Vaal University of Technology

A conceptual framework for crowdsourcing in higher education

Submitted for the degree Magister Technologiae in Information Technology

In the Department of Information and Communications Technology,

Faculty of Applied and Computer Sciences

Thulani W. Shongwe

Student Number: 211019496

Supervisor: Professor Tranos Zuva

Co Supervisor: Mrs Rene van Eck

April 2018

Declaration by Candidate

"I hereby declare that the dissertation submitted for the degree M-Tech Information Technology, at Vaal University of Technology, is my original work and has not previously been submitted to any other institution of higher education. I further declare that all sources cited or quoted are indicated and acknowledged by means of a comprehensive list of references".

ThulaniShongwe

Abstract

The speedy growth of Internet based information and communication tools produced a new field of prospects for educational organizations to reach their aims. One of the options is crowdsourcing. Crowdsourcing was recently the answer to the growth for providing different applications in areas such as education, financing, and entrepreneurship. South African schools are considerably failing in education. A big challenge is when it comes to the mathematics delivery method which ends up affecting the learners' performance. When compared to other middle income nations, South Africa is ranked third from the bottom in terms of its performance when it comes to mathematics. This study designed a conceptual crowdsourcing tutoring framework. The framework defines the use of how crowdsourcing can contribute to tutoring grade 11 and 12 mathematics in order to improve the learners' performance. A prototype was developed to illustrate the crowdsourcing tutoring framework. The *simpleKmeans* algorithm was used in the prototype. The algorithm was used to select learners, tutors and appropriate textbooks for the virtual class. The prototype system proved to be effective as it was able to cluster students according to their performance and tutors according to their student pass rate.Through the usage of a clustering *simpleKmeans* algorithm, this study was able to create a virtual class that illustrated how all the components come together for the proposed crowdsourcing tutoring virtual class.The use of the prototype system was able to fill the virtual class with students who obtained low average marks and educators with the pupils who had the highest pass rate. This study was able to build a virtual class with the following components: learners, tutors and textbooks. Objectives and research questions of this study were fulfilled. In future studies the researcher will endeavor to make the system recommend textbooks without using the textbooks used by the teachers who produced the best results.

Keywords: **Crowdsourcing, Education, E-Learning**

Acknowledgements

I would to take this opportunity to send my utmost thanks to my mentor and supervisor Professor Tranos Zuva who has provided me with the support and guidance needed to complete this dissertation. I really appreciate his time and I am grateful for this amazing learning experience.

I would also like to thank the Vaal University ICT department for providing me with this amazing academic learning experience. I have realized that without the support of this University, my research could not have been successful and thanks again to you Vaal University of Technology for the financial support and for giving me the opportunity to further my studies.

I would like to thank my family and friends for their continuous support. Thank you for providing this ongoing support when I needed it the most. In addition, I would like to thank my fellow students who were there for me when I needed their contribution. They provided a favorable environment for me to work in.

I would also like to thank my daughter Phepsile Shongwe for her support and I wish she can be rewarded for joy she sacrificed during the period of my study. Not forgetting my late mum (Anna Nontombikayise Dludlu) who always encouraged me to put education first.

List of Figures

Figure Number	Figure Name	Page
1	Trends in International Mathematics Science Study	5
2	Current crowd work processes	13
3	Framework for assessing crowd source-ability of a task	15
4	Framework proposed by Kittur et al. (2013) for future crowd work processes to support complex and interdependent work	17
5	Taxonomy in crowdsourcing	20
6	The application of crowdsourcing in software engineering	22
7	A crowd-learn concept	34
8	Innovation project outcomes from crowdsourcing	36
9	Learning outcomes using crowdsourcing as a problem-solving educational tool	37
10	The concept of crowdsourcing for problem-solving	40
11	Proposed crowdsourcing tutoring framework	50

12	SimpleKmeans clustering process	55
13	Picture of students accessing our system via gadgets	59
14	Student registration form	60
15	Multiple choice sample questions to be used for re-evaluation	62
16	Comparison of performance time for SimpleKmeans and Hierarchal algorithms	63
17	Results for classification of students with high average marks and low average marks from the different topics using the weka tool	68
18	Results for the classification of tutors with best results from the different topics using the weka tool	74
19	Picture displaying how the virtual class comes into action once the students have been recommended to one virtual class	78
20	Results classified showing all the components put together	81

Declaration by Candidate……………………………………………………………………i

Abstract……………………………………………………………………………………ii

Acknowledgements……………………………………………………………………......iv

List of Figures………………………………………………………………………… v

Table of Contents……………………………………………………………...……vii

Chapter 1 1

1.1 Introduction...1

1.2 Problem statement..5

1.3 Research questions..6

1.4 Aim and objectives 7

1.5 Research contribution 7

1.6 Research ethics...7

1.7 Dissertation layout 7

1.8 Chapter summary……………………………………………………………………10

Chapter 2: Literature Review...12

2.1 Introduction 12

2.2 Benefits of crowdsourcing..23

2.3 Challenges of crowdsourcing 27

2.4 Crowdsourcing in education..30

2.5 Some of the frameworks used in crowdsourcing...41

2.6 Chapter summary………………………………………………………………….47

Chapter 3: Research Methodology..48

3.1 Introduction..48

3.2 Sample of the study……………………………………………………………………49

3.3 Proposed crowdsourcing tutoring framework…………………………………………49

3.4 Student selection………………………………………………………………………50

3.5 Tutor selection…………………………………………………………………….....51

3.6 Book selection……………………………………………………………………........51

3.7 Virtual class composition…………………………………………………………...51

3.8 Selection of clustering algorithm……………………………………………….......51

3.9 SimpleKmeans algorithm…………………………………………………………...54

3.10 Chapter summary………………………………………………………………….......56

Chapter4: Experiments and Results…………………………………………………..58

4.1 Introduction……………………………………………………………………………58

4.2 Students and access to prototype system………………………………………………58

4.3 Experiment for algorithm selection……………………………………………………63

4.4 Experiment for student selection………………………………………………………64

4.5 Experiment for tutor selection…………………………………………………………71

4.6 Book selection…………………………………………………………………………77

4.7 Virtual class……………………………………………………………………………78

4.8 Crowdsourcing prototype………………………………………………………….....82

4.9 Chapter summary………………………………………………………………………83

Chapter 5: Conclusion and Future Work……………………………………………………..86

5.1 Introduction……………………………………..………………………………86

5.2 Conclusion……………………………………………………….……………..86

5.3 Future work……………………………………………………………………….88

5.4 Chapter summary…………………………………………………………………89

References………………………………………………………………………...91

Chapter 1

1.1 Introduction

This research is intended towards a mission of improving mathematics performance among learners in grade 11 and 12 by designing a proposed crowdsourcing tutoring framework that has been developed to improve the pass rate in mathematics for South African schools. The speedy growth of Internet based information and communication tools produced a new field of prospects for educational organizations to reach their aims. One of the options is crowdsourcing.

The word originated in a Wired Magazine article by Jeff in which he defined the evolving occurrence as subcontracting to a multitude of people and since then the word has progressed and been well-defined in a variety of ways (Keatinga and Furbergb, 2013). Educational organizations have many responsibilities that are crucial, but that entail time and energy that could be used to pay attention to tutoring; therefore crowdsourcing can be the answer (Skarzauskaite, 2012).

The application of crowdsourcing in the academic world can be traced back to January 1859, when the Philological Society in Great Britain started preparing a New English Dictionary whereby the society made an appeal to the English-speaking and English-reading public to read books and contribute extracts for the Philological Society's New English Dictionary (Gupta and Sharma, 2013).

Crowdsourcing was recently the answer to the growth of providing for different applications in areas such as education, financing, and entrepreneurship (Dhiya et al., 2015). Countless debates have been held for a long time in relation to the ability of a large network of people connected through ICT, termed as “crowd”, to perform successfully challenging proposals and problem-solving activities (Yannis et al., 2014). Connected tutoring systems have made substantial growth in recent years, and one can visualize them as "computational" representatives in a mixed human-computer social system (Weld et al., 2012).

There are explanations why education needs to be aligned with crowdsourcing, one being crowd-techniques will be necessary in order to deliver quality education in some areas, another being existing techniques are ready for application to this new area and lastly online education represents a new, relatively unexplored way of creating crowds (Weld et al., 2012). Crowdsourcing can in an effective way produce breakthrough innovations, but it has to be done in a particular way to guarantee great results (Kevin and Karim, 2013).

Though the term crowdsourcing has lately entered the Information Systems literature, the use of a combined number of intelligent people for solving business and academic problems has largely been the subject of attention throughout history (Amrollahi and Alireza, 2016). Crowdsourcing approach has been used for resolving numerous diverse problems up to now and 24 crowdsourcing platforms have been revised for a variety of applications such as: business, city planning, policy development, and event outreach (Amrollahi and Alireza, 2016).

The conceptualization of crowdsourcing introduced a framework that permitted addressing the players and their connections, the arrangement of assets, the method of mobilization of these assets and the purpose of mobilization (Asmolov, 2014). According to Svinicki (2010), why should one care about the conceptual frameworks that motivate research on teaching and learning? He suggests that one would not consider restructuring a model without understanding the essential principles that support and affect it in the first place. An appropriate platform to support the crowdsourcing activities in education is identified after the investigator has defined the target audience and planned an engagement instrument and this platform offers an opportunity for connecting and exchanging valuable ideas with participants (Keatinga and Furberg, 2013).

According to Skarzauskaite (2012), web-based distributed problem solving and manufacturing model for business is a fitting prototype for enabling educational organizations in their procedures. Crowdsourcing has appeared as a vital aspect of education on the web, for it promotes the sincerity and distribution of assets and information contributed by public.

The capability to evaluate and generate effective approaches to applying crowdsourcing and student empowering tools in the classroom is restricted by our willingness and capability to implement those (Hills, 2015). Industries are on the verge of a significant change in the way they innovate. According to Buecheler et al. (2010), during the past era, the Internet has supported societies to associate and work together, generating a virtual world of Combined Intelligence.

Since virtual relations are now likely and have developed outside of educational surroundings, administration of educational establishments must take into consideration labour force found in the crowd (Skarzauskaite, 2012). The purpose of this study was to propose a conceptual framework demonstrating how crowdsourcing content design in online tutoring represents principles of learning for which there is proof and to provide analysis of how this framework can be implemented in higher education focusing on the grade 11 and 12 mathematics syllabus.

The framework defines the use of how crowdsourcing can contribute through the use of a platform to lower the failure rate when it comes to higher education mathematics and the focus will be on grade 11 and 12 so as to tackle the problem earlier before they even get to their grade 12. Here the researcher has designed and developed a framework for effective crowdsourcing tutoring as there are many platforms already available, but still the problem of high failure rate in South African schools continues at an alarming rate especially in science and mathematics, but the focus here will be mainly on the grade 11 and 12 mathematics syllabus.

Instead of planning prototypes and techniques for determining workers' consistency, one must consider developing an intelligent framework that shows how tutors and students can work together in order for crowdsourcing tutoring to be more effective (Singla et al., 2013). There was consideration that there are many trials in developing such a framework, lots of them attributing to the underlying intricate nature of the learning practice of human beings (Singla et al., 2013).

Personnel in crowdsourcing are from various backgrounds, with different skill sets and preferences; concepts may be easy to absorb by one set of students and yet they could be very tough for others to pick up (Singla et al., 2013). These issues, which affect the tutoring capability of the tutors, are unfamiliar to the teaching system and with these trials, an effective conceptual framework for crowdsourcing can be aligned with developing a teaching system that can interactively instruct the tutors and familiarize the teaching as per the abilities, experience, and dynamics of each tutor.

According to Skarzauskaite (2012), many are smarter than a small number, no one differentiates everything, everybody knows something, all facts reside in humanity. Crowdsourcing the learner-mentor interaction delivers an essential human understanding and at the same time generates instructional sections that are inspected for quality by assessing division accomplishment with both the instant learner and upcoming learners (Anderson, 2011). According to Hills (2015), crowdsourcing content manufacturing is learner-centered by plan because it controls student knowledge to create an artefact that expands the base over which students can identify how what they are learning has relevance to their survival. Today's leading learning and motivation concepts are based on learners' thinking and understanding of the circumstances in which they find themselves and this conceptual framework for crowdsourcing will lead to a lot of changes in the way psychologists do investigation in education and in the ways that tutoring is designed (Svinicki, 2010).

1.2 Problem statement

South African schools are considerably failing in education, mostly when it comes to mathematics delivery methods that end up affecting the learners ‘performance (McCarthy and Oliphant, 2013). When it comes to mathematics teaching, it is not up to the expected standard and most teachers are not able to answer questions in the curriculum they are teaching (McCarthy and Oliphant, 2013). When compared to other middle income nations South Africa is ranked third from the bottom. Figure 1 provides information about how dire the situation is.

