more than one available networks. Therefore, during the process of deciding on the contexts in which to teach a particular topic, teachers may need to elicit more alternatives from their students. The number of letters is the solution to this problem. As assumed above, each word given by the participants reflects certain parts of their neural network. Each selection contributes to the big picture of that participant's mind map; more letters presented, then, indicates that he mind map of that participant is more detailed.

Method

This research was conducted in a similar way to a survey; however, the purpose of a survey is to present the big picture of investigated cases and the current study aimed to demonstrate the usability of FLM. To achieve this, a data-collecting tool was created, and Wordle was used to present graphically the data to be analyzed.

The aim of the study

Traditionally, it is the teacher who decides the context in which the subject matter will be taught. It is considered that students do not have adequate knowledge of and experience in the subjects taught in school and university. However, for proper learning to take place, the students need to be interested in the subject matter, and this can be achieved by involving the students in the contexts in which the teaching takes place. FLM allows teachers to gain insight into students thinking and their attitudes towards the curriculum.

The sample

The participants were selected from a university and a high school. They were chosen because they were easiest to recruit and we did not consider picking representatives from a range of universities or high schools. From the university, all the participants who volunteered (N=50) were pre-service physics teachers. The participants from the high school consisted of two 10-grade classes from the science group, which were taking physics courses and were following context-based physics curriculum. In each of the classes, labeled A and B, there were 25 participants.

Timing the study

We did not give attention to timing; we read the written directions to participants, but collected the participants' papers after they finished the task. For instance, one week after giving the task, I went to the high school to gather all the response sheets from their teacher. In the university, the participants brought their reports to my office individually. Entering the collected data took approximately three hours.

The budget of the study

We believe that FLM is the cheapest method to determine appropriate contexts for small cases or country-wide curriculum studies. The data-collecting tool consists of a form that can be printed two to a page. The second stage only requires an internet connection already available in the University. If the students immediately wrote the words on a tablet or mobile phone and sent them by email, then the study expenditure would be zero; however, collecting the data during the course hours does not seem to be efficient.

Data-gathering tool

This tool consists of two parts: directions for the participants and sheets of paper containing boxes in which the participants write the word they associate with the given letter. In the direction part, we explained the purpose of the study and how the participants are to complete the boxes. We did not require any personal information such as gender, name or socioeconomic status. Also, no information was recorded concerning the academic achievement of the participants.

Data analyzing method

The data was analyzed using the same procedure as given in the Wordle website. The data is entered into a box on the website, and then the program transforms the entries into words using different size fonts according to the frequency of occurrence. The output images from different classes or cases can be compared by assigning numbers to specific words.

Results

Four different output images from participants are presented in this section. Various words are labeled with numbers, and the differences are discussed together with a guide to the use FLM. In Figure 3, we put two outputs from 10th-grade students. The red numbers denote the output from class A, and the green numbers present the output from class B. These two outputs are approximately the same. Figure 4 shows the combined outputs of class A and B. The bottom image in Figure 4 numbered in orange consists of mixed data from all the 10th-grade participants and the numbers highlighted in yellow are the outputs from the university participants. In contrast to the data given in Figure 2, there are two different outputs in Figure 4.