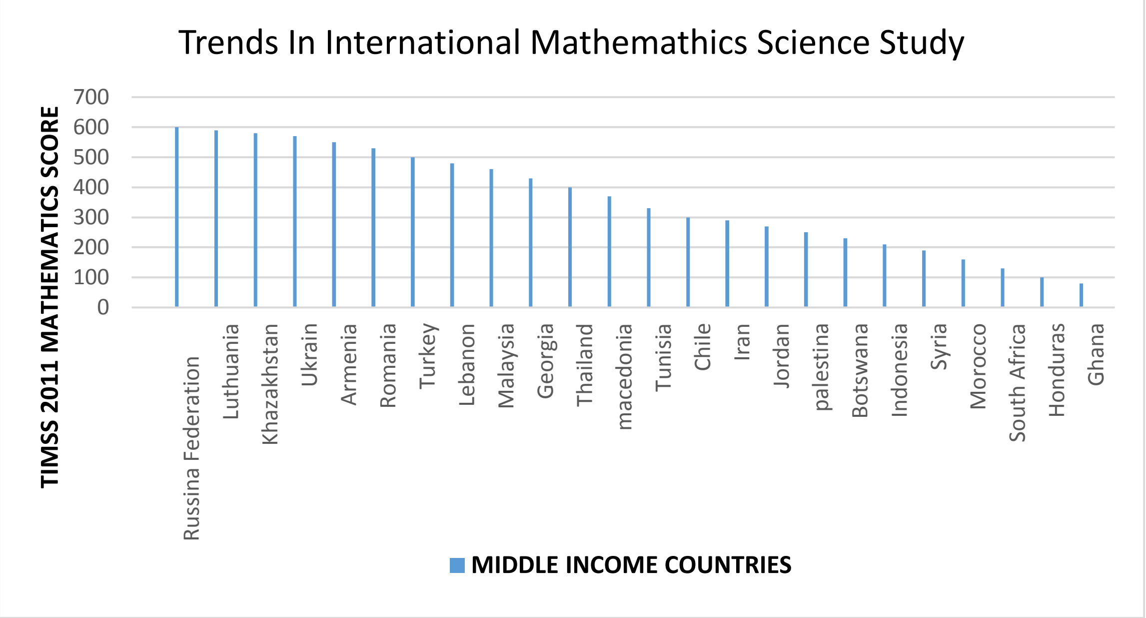

Figure 1 Report for CDE, South Africa’s Education Crisis The quality of education in South Africa 1994-2011(Spaull, 2013)

It is therefore concerning that the TIMSS (Trends in International Mathematics Science Study) established that the average mathematics performance of South African learners in Grade 9 is well below the global yardstick of 500 points (Visser et al., 2015). This study intended to improve mathematics performance of students or learners by designing a crowdsourcing tutoring framework that has been developed to improve the pass rate in mathematics for South African schools.

An application was developed to measure the effectiveness of the crowdsourcing framework for effective mathematics tutoring in South African Schools. In addition, it introduces people with expertise like retired educators and professionals who are not recognized in mathematics and yet they can contribute towards the improvement of pass rate for mathematics in South African schools.

1.3 Research questions

How can crowdsourcing be used to enhance the performance of mathematics pass rate for South African schools using an effective crowdsourcing tutoring framework?

In order to answer the research question the following sub-questions need to be answered:

- What has been done in literature to use crowdsourcing in education?
- How to develop a crowdsourcing tutoring framework for effective mathematics tutoring in South African Schools?
- How can the effectiveness of the proposed crowdsourcing framework be measured?

1.4 Aim and objectives

The aim is to design a crowdsourcing tutoring framework that will be used to improve the pass rate in mathematics for South African schools.

In order to fulfil this aim the following objectives need to be achieved:

- To study what has been done in literature to use crowdsourcing in education
- To propose a crowdsourcing tutoring framework for an effective mathematics tutoring system
- To develop an effective crowdsourcing mathematics tutoring prototype system for South African schools
- To measure the effectiveness of the developed prototype system

1.5 Research contribution

This study presents an effective crowdsourcing tutoring framework that will be used towards the improvement of the mathematics pass rate of learners in South African schools.

1.6 Research ethics

This research does not need ethics as the study will propose an effective crowdsourcing tutoring prototype system that will be used towards the improvement of the mathematics pass rate of learners in South African schools.

1.7 Dissertation layout

Chapter 1 – introduces crowdsourcing in general, crowdsourcing in education, and the motivation for this study. The problem statement for this study is set out clearly which is that South African schools are considerably failing in education, mostly in terms of mathematics teaching and learning. The main aim of this research is stated, which is to improve mathematics performance by designing a crowdsourcing tutoring framework that will be used to improve the pass rate in mathematics for South African schools. This main aim is answering our main question. Sub-questions and sub-objectives are stated to help guide this study towards achieving its goal. The research contribution and research ethics are also covered in this chapter. Lastly, the research layout is presented and a chapter summary is provided at the end of this chapter.

Chapter 2 – this chapter presents literature exploration that was regarded as important and an early step in responding to the research questions. The literature exploration involved finding and studying literature for information on crowdsourcing in general, crowdsourcing in education, benefits of crowdsourcing, challenges of crowdsourcing, and the existing crowdsourcing frameworks that have been used in any other fields to develop a solution. The literature exploration was also used to study various crowdsourcing tutoring frameworks that were specifically used in academic institutions to add value or contribute towards improvement of their results. Lastly, frameworks were compared to see how they were used in terms of their application, strength, and their weaknesses. A brief chapter summary is provided at the end of this chapter.

Chapter 3 – this chapter presents the research methodology whereby in responding to the study questions asked, this study implemented four methods, Literature exploration, modelling, prototyping, and evaluation. The proposed crowdsourcing tutoring framework has been developed and the framework is made up of components like students, tutors, virtual class and books with each having its own interface that is connected to a mini-system. This system will be running in the background of the framework that is controlled by middleware called crowdsourcing prototype. The way students and tutors are selected for the created virtual class is explained in this chapter. The way the books used for content creation are selected, is deliberated on in this chapter. Modelling involved developing an effective crowdsourcing mathematics tutoring prototype system for South Africa that was also used to measure the effectiveness of the prototype developed system.

The algorithm used in the implementation of this proposed framework is introduced and its selection explained giving reasons why this study opted for this kind of an algorithm. *SimpleKmeans* was the algorithm chosen to be used and the reason was that this algorithm is a clustering algorithm and works best with numeric data. A brief explanation of how *simpleKmeans* work was done it picks a number (K) of cluster centers that were centered randomly, assigns every item to its nearest cluster center using Euclidean distance, moves each cluster center to the mean of its assigned items, and repetition happens until convergence. The composition of the virtual class is explained in this chapter. A brief chapter summary is provided at the end of this chapter.

Chapter 4 – Students' access to the prototype system is explained in this chapter. Experiments are conducted for the selection of the algorithm to be used for experiments of students' selection and tutors' selection for the created virtual class. The results are presented where through the use of the weka tool, an algorithm within this tool is used to classify between good and bad performing students or learners in each and every topic of the mathematics syllabus of grade 11 and 12 in South African schools.

In this chapter again the same algorithm is used to classify between best tutors or educators in order for this study to be able to recruit the best tutor for our created virtual class. Books selection for the created virtual class is also done. Lastly, in this chapter in the created virtual class, there is an illustration showing how all the components (students, tutors and books) come together for the proposed crowdsourcing tutoring framework and the topic chosen for this experiment to meet the aim of this study. A comparison of what others researchers have done in relation to this study is also presented at the end of this chapter. A brief chapter summary is provided.

Chapter 5 –this chapter presents the conclusion where attention is given to the problem or question asked and tells if the main aim of this study has been achieved. All the objectives are discussed on how they contributed to answering some of the questions that were to be answered in relation to this study. Future work is also discussed, recommending that the books used to extract the content used by tutors must also be looked at to create a large pool from which more information can be extracted for future use. A brief chapter summary is provided.

1.8 Chapter summary

In summary, this chapter introduced the intent of this study. The mission or aim of this study was to improve students' performance in mathematics among students in grade 11 and 12, by designing and developing of a crowdsourcing tutoring framework to improve the mathematics pass rate of South African schools. The introduction of crowdsourcing in general, crowdsourcing in education, and the motivation for this study was discussed in this chapter.

The problem statement for this study was also discussed with an illustration of the trends in International Mathematics Science Study showing how dire the situation is in South Africa. It also stated clearly how South African schools were significantly failing in education, mostly in mathematics teaching and learning.

The main aim of this research was stated, which was to improve the mathematics performance pass rate by designing a crowdsourcing tutoring framework that will be used to improve the pass rate in mathematics for South African schools. The main aim answered our main question. Sub-questions and sub-objectives were discussed to assist in guiding this study towards achieving its goal. Research contribution and research ethics were also discussed in this chapter. Lastly the research layout is provided with a brief discussion of what to expect from each chapter.

Chapter 2: Literature Review

2.1 Introduction

Crowdsourcing can be used in several ways, whereby the sharing of information, knowledge, time and assets can be something that can be helpful for people in attaining positive goals with a number of tools that can help to use crowdsourcing online (Skaržauskaitė, 2012). Crowdsourcing is an idea used in generating other tasks for end users in an organizational process enabled by social IT; it can also control the work or the vision of a big group of individuals for a promoter using an open call to contribute via the Internet (Oliver et al., 2015).

According to Weld et al. (2012), crowdsourcing can provide the best educational understanding for it is a good framework for learning. Online tutoring systems have made significant advancement in recent years, and one can visualize them as a computational demonstration in a mixed human-computer social system (Weld et al., 2012). Stakeholder system actors from public society, business, and governmental organizations come together in order to discover a common method for an issue that affects everyone. This includes public-private collaboration as suggested in public-private social problem-solving alliances (Hyvärinen and Vos, 2015).

According to Asmolov (2014), conceptualization of crowdsourcing entails a meaning that differentiates it from other ICT presentations and addresses the challenges in every educational sector. The operational properties of crowdsourcing tools and deployments always connect communication to mobilization, which in turn makes crowdsourcing projects to be action-oriented tools that by classification are used to mobilize and involve Internet users with a variety of prospective audiences (Asmolov, 2014).

Complementing volunteer-based crowdsourcing, a paying crowd work business is now rapidly growing in scope and ambition. Crowd work nowadays extends to an extensive range of skillfulness and pay levels, with moneymaking vendors providing interaction to a range of workers and attentive support for various tasks (Kittur et al., 2013). Crowd effort has the prospective of backing a flexible workforce and alleviating encounters such as deficiencies of experts in specific areas or geographical locations (Kittur et al., 2013).

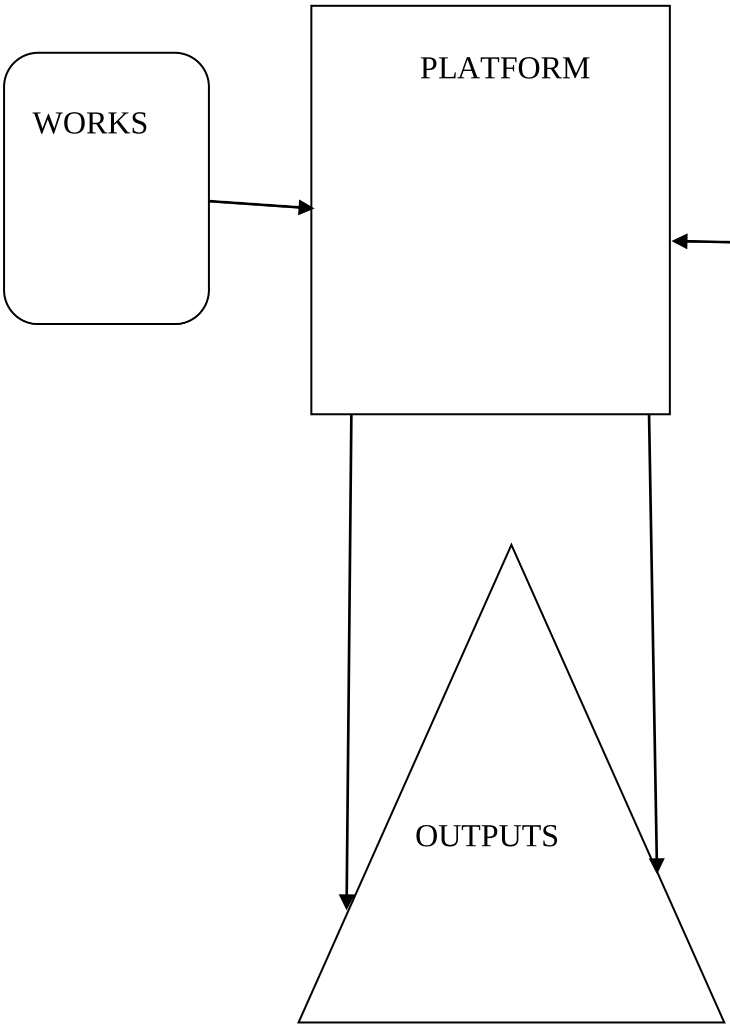

Figure 2: Current crowd work processes (Kittur et al., 2013)

According to Kittur et al. (2013), recent crowd work normally comprises of small, independent, and similar tasks, as shown in Figure 2. Workers in this figure were combined with an occurrence of each assignment to produce an output. Some techniques, established in the computer-supported cooperative work and crowdsourcing societies offer possible benefits to online education, and certain approaches have already been directly applied (Weld et al., 2012). Crowdsourcing content formation is learner-centered by intention because it stimulates student knowledge to build a database with information relevant to the students to improve their pass rate and to provide additional learning for topics relevant to their everyday circumstances(Hills, 2015).

To add to that Buecheler et al. (2010), state that crowdsourcing can be applied in a scientific method in order to make the most of the knowledge that can be added and spread. Two terms were used when developing a framework that can be used to apply crowdsourcing in a scientific method, collective intelligence and crowdsourcing (Buecheler et al., 2010).

It is demonstrated that crowdsourcing writing surveys should be possible with high accuracy and cost estimates that are reasonable compared with current costs (Krivosheev et al., 2017). Extraordinary calculations can be utilized to distinguish the parameters of the crowdsourcing task and is the best calculation recognized in view of a multi-run technique (Krivosheev et al., 2017).