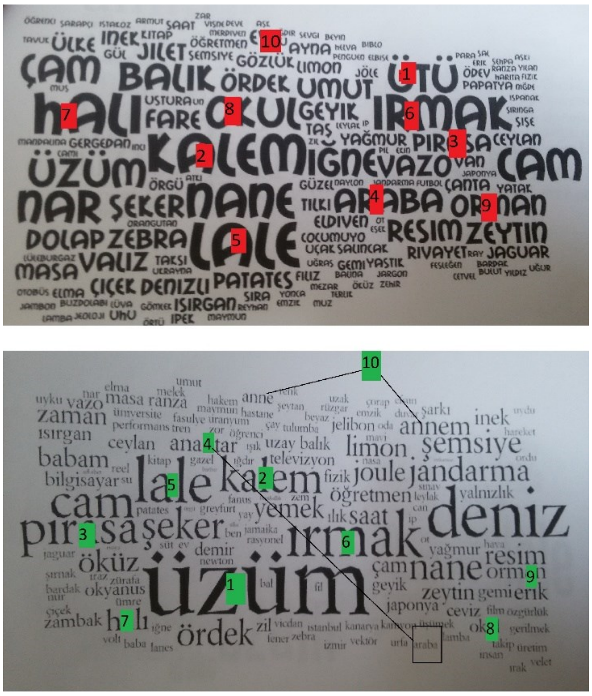

Figure 3. Output images of high school classes

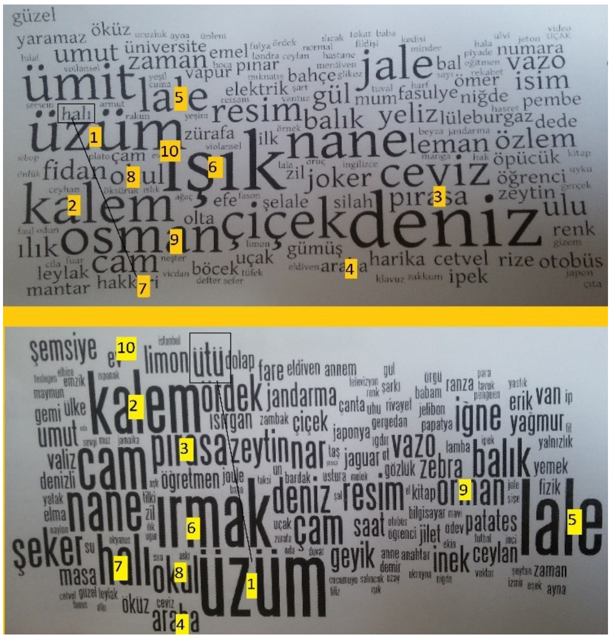

Figure 4. Output images of high school and university participants

The following ten descriptions refer to the four output images given in Figures 3 and 4.

When participants were asked to give a word starting with Ü, the Turkish word 'üzüm' (grape) was the most common response in Figures 3 and 4 from both the high school and university students. Another word for the letter 'Ü' was 'ütü' (iron). The occurrence of this word was more frequent in the responses of the high school students than any other. This difference may be due to high-school students being forced to wear neat, tidy uniforms. In the output of university students, there were more given names such as 'Ümit' as shown in Figure 4. Interestingly, no given names appear in Figure 3; the responses are all objects or concepts.

In Figure 3, although the frequency of the word 'kalem' (pencil) shown in the size of the font changed from group to group, both school and university participants gave the word 'kalem' when asked for a word starting with 'k.' Since all participants were using a writing implement to write the words, this frequency supports the idea that participants associated their response with a near object. Thus, the suggestion is that the Turkish physics curriculum topic of floating bodies according to FLM teachers should engage in a discussion of floating pencils.

Responding to the letter 'p,' both groups of students gave the word 'pırasa' (leek). In 10th-grade physics curriculum, there is a suggestion for teachers in head and temperature unit, soup. If a teacher uses FLM and the students give the word 'pırasa,' then which kind of soup do you discuss during the course? Leek soup, of course, since it is familiar to all.

For the high school class A, for the letter A, the word 'araba' (car) was the more frequent; however, for the high school class B, the word 'anahtar' (key) occurred more often than others. Thus, for class A, the teacher should make cars as the context for a unit on motion and friction, according to the FLM results. If the physics teacher of class B needs a context about moving vehicles, FLM suggests watercraft because this class wrote 'deniz' (sea) for the letter 'd.'

In Turkish, 'lale' is very popular in both cases. This is a kind of flower and means 'tulips.' If we asked them to say a flower, probably they would say 'rose'; however, according to this FLM application, both cases know 'tulips' with high frequency. In Turkish culture, beyond being a flower, tulips have artistic, historical and religious meanings. This widely connected network of Turkish participants produced 'tulips.'

When presented with the Turkish letter 'ı', many of the high school participants wrote 'ırmak' (river); however, many of the university participants wrote the female name 'Işık.' This result suggests that teachers could use a 'river' context for these high school students and 'sea' for the university case. For the high school students, this result contrasts with the suggestion in the 10th-grade science curriculum that waves could be discussed using examples in lakes or the sea. From this FLM application, we suggest that teachers use waves on rivers as a context.

As can be seen in Figure 4, university participants wrote 'Hakkari,' which is a city of Turkey. This is probably because, when spelling a word phonetically in Turkish, people mostly use city names (R – Rize), but participants of the high-school wrote 'halı' which means 'carpet.' University students do not regularly use carpets in dormitories. Moreover, if you write any letter to Google, you usually see the names of cities (It is valid for ordinary users not for using 'Search Engine Optimization').

For the letter 'O', the high school participants, especially class A, wrote 'okul' (school) but university participants wrote 'Osman' (a male name). In high school years, the effect of your family continues but in university, you are probably away from home and need more social skills. In this perspective, there is a consistency of results regarding being a high school or university participant.

In Figure 3, the size of the font for the word 'orman' (forest) is smaller than 'okul'. The high school students wrote 'okul' and 'orman'; however, the former school is in a large font size than the latter. This could imply that school is a more relevant context than forest for those students.

Similar to the previous result, home (ev) seems to have a smaller presence than school (okul) in high school students' mind, but this word is in a bigger context in high-school students' mind compared to university students' ones. In other words, among university participants' responses, it is hard to see the word 'ev' (home).

Conclusion

A few important points that teachers should take into account when using FLM to construct a context for their course should be mentioned.