Group applications likewise work for departmental applications where it is feasible for the explanation of the job to engage in or be valuable to the group overall and the space of crowdsourcing is without a doubt developing (Wang et al., 2013). An unmistakable pattern in the advancement of crowdsourcing is that the space of possible explanation stages is extending to incorporate more indications that permit specialists’ exchange of expenses in one measurement or another (Wang et al., 2013).

In Figure 3, there are variables that satisfy the connection among scientific productivity and collaboration in a scientific situation and it gives a schematic overview of all the relationships of the different elements of this framework.

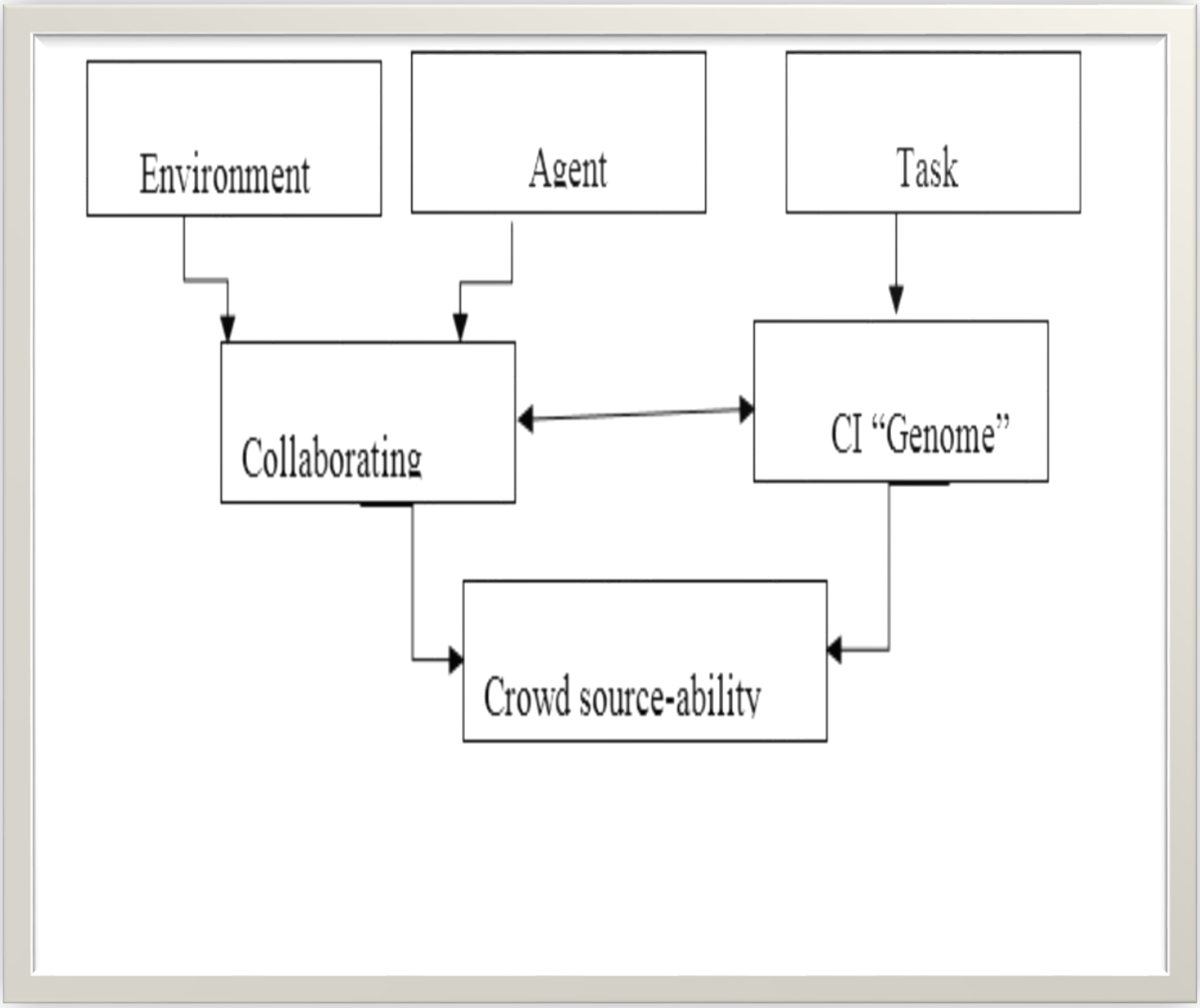

Figure 3: Framework for assessing Crowd source-ability of a task (Buecheler et al., 2010)

Saxton et al. (2010) state that by taking advantage of the rising recognition of social web technologies, business executives big or small are frequently producing and testing ground-breaking sourcing models. Crowdsourcing is one of the models whereby companies use the Web to connect the hard work of a virtual crowd to attain exact organizational tasks (Saxton et al., 2010).

Crowdsourcing is another way where data is being clarified, re-contextualized, confirmed, arranged, and shared through movement of the social media linked crowd (Starbird, 2012). Crowdsourcing efforts have continuously been credited equally with supporting important roles in response to hard work for numerous events (Starbird, 2012).

Crowdsourcing involves empowering a different group of individuals through the use of tools to contribute to a larger effort (Bott and Young, 2012). Encouragement to contribute must be stimulated to invite the best operative collaborators, and the crowd's incentive need to be in parallel with the long-term objective of the crowdsourcing resourcefulness to guarantee reliable, good quality participation (Bott and Young, 2012).

Merging views from human and computer group theories may thus deliver corresponding strengths and address similar weaknesses over using either of the two (Kittur et al., 2013). Crowdsourcing is another Web 2.0 based phenomenon; it turns into an apparent sourcing instrument for problem-solving in associations by outsourcing issues to an unclear constituent or to the crowd (Zhao and Zhu, 2014). Much like in its demonstration, it is considered that crowdsourcing research is a dynamic and lively research area as it has been consistently developing throughout the years (Zhao and Zhu, 2014).

In the crowdsourcing world, tasks are transferred to an organized group of people to perform with the aim that an organization's generation cost can be extremely reduced (Yuen et al., 2015). In crowdsourcing models, task planning can help workers to locate their correct assignments quicker as well as help requesters to get great quality returns faster (Yuen et al., 2015). One of the concerns for workers is that it is difficult for experts to discover suitable assignments to perform since there are a lot of unnecessary accomplishments out there (Yuen et al., 2015).

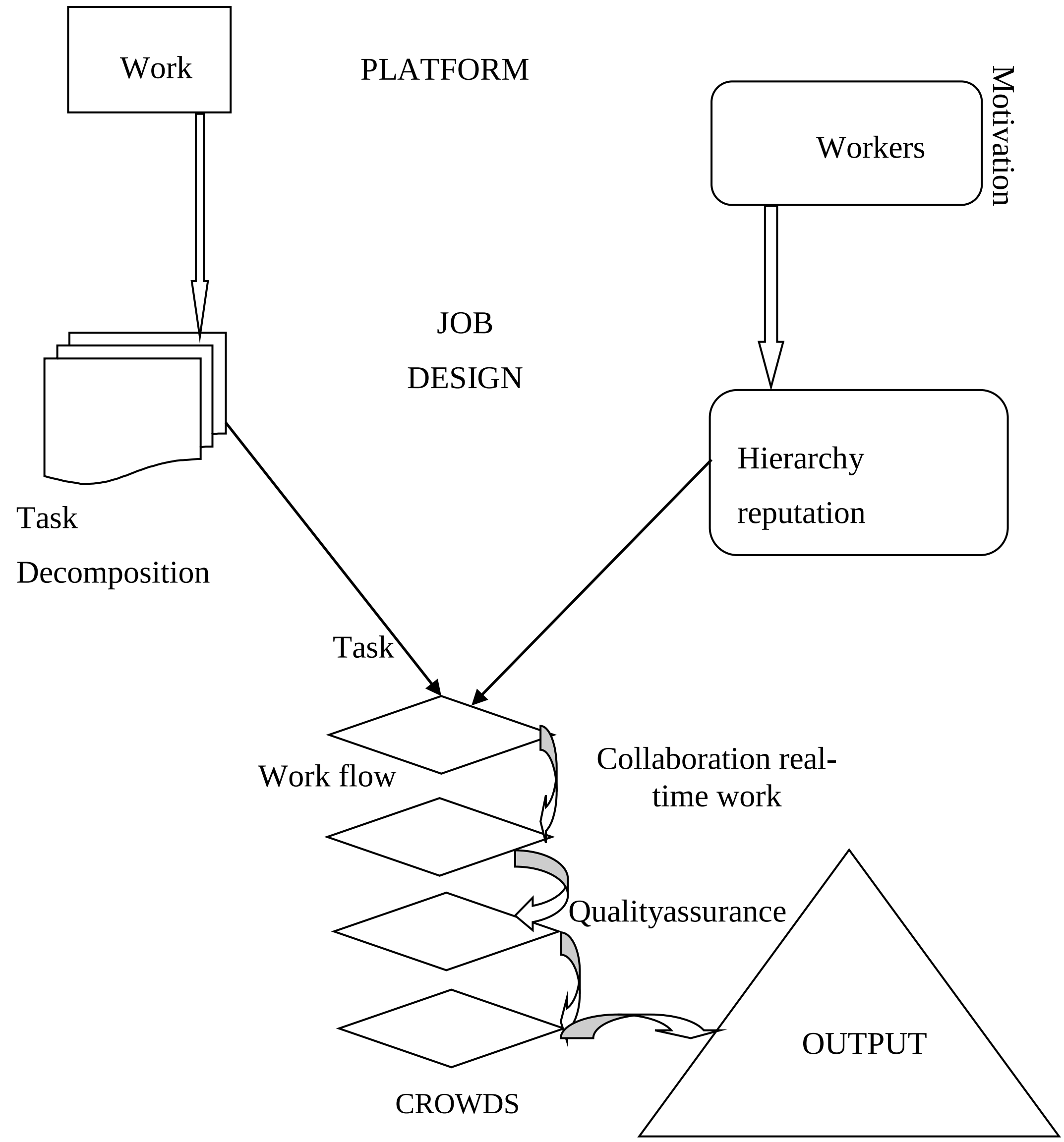

Figure 4: Framework proposed by Kittur et al. (2013) for future crowd work processes to support complex and interdependent work

The aim of this framework in Figure 4 is to visualize the prospect of crowd work that can sustain additional complex, innovative, and highly prized work (Kittur et al., 2013). At the top of the platform, it displays how many tasks and workers are managed (Kittur et al., 2013). Kittur et al. (2013), add that complex tasks must be broken into smaller subtasks, each planned with specific requirements and features that must be allocated to appropriate sets of workers who themselves must be appropriately encouraged, nominated through status, and planned through hierarchy. Kittur et al. (2013), further state that jobs to be embarked upon may be arranged through multi-stage workflows in which personnel may join forces either synchronously or asynchronously Kittur et al. (2013), conclude about Figure 4 by saying that quality assurance is required to guarantee each staff member's output is of high quality.

Chhabra et al. (2015), enhances that crowdsourcing in education through analysis of case studies on open online learning communities has played an important role, where certain roles are played by the individuals involved and also future roles are proposed that could be generated through these organizations. Chhabra et al. (2015) added that they found that the crowdsourcing model delivers all the roles that are recognized in the traditional environments, as well as offering some extra value due to the social networking tools.

Keatinga and Furberg (2013), say that an appropriate stage to support the crowdsourcing activities is recognized after the investigator has defined the target audience and planned an engagement instrument, as this stage offers an environment for collaborating and exchanging value with members.

Selection criteria for a crowdsourcing platform must contain the accessibility of the resource to participants of the target audience, the capability to incorporate appropriate engagement tools to drive continuing contribution, and the means to allocate incentives after completion of events (Keatinga and Furberg, 2013). The online education and crowdsourcing communities are addressing similar problems in educating, motivating, and evaluating students and workers (Williams et al., 2015).

The study of crowdsourcing is a fundamental research area that has been growing over the previous years (Hetmank, 2013). Crowdsourcing is a developing disseminated critical problem solving model on the merging of human and machine computation (Mao et al., 2015). The quick development of inventive Internet centered information and communication technologies opened up another playing field for associations to achieve their objectives (Skarzauskaite, 2012).

Crowdsourcing is quickly picking up acknowledgement and has even been used to back political decision making (Thebault-Spieker et al., 2015). Thebault-Spieker et al. (2015) say that one must ensure that it is mainly a practical thought and that it does not empower shared discrepancies (Giudice, 2010).

We may subject everything to danger by either missing an exceptional opportunity or getting carried away from a development bubble; separating the wheat from the chaff requires effort (Salminen, 2015). Organizations with imaginative business thoughts have expanded in this manner making a new flow of advancements, thus getting the most out of the Internet's together with individuals' ability to contribute and relate themselves through normal interests by means of crowdsourcing (Linkruus et al., 2012).

According to Llorente et al. (2015), the most accessible type of crowdsourcing stages are committed to collecting ideas about a given topic and vote on the most well-known alternative. Some crowdsourcing application cases are the proposition of an item logo or the proposal perfect for promoting a drive (Llorente et al., 2015). Crowdsourcing has influenced large crowds of people to complete small tasks thus drawing a lot of attention as a research tool and researchers have found crowdsourcing as a tool for assembling subjective decisions about various online media very useful (Egelman et al., 2015).

Figure 5 shows the arrangement or divisions of how crowdsourcing has been used in different applications like on the development of crowdsourcing applications and systems that have been used on various environments (Yuen et al., 2011). The other aspect of crowdsourcing involves computational strategies, advancement, quality control instruments, data sharing frameworks and executing investigations in view of the scientific classification of crowdsourcing as it appears in Figure 5 (Yuen et al., 2011).