Written words from different cases may not be the same, and these words may change from school to university or/and from class to class in the same school. From early word learning (Hills et al., 2010) to adult education at University (Jõgi, Karu & Krabi, 2015), context-based education is of critical importance to make subject matter understandable, but teachers have some difficulties to find contexts (Walan, Ewen & Gericke, 2015). FLM may help teachers to find contexts for different classes or diversity learning environments. We understand that FLM creates a small gateway into the learners' world/mind that allows researchers and teachers to understand student's needs and learning styles. Teachers can use FLM before starting a new issue or textbook authors may use these word clouds. This study suggests that FLM allows users to read output image from Wordle produced by big data visualization.

Although there are different types of educational approaches such as inquiry-based learning, we focused on context-based learning to explain how a big data source can be visualized. FLM can be used with different methods, but this article only gives one example. From classroom or school-wide cases, we generated a practical method that is easy to use, costs little and takes minimal time. The brain records what is happening around it, and FLM gives a big picture of all the participants' brain records. From this perspective, it is suggested that FLM, a big data source analyzing procedure, may work with the brain.

FLM can help teachers decide on contexts correlated with participants in particular classes or across the whole school. Additionally for textbook writers, FLM can be used to guide and modify new books. Furthermore, writers can give a description of the use of FLM in the teacher's book to encourage teachers to use this approach). Finally, curriculum developers can use FLM to determine contexts to recommend to teachers and give a description of FLM in the teacher's handbook.

It is possible to reconstruct this study in other languages and other groups of participants. In addition, studies about students with low and high achievement will give us highly valuable information. Examining the potential difference in the responses from male and female students may also provide new insights.

References

Alimisis, D. (2013). Educational robotics: Open questions and new challenges. *Themes in Science and Technology Education*, 6(1), 63-71.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, 20(03), 679-685.

Asim, S. (2015). Using Wordles as the Visual Data to Explore Prospective Elementary Teachers' Conceptions of Informal Science Education. In D. Slykhuis & G. Marks (eds.), *Proceedings of Society for Information Technology & Teacher Education International Conference 2015* (pp. 2779-2783). Chesapeake, VA: Association for the Advancement of Computing in Education (AACE).

Bennett, J., Hogarth, S., & Lubben, F. (2003). *A systematic review of the effects of context-based and Science-Technology-Society (STS) approaches in the teaching of secondary science*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Bennett, J., & Lubben, F. (2006). Context-based chemistry: The Salters approach. *International Journal of Science Education*, 28(9), 999-1015.

Berlanga, A. J., Van Rosmalen, P., Boshuizen, H. P., & Sloep, P. B. (2012). Exploring formative feedback on textual assignments with the help of automatically created visual representations. *Journal of Computer Assisted Learning*, 28(2), 146-160.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist*, 26(3-4), 369-398.

Borge-Holthoefer, J., & Arenas, A. (2010). Semantic networks: structure and dynamics. *Entropy*, 12(5), 1264-1302.

Chan, K. Y., & Vitevitch, M. S. (2010). Network structure influences speech production. *Cognitive Science*, 34(4), 685-697.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, 19(2), 171-209.

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *The American Journal of Distance Education*, 19(4), 215-227.

Dadzie, A. S., & Rowe, M. (2011). Approaches to visualizing linked data: A survey. *Semantic Web*, 2(2), 89-124.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*, 8(3-4), 391-450.

Gilbert, J. K. (2006). On the nature of 'context' in chemical education. *International Journal of Science Education*, 28(9), 957-976.

Griffiths, T. L., Steyvers, M., & Firl, A. (2007). Google and the mind predicting fluency with PageRank. *Psychological Science*, 18(12), 1069-1076.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63(3), 259-273.

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The Qualitative Report*, 15(3), 630-643.

Parchmann, I., Gräsel, C., Baer, A., Nentwig, P., Demuth, R., & Ralle, B. (2006). 'Chemie im Kontext': A symbiotic implementation of a context-based teaching and learning approach. *International Journal of Science Education*, 28(9), 1041-1062.

Picciano, A. G. (2012). The evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. In G. Fox & W. W. Smari (eds.), *International Conference on Collaboration Technologies and Systems (CTS)* (pp. 42-47). USA: IEEE.

Walan, S., Mc Ewen, B., & Gericke, N. (2015). Enhancing primary science: an exploration of teachers' own ideas of solutions to challenges in inquiry-and context-based teaching. *Education*, 3-13, 1-12.

West, D. M. (2012). *Big data for education: Data mining, data analytics, and web dashboards*. Governance Studies, The Brookings Institution, Washington. Retrieved December 10, 2015 from http://www.brookings.edu/research/papers/2012/09/04-education-technology-west.

To cite this article: Bülbül, M. S. (2015). Big Data with small cases: A method for discovering students centered contexts for physics courses. *Themes in Science and Technology Education*, 8(2), 105-113.

URL: http://earthlab.uoi.gr/theste