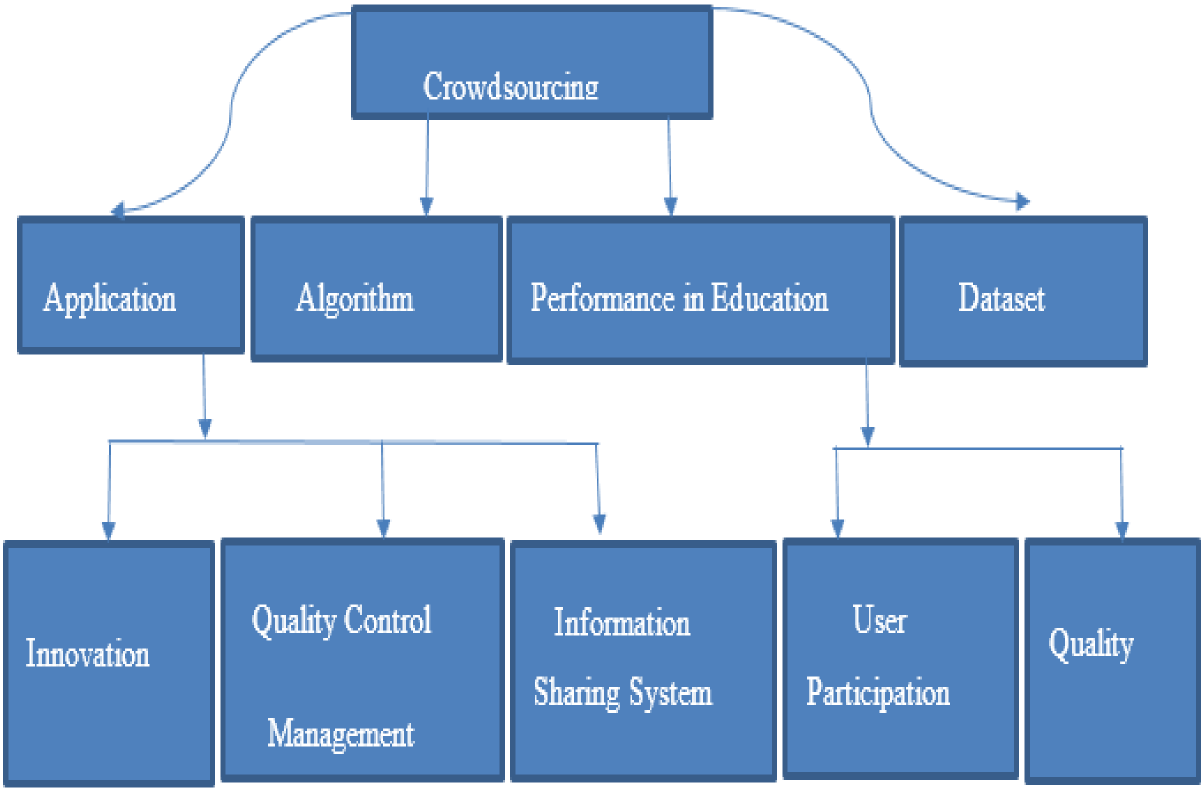

Figure 5: Taxonomy in crowdsourcing (Yuen et al., 2011)

Crowdsourcing has been in operation for quite some time now and McIntosh(2015), trusts that when crowdsourcing is utilized as a convenient instrument for business, instruction becomes significant. He additionally expresses that when contrasted with old-style classrooms, it delivers much better since it offers generally safety, a minimal effort approach to achieve high satisfaction and proceed with engagement in the classroom given existing challenges in holding students' interest (McIntosh, 2015).

Interestingly Yuen et al. (2011) produced the ESP game to get Internet users to reach a decision on image labels, Dalvi et al. (2013) used crowdsourcing to organize various website layouts. Bernstein et al. (2012) generated a word processor that uses crowdsourcing for editing, Estellés-Arolas and González-Ladrón-De-Guevara (2012) displayed how crowdsourcing can add value to the design process, and in their study, they utilized crowdsourcing to get a response about different types of banner advertisements (Egelman et al., 2015).

Crowdsourcing when put into practice can be useful for enhanced knowledge development and today, this idea is stated as crowdsourcing learning or crowd learning (Llorente et al., 2015). In crowdsourcing education, a group of pupils from the same or various institutions can work together in self proposed combined projects (Llorente et al., 2015).

Crowdsourcing has been used in software engineering and is a way of undertaking any external software engineering work by a projected, possibly large group of online workers in an open call format (Mao et al., 2015). Crowdsourced software engineering usually includes three types of players, namely the owners who have software development work that needs to be done, workers who partake in developing software and platforms that provide an online market within which requesters and workers can meet as shown in figure 6 (Mao et al., 2015).

Tan et al. (2015) in their study defined a manageable crowdsourcing open application which is particularly practically built from the direction of an institution, and important views have been recorded from the information investigation team. Recommendations were created for the useful application framework as shown in Figure 6 and the paper model has combined these proposals. Furthermore, two extra highlights were added with one being the capacity to share an event report, and secondly the empowering of participants to effectively report on events (Tan et al., 2015). It was discovered in one of their meetings that the paper model is encouraging, including remarks which projected that the framework was up to standard and informative (Tan et al., 2015).

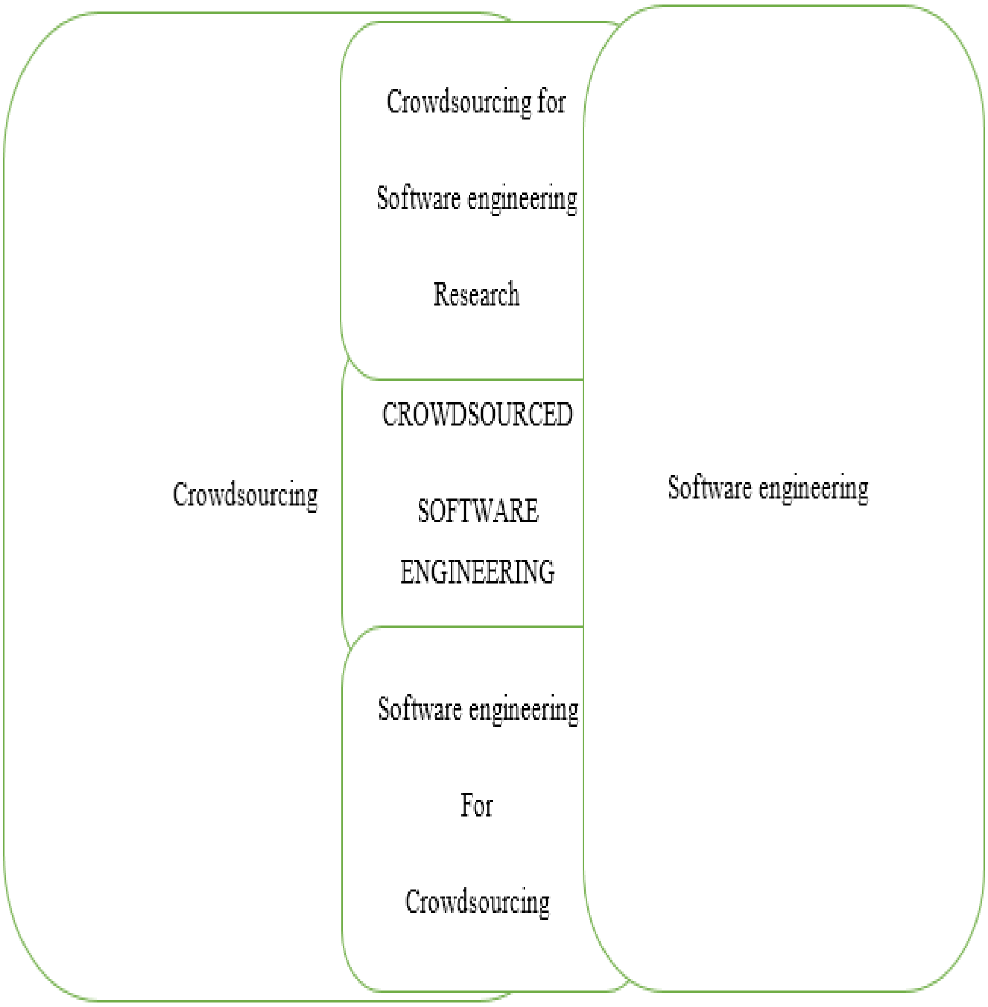

Figure 6: Crowdsourcing and software engineering (Mao et al., 2015)

2.2 Benefits of crowdsourcing

Papadopoulou and Giaoutzi (2014) elaborate by saying crowdsourcing is a new approach for knowledge acquisition, information dissemination, and the interchange of opinions amongst professionals and the crowds. There are several benefits of crowdsourcing and noticeable the most vibrant one is being a problem solving and invention mechanism (Way et al., 2011). Designed on this approach, many problems can be distributed, and decided over the implementation of a sensible online stage planned for such an initiative (Papadopoulou and Giaoutzi, 2014). The management of shared knowledge and intelligence outcome in the establishment of state-of-the-art designs can result in participants being compensated with enticements as an acknowledgment of their contribution (Papadopoulou and Giaoutzi, 2014). The demanded benefits of crowdsourcing include an easy way to access a diverse personnel, several solutions, lower labour rates and condensed time to market (Mao et al., 2015). The benefits of crowdsourcing as derived from literature reviewed are discussed below:

FASTER RESULTS

The researchers Barbier et al. (2012), express that a multitude of individuals can decide on a few encounters faster than single persons or rather little gatherings. He keeps on expressing that a group can conveniently provide information about circumstances touching on the group and crowdsourced data can be utilized to profit the group by giving information or results quicker than through customary means (Barbier et al., 2012).

The crowd is normally quick to answer in solving the problem as they are always willingly waiting for a new task to be posted on the crowdsourcing platform. Some high tech platforms even have improvement time of about an hour, producing a wild range of responses immediately (Ellero et al., 2015). Crowdsourcing is a beneficial method for it can be used by almost any crowd for a variety of aims.

An example is when the Defence Advanced Research Projects Agency (DARPA) released ten helium balloons and tested people to trace their whereabouts. The results confirmed that a combined group can solve time sensitive problems (Barbier et al., 2012). Another case is that in 2012, DARPA took a resolution to crowdsource its next generation amphibious fighting vehicle after terminating its old-style procurement that had already cost taxpayers $13 billion. The appealing work of art was declared in April 2013, just six months after the competition was launched (Boudreau and Jeppesen, 2015).

MORE AFFORDABLE

This is as a result of the crowdsourcing reducing the fee of getting the project done. These include fees from salaries because the individuals managing the project would want to be paid; if you opted to try and do subcontracting on the challenge an independent consultant is expensive since it is profit oriented (Maheshwari and Janssen, 2014). For such situations, additional inspiration utilized for the crowd must be considered, which can be more in a form of payment, goodwill, satisfaction, and reputation, among others (Chatzimilioudis et al., 2012).

Crowdsourcing can likewise lead an association to lower expenses from their crowdsourced thought, e.g. Colgate Speed Stick utilized the group to deliver a Super Bowl advert for the bargain basement cost of $17,000, compared with the charges normally connected with customary agencies (Boudreau and Jeppesen, 2015).

Crowdsourcing is believed to have undoubtedly controlled lowering the expenses of doing tests while at the same time refining the nature of quality output; this has allowed all entities, not only the biggest partnerships, to consider the advantages they can obtain from the crowd (Boudreau and Jeppesen, 2015). Advantages of crowdsourcing are that it gives associations access to a potentially huge amount of work outside of the association which can finish fundamental assignments regularly in a small amount of time and at a small amount of the costs than if similar exercises were conducted within the organization (Whitla, 2009).

OPTIMIZE CREATIVITY AND INNOVATIVENESS

This is because crowdsourcing provides a competitive room for experts to bring their best in order to win the reward or earn recognition in developing the winning task. Hence the professionals tend to be more inventive and innovative in completing the task at hand. There is also improved assignation and safeguarding of internal talent; Thomson Reuters practices crowdsourcing within to tap into the skills of its 17,000 technologists resulting in new problem solvers and breaking down department silos. Crowdsourcing offers attractive prospects for gaining new perceptions on a variety of problems (Esteves et al., 2012).

It has also been found that some firms are utilizing crowdsourcing to trace large numbers of individuals eager to complete new invented repetitive tasks for constrained financial compensation (Whitla, 2009). Although crowdsourcing may not be the only answer, if it is planned, ongoing managed and integrated according to a different strategy, then it offers organizations a better method to test with new problem solving strategies to design and develop new product prototypes to make new connections in order to develop talent including to gauge what works (Boudreau and Jeppesen, 2015).

THERE IS NO RISK AND EXTRA FEES

This is because you will only pay if you have found what you were looking for as the deliverables in the tasks. If the deliverables are not met then you will not make the payment on the task; not like in the traditional model where you will pay an employee or a private entity to do the job and risk not getting what you intended to get in the project or task. Most likely the private entity will put in hidden costs to the project so that they can benefit (Esteves et al., 2012).

It has furthermore been found that a couple of organizations are utilizing crowdsourcing to follow great quantities of people eager to complete large amounts of repetitive work for constrained financial repayment (Whitla, 2009) . Despite the fact that crowdsourcing might not be the main resolution, if it is applied correctly, controlled and joined by a good strategy, it offers associations a superior technique to test with new critical thinking methodologies at a reasonable cost (Boudreau and Jeppesen, 2015).

MINIMUM MANAGEMENT

Since the task is taken out of the business' daily operations, there is less supervision of the whole process of coming up with an idea. You can only offer half technical assistance to guarantee quality in the received ideas or data (Hofstetter et al., 2018). Crowdsourcing follows the next technique in its execution where firms find a task or many tasks that are currently being conducted within the organization with minimum supervision (Whitla, 2009).

Instead of continuing to complete this activity within the organization, the tasks are released to a group of strangers who are requested to do the task on the firm's behalf for an agreed amount of cash (Whitla, 2009). Crowdsourcing gives an extensive variety of working arrangements as groups of individuals work on the task. This enhances coordinated efforts that are vital for thinking of a wide choice of best answers for the issue at hand (Bergvall-Kåreborn and Howcroft, 2013).

2.3 Challenges of crowdsourcing

Although from above there was discussion of some of the advantages that crowdsourcing can deliver to problem solving, now the focus will be on the challenges of utilizing crowdsourcing as derived from literature reviewed and they are discussed as follows below:

LACK OF CONFIDENTIALITY

Sending problems out to a huge group of outsiders seems at odds with conservative corporate wisdom and supervisors who have by tradition looked inward for solutions; they are reasonably suspicious of whether intellectual property can be protected if it is exposed so publicly (Boudreau and Jeppesen, 2015). This is because with crowdsourcing you have shared the problem with a high number of people where some of them may be your rivals or a threat to the thoughtful solutions you are to get from the task (Sullivan, 2016).

As indicated by Buecheler et al. (2010), crowdsourcing is in an unsafe situation of dealing with the unknown, where emergence and the responses to developing behaviour play a significant part. The individuals from the crowd are strange and eventuality plans for the surprizing behaviour of this interacting cannot be fully prepared in advance (Buecheler et al., 2010).

LACK OF COMMUNICATION

This is because the issuer of the task only sends the problem to the crowd working on the task without communicating during the process. Hence if they do not clearly understand the task they may make the wrong guess decision and find themselves delivering a job out of the aims of the task because they never had a chance to interact with the source of the task (Esteves et al., 2012). Another challenge associated with utilizing crowdsourcing is that, although the method works on the principle that two heads are better than one, sometimes a crowd can return a vast amount of noise that may be of little significance (Whitla, 2009).

LAWSUITS

This sets a challenge because the partakers may be affiliated with institutions that do not allow the use of intellectual property of the participant in any other way than in the organization itself. This may pose a problem when there will be a need of transferring the right of the invention or idea to the source of the task.

Partakers who face such a challenge include lecturers and professors from universities where their contracts do not allow them to engage in activities like crowdsourcing (Sullivan, 2016). Lawful matters in regard to the possession of ideas also need to be clearly addressed and in some types of work crowdsourcing will not be functioning, for example there is a limited ability to use the methodology where the information to be gathered or project being worked on is private in nature (Whitla, 2009).

IDEAS ARE NOT SOLUTIONS

This is a trial because the crowd may offer ideas to solving a problem but lack in practicality such that they can form resolutions. It sets a challenge to the institute or the source of the task in working on the application strategies of the new ideas rather than the solution. Such an incidence will still be costly to the organization in getting a team to make the idea a solution (Bergvall-Kåreborn et al., 2009).

UNFAIR TO WORKERS

When observing crowdsourcing, ethical concerns are still related to the trend of outsourcing where organizations are substituting their own highly paid workers with much lower paid workers from outside the institute (Whitla, 2009). Crowdsourcing is very partial to the partakers who do not get compensated or credited even after devoting their time and resources in trying to come up with the solution. This provides room for them to be dispirited to additional contest as implementation of the task takes up some of their resources (Esteves et al., 2012).

According to Whitla (2009), many marketing professionals, from copywriters to photographers, account executives to researchers, may discover that their services could to a greater or lesser degree be crowdsourced and also find how companies will treat present workers (Whitla, 2009). The very low pay rates given to those crowds for the typical success, raises questions of exploitation of workers for the organisations' benefit (Cove, 2007).

COST OF SETTING THE INFRASTRUCTURE AND REWARD

Although the usual cost of crowdsourcing is known to be lower than the outdated way of executing a task, as an institute there are still costs related with crowdsourcing. This will consist of getting the right technology to harness all the information and organize a reward for the winning contributor. The cost of equipment lies in setting up a strong platform to create collaboration between the source and participants practical (Dhiya et al., 2015).

LACK OF CREDIBILITY OF THE CROWD

Although the selection is done on some crowdsourcing prototypes, the reliability of the credentials that they submit as contestants is debatable. This is because in crowdsourcing one is likely to deal with a pool of unknown crowds from all over the globe or locality which poses a threat from contestants with fake identifications (Hofstetter et al., 2017).

CYBERCRIME ISSUES

With the ever growing complexity of internet connections, offenders are also finding an opportunity to participate in criminal activities like cybercrime. The criminals may get their hands on the established solutions and use them as their own or divert the payment of the reward from the winner to them. This vulnerability comes from the complication of running a crowdsourcing promotion as the players are less likely to make personal collaborations since they mostly rely on cyber collaborations (Esteves et al., 2012)

2.4 Crowdsourcing in education

Crowdsourcing as defined by Jeff Howe (2006) is thought to be the action of taking a job initially done by a selected worker and subcontracting it to an open ended, generally big group of workers in the form of an open call (Skarzauskaite, 2012). Educational organisations have many tasks that are vital, but they want time and energy that could be used to focus on teaching and crowdsourcing is the required resolution. The term being new shows that there is no key literature on the topic of how this directive is operated in educational proceedings (Skarzauskaite, 2012).

Crowdsourcing can be used in numerous instances in higher education institutions (Solemon et al., 2013). It can be mapped into four methodologies of crowdsourcing, as deliberated in Howe (2008). Skarzauskaite (2012) also has presented numerous other additional strategies such as crowd democracy, and crowd reviews. Solemon et al. (2013) discuss the four crowdsourcing approaches as follows:

Collective intelligence or crowd wisdom includes dissemination of knowledge and gathering thoughts from the crowds to resolve problems, or predict future results or help direct a university's corporate tactic. This method in higher education institutions may include crowdsourcing projects to come up with marketing tactics for newly intended products and thoughts for achieving constant development to increase academic standards. In addition, universities also include those who focus on crowdsourcing by collecting intended initiatives from students, staff and community to achieve the university's goals.

Crowd creation is another method where those who trust in crowdsourcing use it to create a service. There are numerous projects linked with this type of crowdsourcing in education and samples include those investigators who go to the public to create a database of online historic exhibits, bibliographic data, linguistics, and biology as well as image classification. Some investigators also rely on the public to produce online schoolbooks, class content and other materials. Lastly, investigators present projects that offer on-campus support by the masses and the operational charge of engaging permanent support staff is reduced.

Crowd funding is a method that refers to the willpower of people who network and pool their money through minor contributions from several crowds. It is usually conducted online to support the necessary funding for educational developments. Requirements for financing student education and funds for scholarships, need this approach because it is believed to be the most suitable and comprehensive strategy. Lately, a new development in this method is raising money for educational investigation activities.

Crowd voting includes any technique of gathering the views, thoughts and decisions of the community by way of voting. Tertiary organizations and colleges are seen to have applied crowd voting for their competition based creativities with the objective to empower learners in making choices.

Though crowdsourcing has developed as a common term for trades, writers, and media, nothing much has been in print worthy of being an educational technique to improve learning (Way et al., 2011). Applying the joint knowledge of a larger crowd internal or external of the institution for carrying out organizational everyday jobs has always been a valuable method for organizations (Amrollahi, 2016). Nevertheless, after the word crowdsourcing was first made known by Howe (2006) as a newly discovery method, it attracted much more attention in both school and training institutions (Amrollahi, 2016).

When this study explored why educational research and crowdsourcing can be exciting Weld et al.(2012) brought about three reasons why education is an exciting direction for crowdsourcing exploration. The reasons are that crowd techniques will be vital to deliver quality education on some levels, current techniques are ready for being applied to this new area, and lastly online education embodies a new relatively unexplored way of forming crowds (Weld et al., 2012).

Tarasowa et al. (2014), indicate that crowd learning considers the information, inventiveness and effectiveness of the masses for the creation of rich semantically organized e-learning content. They carry on to elaborate that crowd learning combines the wiki style of joint content authoring with Shareable Content Object Reference Model necessities for re-usability. On another note, they say it permits splitting the learning material into learning items with an adjustable level of granularity. Lastly, the crowd learn concept is established on five vital components which are standard compliance, semantic structuring, enhanced possibilities for re-use, crowd-sourcing and social networking (Tarasowa et al., 2014).

Let us briefly look at how these components are vital for the e-learning concept. Standard compliance is payments related to building high quality e-learning content which at times are extraordinary. Semantic structuring is the dealing with vast learning objects and breaks them into fine-grained learning items. Re-use deals with growing the cost productivity of content creation, developing the benefit of e-learning substance, supporting the movement and variation to new requirements.

Crowd sourcing deals with considering the view that there are even now large numbers of unprofessional and professional users who are working together and contributing to the Social Web. Involving the impact of such crowds can meaningfully advance and widen the spreading of e-learning content. Social networking contracts with the theoretic fundamentals for e-learning 2.0 are drawn from societal constructivism as shown in Figure 7. Again social networking deals with the fundamentals for e-learning 2.0 which is presumed that students absorb as they work together to understand their experiences and create meaning (Wang et al., 2012).

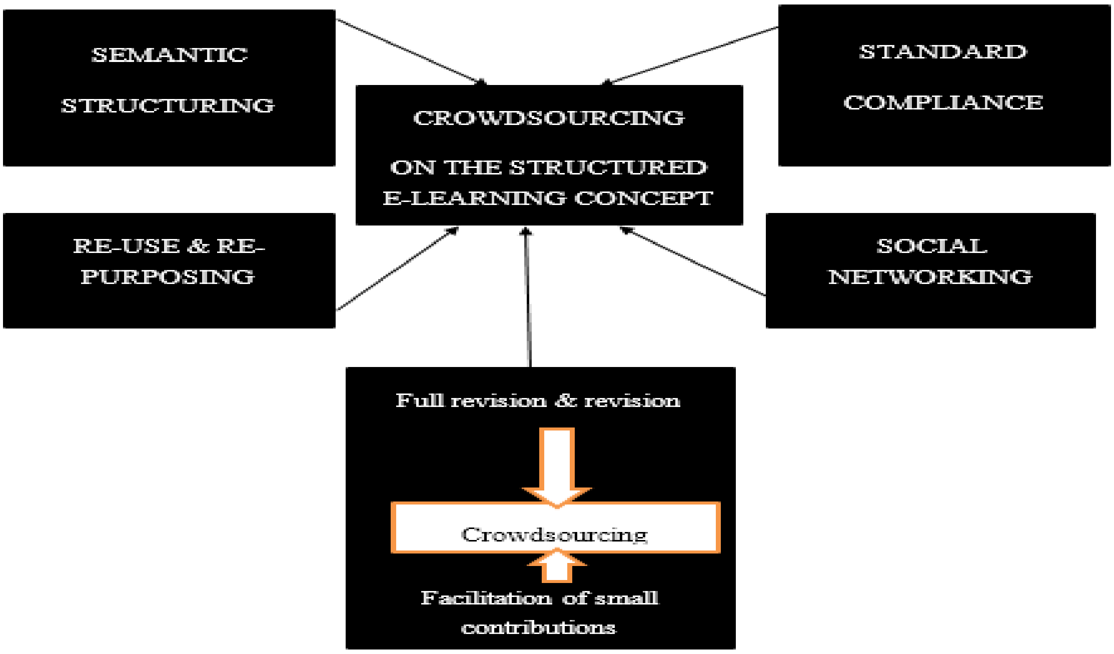

Figure 7: Crowd-learn concept (Tarasowa et al., 2014)

There is a taxonomy for crowdsourcing which, when applied appropriately, can be a resolution in the design of efficiently applying crowdsourcing in learning activities (Farasat et al., 2017). Regrettably, many crowdsourcing projects also were ineffective primarily due to lack of background of the participants, or lack of structure and management (Farasat et al., 2017). In order to understand the success criteria, here is a presentation for a taxonomy for crowdsourcing in education based on five elements and they are as follows:

- Source: which aids in understanding who is producing the content across experimentations as this may have been current students, alumni, crowds, and teachers. These may be from a residential course. Furthermore, within those groups, participants were sometimes pre-filtered.

- Educational content knowledge: where tasks involve different levels of educational content knowledge as well as a background in instruction of knowledge.

- Effort and complexity of task: where inspiring pupils put in small amounts of time; contributing a suggestion is different from cheering them to contribute substantial time.

- Domain knowledge required: where clarifying the significance of information, or tagging info with perceptions and learning objectives involves a high level of domain knowledge.

- The structure provided: where crowdsourcing regularly works better with a high level of guidance, and as the scale of classes develops these do not necessarily limit our capability to crowdsource content (Farasat et al., 2017).

Way et al. (2011) envision whether crowdsourcing has the prospect of being beneficial to any crowdsourcing thought to increase improvement and learning outcomes in culinary and hospitality training. In their study they use a ground breaking student project joining collaborative problem solving through a technology enabled format (Way et al., 2011). They added that this educational approach appears mainly applicable to pupils studying in temporary economies or distance arrangements by creating a global style to learning. Lastly they add that this way may be useful for teaching competence in students to facilitate ground-breaking behaviours independently or collaboratively (Way et al., 2011).

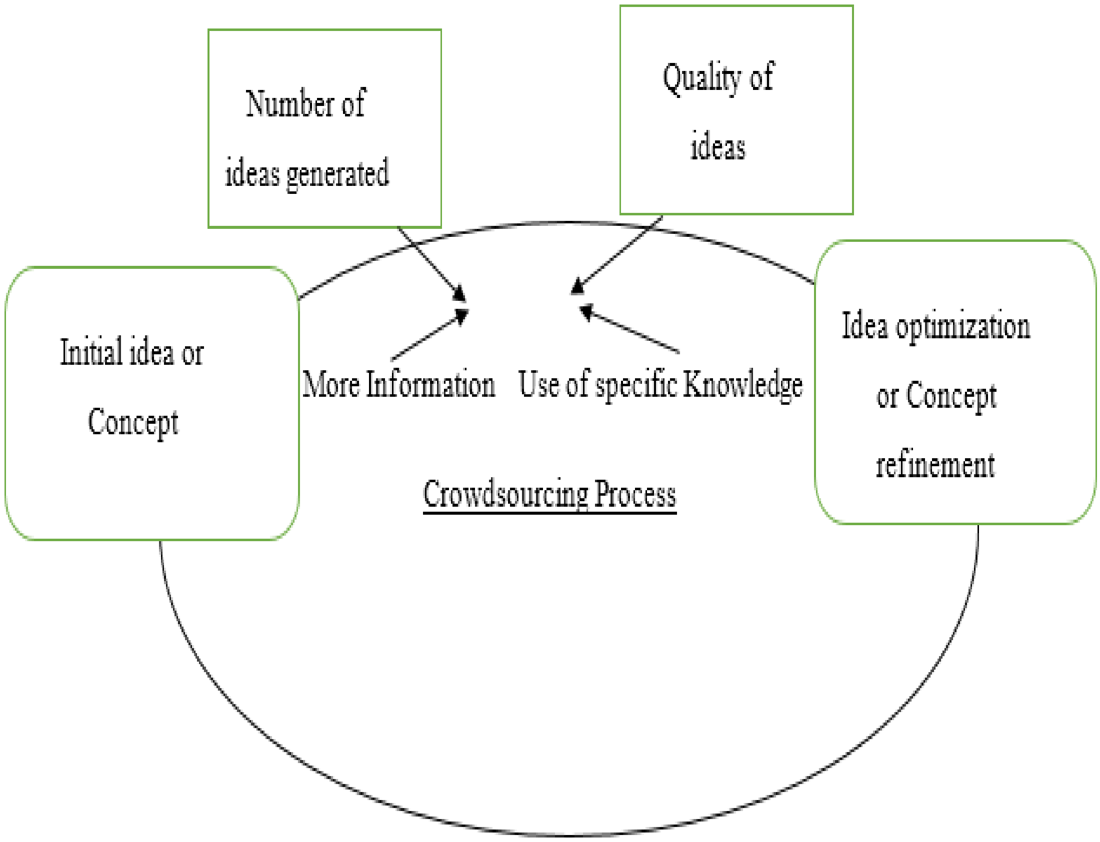

Figure 8: Innovation project outcomes from crowdsourcing (Way et al., 2011)

Figure 8 displays how these theoretical relations in the crowdsourcing process would also permit a better use of precise information and allow those with superior proficiency to share that information in a useful manner (Way et al., 2011). A key benefit of crowdsourcing in an educational innovation condition is the cross-functional team knowledge; this knowledge can be useful in learning how cross-functional variety can be achieved and supported during crowd problem solving. Peer learning has demonstrated to be a major result that increases learning strengthening by adjusting to various learning style preferences (Way et al., 2011).

Figure 9 shows the key learning outcomes, which involve thought improvement and feedback problem-solving practices. These techniques are vital components of innovation managing process models (Way et al., 2011). There are vital benefits in the crowdsourcing model using a learn-by-doing capability approach and these benefits are also proposed to be facilitated using a crowdsourcing approach (Way et al., 2011).

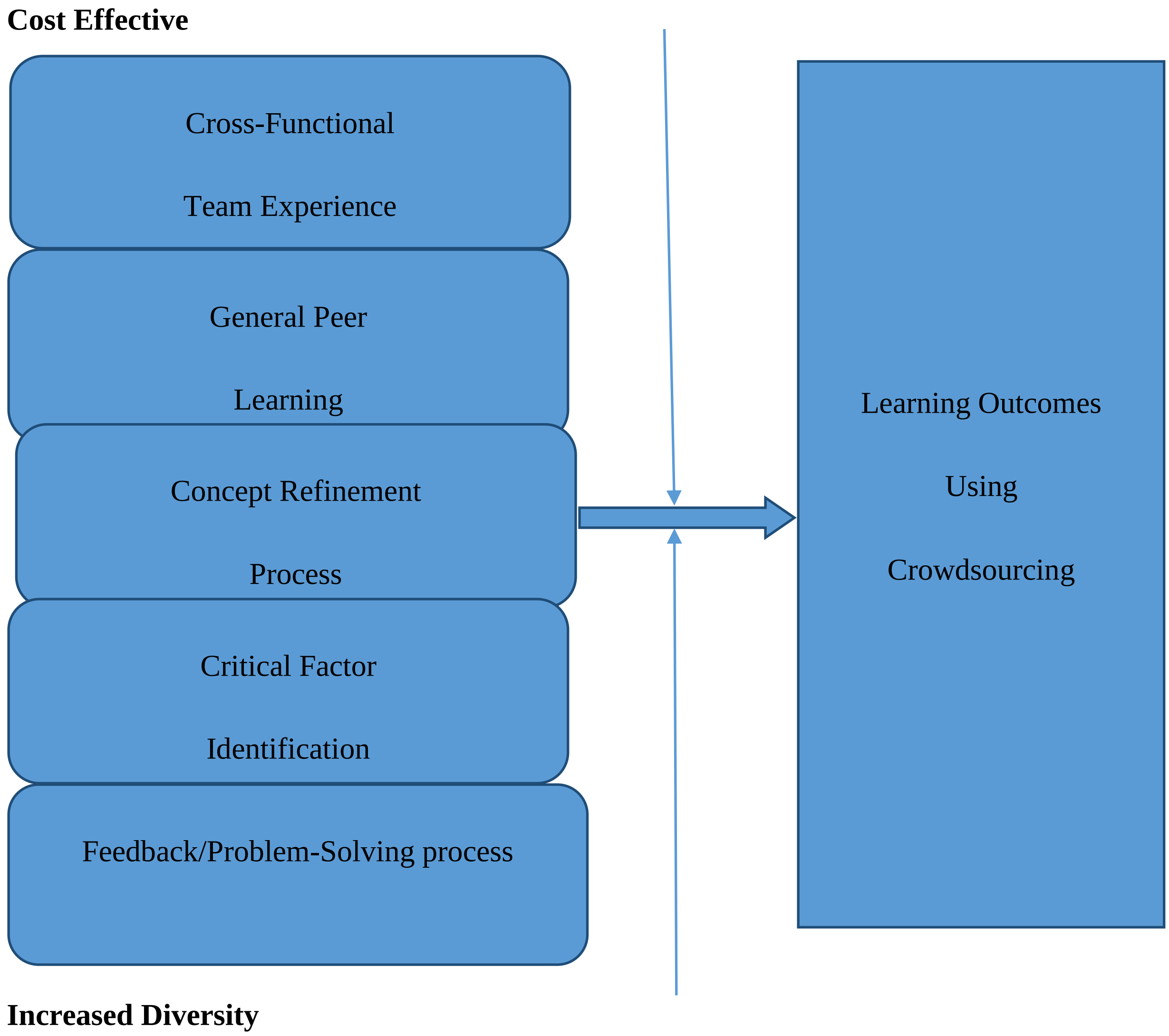

Figure 9: Learning outcomes using crowdsourcing as a problem-solving educational tool (Way et al., 2011)

Many writers have done a lot of reviews when it comes to how crowdsourcing has been successfully applied in educational events and most of them agree that crowdsourcing is indeed a successful model (Skarzauskaite, 2012). Crowdsourcing content creation motivates learners’ commitment at the frontline between educational content and students’ interests (Hills, 2015).

Crowdsourced content is likewise utilized as a part of delivering programs for classes and reading material. Composing and producing a complete book is and has been a problematic practice for one or a few people to adapt to (Skarzauskaite, 2012). The goal to which learners focus their liveliness is openly defined as the task of realising classroom content outside the classroom, building related content, clarifying how the content and the real-world applications are connected (Hills, 2015). For the fact that it is a web-based, distributed problem solving and structure model for business, it qualifies as a suitable model for allowing educational establishments in their processes (Skarzauskaite, 2012).

Crowdsourcing has emerged as a dynamic part of instruction on the web since it supports the openness, distribution of assets and information contributed by groups (Skarzauskaite, 2012). Crowdsourcing in the development and utilization of educational resources permits web tools to effect collaboration and produce resources with the help of user groups and other participants. Crowdsourcing in this instance takes the learner as producer model, which is an evolving approach to education that borrows much from earlier literature on learner-centred education (Hills, 2015). One of the good and practical examples is the University of Alabama where it allowed the community to participate in their library projects.

They allowed users to tag materials from current collections, which means that from the luxury of their own workstation users can tag people, places and accidents with the material, resulting in the creation of crowdsourced content (Skarzauskaite, 2012). Although crowdsourcing has brought about a lot of progress in the way educational content is collected, there is also progress in the educational aids in terms of an effective teaching-learning process (Chhabra et al., 2015). Most of the educational content available to the learners is optimal in the perspective of being up to date, complete and easy to understand. There is a necessity to iteratively develop the educational material based on the feedback collected from the students' learning experience through the application of crowdsourcing methods (Chhabra et al., 2015). In one example, Chhabra et al. (2015) created a technique of textbook creation and refinement joined with the benefits of an annotation system. They provided an incorporated platform in which current content is joint with a collaborating environment where the people in the crowd collaborate so the quality of material can be greatly improved on a constant, ongoing basis.

There is still a persistent need for scientific engagement in this field and currently many tasks that are inappropriate for individuals endure to challenge even the most sophisticated computer programs like image annotation (Yuen et al., 2011). Analysis of crowdsourcing can also rely on many main theoretic frameworks that are often applied to the study of ICTs (Asmolov, 2014). The idea of connective activity created by Bailard and Livingston (2014) can add immensely to understanding the dynamic of the procedure behind crowdsourcing and its systems. Systems join the administrative properties of crowdsourcing devices and it is its disseminations that always connect deployment to communication (Asmolov, 2014).

At present, there are a lot of structures that are being created for various purposes. One reason being a need for considering how existing technology can develop students educational content knowledge (Cherner et al., 2014). The TPACK structure was developed for this reason (Vornanen et al., 2016). TPACK is a framework clarifying how learning of substance, teaching method, and innovation can be seriously incorporated into the guideline educators give to learners (Vornanen et al., 2016).

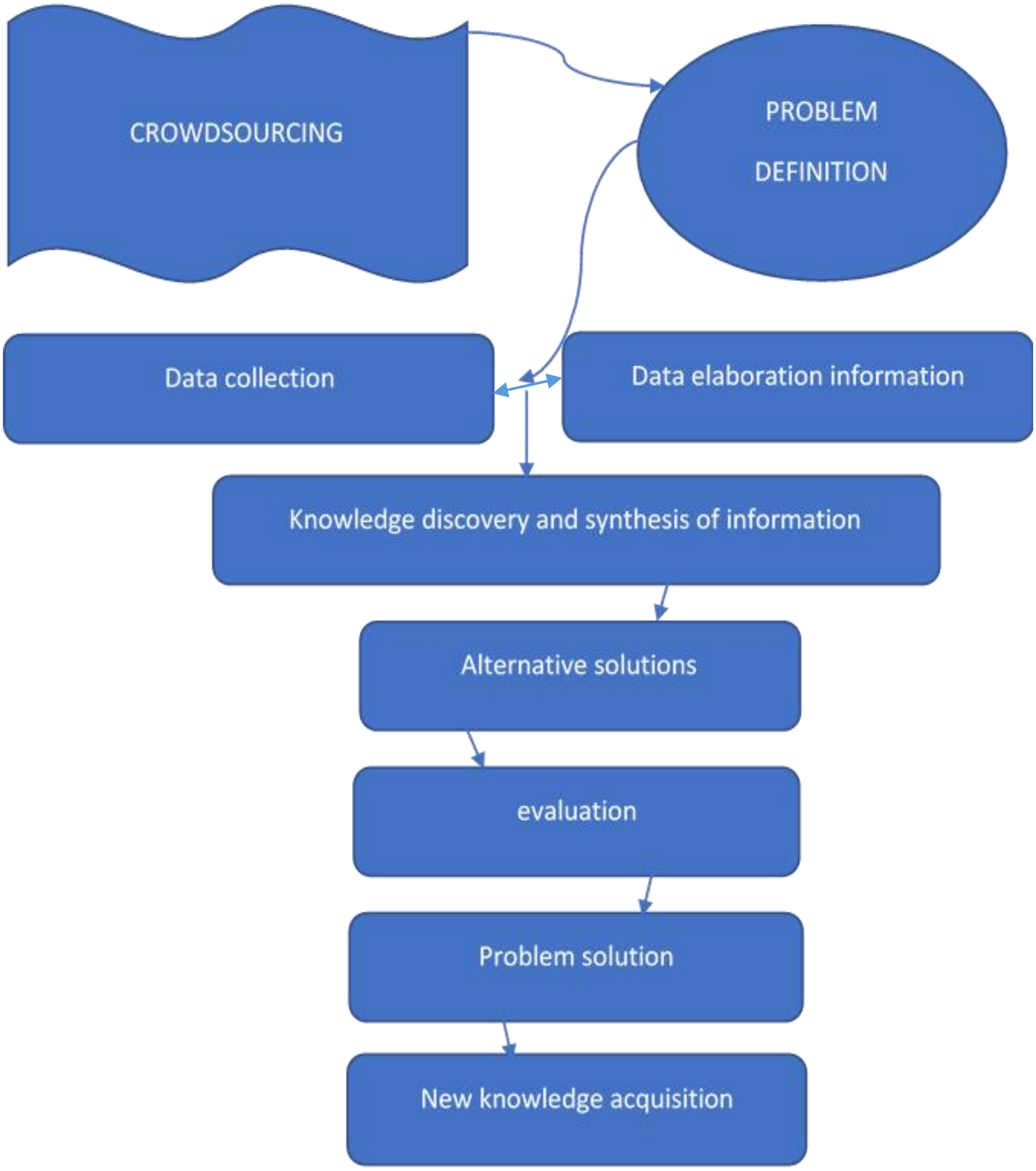

Figure 10: The concept of crowdsourcing for problem-solving (Papadopoulou and Giaoutzi, 2014)

Figure 10 gives an illustration of how crowdsourcing is another strategy for information securing, data conveyance, trading of sentiments together with translations among experts and the crowd. Figure 10 further shows an approach on how extremely different sorts of issues can be spread and settled over the utilization of a proposed web-based solution. According to Barbier et al. (2012), crowdsourcing is an operational creation display that has been created as of late and the term crowdsourcing characterizes an imaginative electronic plan of action that associates the innovative resolutions of a spread system of people through amounts of work to an open call for recommendations (Barbier et al., 2012).

Crowdsourcing today is generally used to characterize a procedure which may include means or strategies for information and data usage in an educational setting. This may include a gigantic gathering of laborers or students who are not organized to deliver a normal substance (Papadopoulou and Giaoutzi, 2014). Crowdsourcing is similarly observed as a sort of participative online activity in which an individual or an educational institution proposes to a social occasion of individuals with fluctuating learning to join so they can provide a solution via a method of a versatile open call (Hörler, 2014).

2.5 Some of the frameworks used in crowdsourcing

We look at these frameworks in terms of how they have been utilized and our focus is mainly on their application, strength and weaknesses.

Framework 1 Adaptive learning class assistant (Hattem et al., 2003)

APPLICATION

This framework uses adaptive learning measures into creating a web-based student learning experience and teacher's ability to track student progress through the framework. This platform has a reasonable way of teaching students in a progressive and adaptive way. The platform will generate questions that will progressively get easier and more difficult depending on the answers given by the students, which would be critical in evaluating their understanding. This method also proves to be a good grading book and a reporting feature to its users. Each user has its own profile in the platform proposed by the framework. These different profiles are categorized based on the different parties involved in the system. There are profiles for teachers, students and teaching assistants. This shows how the proposed framework operates and the components involved in it (Hattem et al., 2003).

STRENGTH

This framework provides the students with a model of evaluating their progress as they are working through the problems presented on the platform. This provides efficiency as the students through the platform can be able to know if they are progressing given the different scaling levels of difficulty of their questions. This is because the platform, through its scaling, can move students to more difficult problems if they prove to have understood the less difficult problems. The platform also includes teaching assistants who can view student's information ranging from their questions and answers to stand as a mediator and give comments to make the exercise efficient if delivering well answered questions (Hattem et al., 2003).

WEAKNESSES

The platform is not specifically clear as to how it is going to cater for the IT expertise of the teachers to configure the platform with newer problems. In that regard outsourcing or getting an IT expert will increase the cost of running the platform and delivering on the service (Hattem et al., 2003).

Framework 2 Crowdsourcing annotation System (Chhabra et al., 2015)

APPLICATION

This system gives an online stage to the writers to audit their books through collective assistance from educators and their learners. In this stage, the book is posted on the platform with the goal that the students can give group sourced comments to the others to consider while auditing the book. The evaluating is key so the book incorporates the perspectives of the clients of the book with the goals that are incorporated in the updated adaptation.

The remarks that the students provide are under the supervision of their educators. The educators see the students' suggestion and then discuss the content, if required. In its practical application, the book is posted on the web, where alongside it there will be a window permitting to post your remarks. You will likewise discover different explanations in that window from different learners looking into the book as it meets the student's desires (Chhabra et al., 2015).

STRENGTH

This system expresses that at this stage it has the favorable position of lessening the normal costs of investigating the book because of its writers forming a crowdsourcing model to decrease the cost as the student's remarks are open on the web. Consequently, effectiveness is accomplished in such a manner. The crowd of collaborating learners gives a widely ranging assortment of remarks with the end goal being that a decision of good remarks from the book clients will be utilized to get the recently explored book (Chhabra et al., 2015).

WEAKNESS

Its primary shortcoming is that a few students might be too modest to voice their remarks in a gathering setting as the remarks are noticeable to all users at this stage, giving less classification. Another challenge in the structure is that at that stage no one is sure about what will draw the pack even regarding rewards for the students offering remarks (Chhabra et al., 2015).

Framework 3 Decision Making Support System (Chiu et al., 2014)

APPLICATION

This framework is for making and measuring how to make a decent crowdsourcing system for basic leadership at all levels of association's instruction comprehensiveness. The system has 4 segments that pioneers ought to consider in making their own particular structures on crowdsourcing. As the system delivers efficiency to the task, crowd, process and outcome, issue may arise that requires decisions to be made at the time (Chiu et al., 2014).

STRENGTH

The framework calls attention to the fact that the undertaking ought to be unmistakably characterized and imparted to the group laborers; this is critical to complete the task effectively. The group ought to likewise be given the required skills and attention have a satisfactory sized group. In the process part, the proposed system must have a powerful data innovation foundation to run the platform and every one of its activities. Ultimately, the task ought to have methods to assess all stages as realistic so as to give the ideal answer to the issue.

To accomplish this, there must likewise be legitimate administration to run it, and there should be an appraisal of the behavioral measurements considering how the workers of the association will respond to the new crowdsourced task regarding their professional stability and reactions to the general gatherings engaged with the stage. Looking into the IT skills, it is basic in completing an ability on any online stage to sustain the productivity (Chiu et al., 2014).

WEAKNESSES

The framework is reviewed for generalizing crowdsourcing models in its discussion; this is because they differ greatly on the category of information the platform is working with. Therefore, it lacks the diversity in its formulation and there is a need to be brief on a certain industry (Chiu et al., 2014).

Framework 4 Hybrid training solutions, data collection and feasibility analysis (Sisilia et al., 2015)

APPLICATION

Sisilia et al. (2015) proposed a framework that combines universities with collaborating businesses. The Abdis Telkom University remains as the screen of information and its investigation. The system includes TDC Telkom which is a huge telecom organization in Indonesia utilizing crowdsourcing models to prepare its staff and enhance the SME office. CDC, being the requester for this situation, is the one setting out the assignment representatives. Abdis Telkom University act as a facilitator and a go-between in the stage; this is because the university has the correct skills and framework to carry such an assignment (Sisilia et al., 2015).

According to the structure of the preparation program, the staff at the organization gets the benefit of preparing to enhance the representative's capability. Abdis gives a web-based learning knowledge to the workers while preparing material including recordings, introductions, and internet scrutinizing to new or existing staff individuals. The site likewise provides cooperation between addresses with the adequate information and involvement in the field. This will be basic for the credit investigation of the organization as they are requiring new data about how to assess a chance for the SME's who are profiting from the organization subsidizing a certain division (Sisilia et al., 2015).

For direct information accumulation and practicality investigation the Abdis Telkom University encouraged crowdsourced jobs to different business schools where the organization has branches. Alternate schools are in charge of gathering data about the supported SME activity. The universities can utilize learners to lead the activity and send information to Abdis for examination and report back to the organization (Sisilia et al., 2015).

STRENGTH

Live coaching is available for staff who need help and is provided by the university to those employees who need extra work time and personal experience; hence efficiency is achieved to better understand the content. Another approach to cater for efficiency is the framework minimizes the cost for the company to employ a private consultant to train its staff and monitor their SME fund. Hence, crowdsourcing provides efficiency to the company in that regard (Sisilia et al., 2015).

The position of the university to go about as a facilitator, arbiter and investigator of the assembled data gives a surety to better data and regulation of the activity. This additionally gives the organization the best experiences from leaners who at present participate in such a course using a loan. The data is sent on the web to spare the cost of discovering and going to every one of the locations by an officer from the organization; this is checked for by the university and the business colleges in the different destinations where the SME's work (Sisilia et al., 2015).

WEAKNESS

The information gathered may be intercepted and used by competitors as it moves through a large crowd including students who might be careless about its sensitivity. Another unclear part is the arrival time of the task as it is a legitimate concern for the learners and the motivators to lead such a responsibility (Sisilia et al., 2015).

2.6 Chapter summary

In summary this chapter looked at literature exploration which was done on crowdsourcing in general, crowdsourcing in education, benefits of crowdsourcing and challenges in using crowdsourcing. Literature exploration was also used in finding and studying literature for information on the existing crowdsourcing frameworks that have been used in any field to arrive at a solution.

The literature exploration used also covered studying various crowdsourcing tutoring frameworks that were specifically used in academic institutions to add value and contribute towards improvement of their results. Existing tactics and suggested approaches of addressing the problem were important when conducting the literature exploration. These tactics were analytically studied in an effort to give a crowdsourcing tutoring framework that qualifies responding to the research questions.

The work done by Chhabra et al. (2015) in "A Framework for Textbook Enhancement and Learning using Crowdsourced Annotations" was studied more as it has motivated this study. It is about the improvement of the end product in the form of a text, hence in our case the study intends to improve the pass rate among learners in higher education through crowdsourcing. Lastly other frameworks were compared in order to see how they were used in terms of their application, efficiency, and their weaknesses.

Chapter 3: Research Methodology

3.1 Introduction

In response to the research questions asked, this study implemented four methods which were: literature exploration, modelling, prototyping, and evaluation. Literature exploration was the most important step in responding to the research questions and it has been done. The literature exploration involved finding and studying literature for information on the existing crowdsourcing frameworks that have been used in any field to come up with a solution. The literature exploration was also used to study various crowdsourcing tutoring frameworks that were specifically used in academic institutions to add value and contribute towards improvement of their results.

Existing tactics and suggested approaches of addressing the problem were important when conducting the literature exploration. These tactics were analytically studied in an effort to deliver a crowdsourcing tutoring framework that qualifies in responding to the research questions. Much has been done to arrive at the newly developed workable framework that is to govern the functionality of the tutoring platform in terms of reviewed literature.

Insights from other authors were used when designing the framework including Cresswell (2016) who developed the TPACK educational platform and Chhabra et al. (2015) who used annotations and others stated in the literature. Corneli (2013) also looked at the overview of the last ten years of PlanetMath.org, which gave a good picture on how best this framework can be developed. Planet Math created a platform for a useful comprehensive mathematics encyclopedia freely available online that was filled as quickly as possible. The work done by Chhabra et al. (2015) in “A Framework for Textbook Enhancement and Learning using Crowdsourced Annotations” has motivated this study as it is about the improvement of the end product in the form of a text book; hence in this case the study intends to improve the pass rate among learners in higher education through crowdsourcing.

3.2 Sample for the study

Data was collected from schools around the Gauteng province. An average of about 400 students were chosen for this research from 10 different schools around the Gauteng province. Also an average of about 182 tutors from the 10 schools around the Gauteng province were chosen for this experiment. For this experiment only schools around Gauteng province were considered but this can be escalated to anywhere around the whole country.

They are high school students between Grade 11 and 12 and were selected randomly from any school in the Gauteng province in the Republic of South Africa regardless of their nationality as long as they are a part of the school system of South Africa.

3.3 Proposed crowdsourcing tutoring framework for mathematics

The proposed crowdsourcing tutoring framework has been developed and the framework includes components such as students, tutors, virtual class and books with each having its own interface. The components are as shown in Figure 11 in the proposed crowdsourcing tutoring framework.

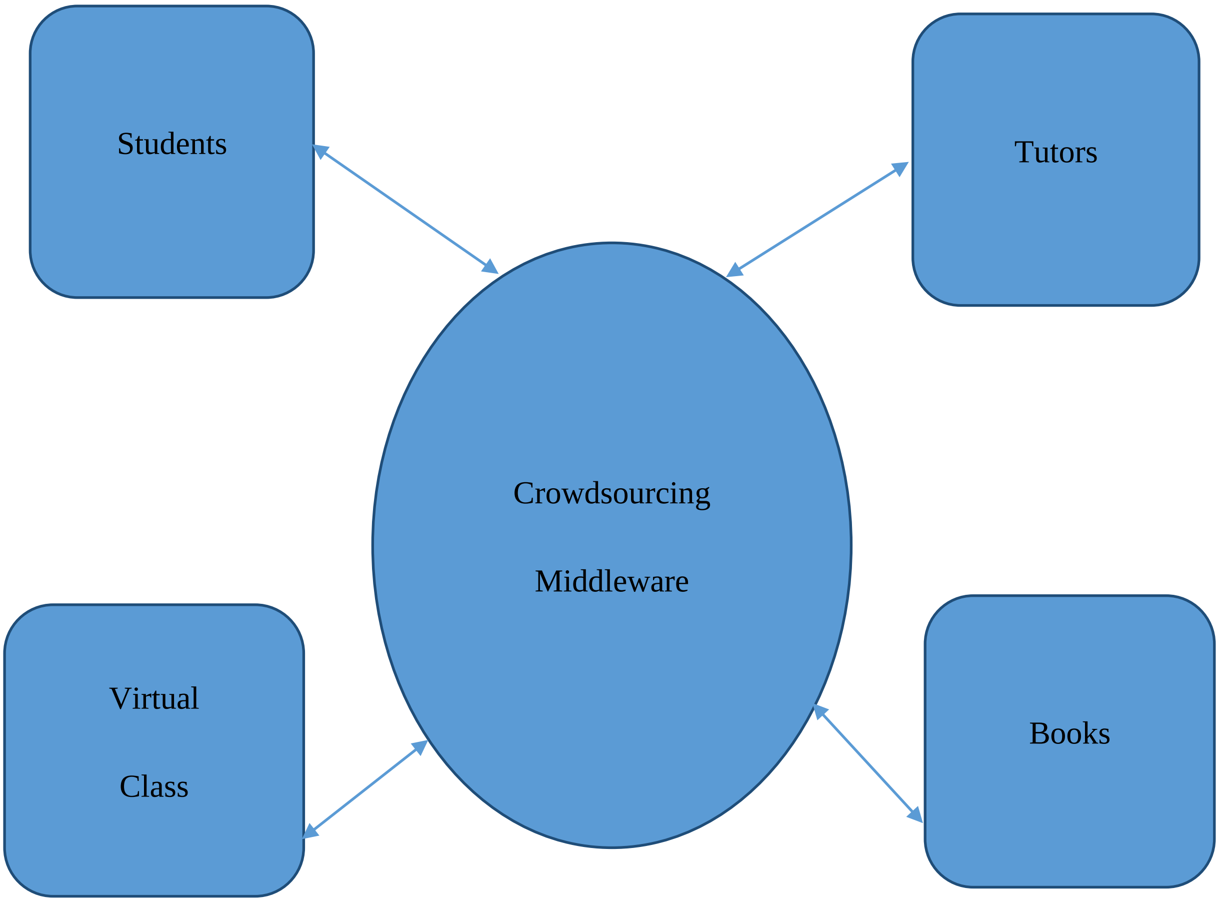

Figure 11: Proposed Crowdsourcing tutoring framework adopted from *(Tarasowa et al., 2014)*

3.4 Student selection

Students were selected from schools around the Gauteng province for this research. This method can also be applied to schools across the country. The chosen students wrote tests after each topic taught in grade 11 and 12 mathematics syllabus and their average marks were recorded.

The students who obtained low average marks and high average marks were grouped into clusters. The cluster with students that obtained low average marks were then categorized into clusters again now being classified by the topics they performed badly. These clusters constitute the virtual classes.

3.5 Tutor selection

Tutors whose students obtained the highest average marks, from the different topics covered in grade 11 and 12 mathematics, were then recommended for the virtual classes.

The tutors who emerged in the group with the highest pass rate, on a specific topic, were then allocated to the virtual class if they are interested in helping the virtual classes. These tutors brought much-needed expertise into the virtual classes.

3.6 Book selection

Books to use were mapped to tutors, when a tutor has been identified as a candidate for the virtual class then the book that the tutor was using in his/her class is also recommended for use in the virtual class.

3.7 Virtual class composition

A virtual class is made up of students, tutors and books put together.

3.8 Selection of clustering algorithm

A clustering algorithm was chosen to conduct experiments and the selection of the method was based on the literature studied where many instances were discovered and had been used for almost similar reasons. The reason this algorithm was chosen was because it is an algorithm that defined collections of items such that the items in one collection are like one another and different from the items in another collection.

Clustering can be seen as the most significant learning method and in learning data mining, clustering has been used to cluster the students according to their performance e.g. clustering can be used to differentiate active students from non-active students according to their performance in activities (Aher and Lobo, 2011b). During this study the algorithm chosen was used to differentiate students with low average marks from students with high average marks. The students with low average marks were then given attention after they were selected using this method.

The algorithms that were candidates to be used for clustering in this study are the *Hierarchical* Clustering Algorithm and *SimpleKmeans.* Upon reviewing both algorithms, it was discovered that they had commonalities although one had an advantage over the other; hence this study concluded to choose an algorithm that would correctly suit the direction of this study's experiments.

The *simpleKmeans* algorithm was chosen for this study.We learned that the most common applied clustering algorithm was the *simpleKmeans* algorithm and this algorithm had been used in many practical applications (Virmani et al., 2015). The reason this study chose *simpleKmeans* algorithm over *Hierarchical* clustering algorithms was because of its ability to process large data sets. The *simpleKmeans* often terminated at a local optimum and generated tighter clusters than *Hierarchical* clustering, especially when clusters were globular. It is a popular algorithm because of its observable speed and simplicity (Virmani et al., 2015).

Oyelade et al. (2010) used *simpleKmeans* clustering algorithm in combination of the deterministic model to study the students' results at a private Institution in Nigeria. It provided a good yardstick to carefully monitor the progress of educational performance amongst students in higher institutions, for the determination of making an operative decision by the academic architects **(**Oyelade et al., 2010)**.**

In this scenario *simpleKmeans* clustering algorithm was presented as a simple and efficient tool to monitor the progression of students 'performance in higher institutions (Oyelade et al., 2010)**.** Therefore, this clustering algorithm worked as a successful benchmark to monitor the progression of the students' performance in higher institutions, enhanced the decision making by academic planners to monitor the candidates' performance semester by semester and improving on the future academic results (Oyelade et al., 2010).

Bijuraj (2013) applied *simpleKmeans* to food items to classify those that were purchased fewer by customers versus those purchased more frequently by customers. This also helped companies deal with a variety of products who may need to know the sale of all of their products in order to establish which product was being extensively sold and which was lacking (Bijuraj, 2013). The study of automobile trajectories using the *simpleKmeans* clustering technique discovered patterns that helped the decision making and the results of the experiment showed that data mining techniques improved the decision making process and leverage the business model under the study (Zambrano and Veliz, 2016).

Abernethy (2010) applied *simpleKmeans* algorithm by answering the following questions when it came to customers since he was doing a study on the behaviors of different age group customers. He reviewed which age groups favored the silver BMW M5 and the data obtained was used to compare the age of the procurer of previous cars and the colours purchased (Abernethy, 2010). The following results were generated from this data where it could be found that certain age groups (22 to 30-year old individuals, for example) had a higher tendency to order a specific colour of BMW M5s (75 percent purchased the blue colour).

Furthermore, it showed that a different age group (55 to 62-year old individuals, for example) tend to order silver BMWs (65 percent purchase silver, 20 percent purchase grey). This generated data, after using the *simpleKmeans* algorithm, showed that when extracted would default to cluster around certain age groups and certain colours. This allowed the user to swiftly determine patterns in the data (Abernethy, 2010).

After studying the literature review on how *simpleKmeans* had been applied, this study then decided to use this algorithm in the implementation of this proposed framework because it is a clustering algorithm and works best with numeric data. The data to be classified in this study was of numerical value since it was data from students' average marks and tutors' pass rate on different topics. The *simpleKmeans* algorithm was used to classify between the students' performance and group them according to their capabilities which made it easier for this study to be able to tell who excelled in a specific topic within the mathematics syllabus.

3.9 SimpleKmeans Algorithm

SimpleKmeans is a prototype based partitioned clustering technique that attempts to find a user specified number of clusters (K). The specified number of clusters are represented by their centroids and *simpleKmeans* performs best with numerical data (Stefanowski, 2009). This is how *simpleKmeans* works:

It picks a number (K) of cluster centers that are centered randomly, assigned each item to its nearest cluster center using Euclidean distance, moved each cluster center to the main of its assigned items. The repetition occurred until convergence which brought a change in cluster assignments less than a threshold (Stefanowski, 2009).

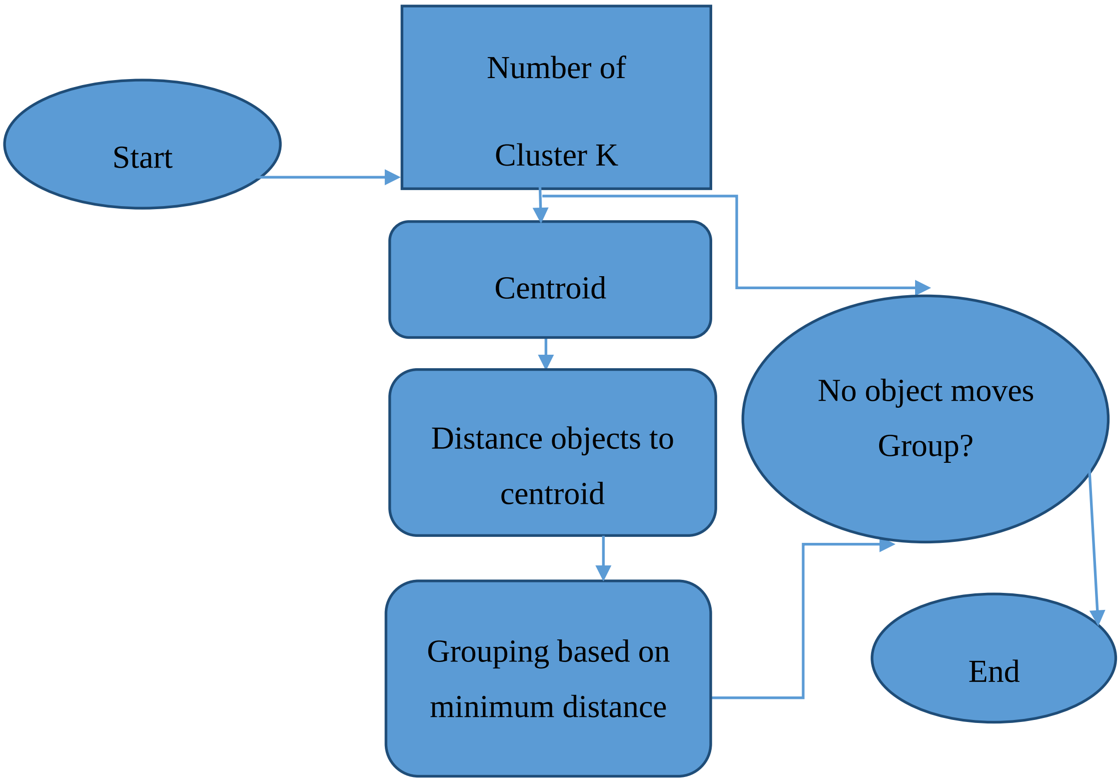

Figure 12: SimpleKmeans clustering process (Jain et al., 2010)

As shown in figure 12, *simpleKmeans* decides the centroid coordinate, distance of each object to the centroids and then group the objects based on minimum distance with which to find the closest centroid. This will happen until the merge is complete and stable with no object that move (Jain et al., 2010). The following is the generalized pseudocode of traditional *simpleKmeans* explaining on how *simpleKmeans* works and the reason why this study has chosen to use *simpleKmeans* since it has proven on many occasions to be the simplest method of grouping numerical values.

SIMPLEKMEANS PSEUDOCODE STEPS ON HOW IT PROCESSES

Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values

Step 2: Initialize the first K clusters

- Take first k instances or

- Take Random sampling of k elements

Step 3: Calculate the arithmetic means of each cluster formed in the dataset

Step 4: K-means assigns each record in the dataset to only one of the initial clusters

- Each record is assigned to the nearest cluster using a measure of distance (e.g. Euclidean distance)

Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset (Oyelade et al., 2010).

When this study reviewed the properties of this algorithm, it also discovered that it was a simple and easy way to group items especially for numerical data. There is always K clusters, at least one item in each cluster being non-hierarchical and does not overlap. Lastly, each member of a cluster is closer to its cluster than any other cluster for closeness does not always involve the center of clusters (Bijuraj, 2013).

3.10 Chapter summary

In summary this chapter introduced the methodology that was used to be able to create a framework that would be used to answer the main objective of this study. Four methods were implemented which were literature exploration, modelling, prototyping, and evaluation. Literature exploration was the most important step in responding to the research questions and it has been done. The literature exploration involved finding and studying literature for information on the existing crowdsourcing frameworks that have been used in any fields to provide a solution.

The literature exploration was also used to study various crowdsourcing tutoring frameworks that were specifically used in academic institutions to add value and contribute towards improvement of results. The proposed crowdsourcing tutoring framework has been developed and the framework comprised of components such as students, tutors, created virtual class and books. Each component had its own interface that is connected to a mini-system that would be running in the background of the framework controlled by crowdsourcing middleware.

Modelling involved developing an effective crowdsourcing mathematics tutoring prototype system for South Africa which was also used to measure the effectiveness of the prototype system developed. The method for the selection of the algorithm used for conducting experiments, selection of students, selection of tutors and selection of books for the virtual class was specified and the criteria used for formulating this method was explained. Students with low average marks on the different topics became our candidates for the created virtual class. Tutors with the highest pass rate on the different topics were the ones recommended for our created virtual class. The books that the recommended tutors were using whilst still at school, were the same books recommended for our created virtual class.

The algorithm used in the implementation of this proposed framework was introduced. *SimpleKmeans* was the algorithm chosen to be used and the reason was that this algorithm is a clustering algorithm and worked best with numerical data. A brief explanation, of how simpleKmeans works, was done. It selects a number (K) of cluster centers centered randomly, assigned each item to its nearest cluster center using Euclidean distance, moved each cluster center to the main of its assigned items and continually repeated until convergence.

Chapter 4: Experiments and Results

4.1 Introduction

This is the stage where the conducting of experiments to be used to obtain results, as outlined in our methodology, is presented. It illustrated how the proposed framework would work and how the algorithm was used to classify tutors and students. Weka was used in this study and it is an open source software issued under the GNU General Public License. Weka is a collection of machine learning algorithms for data mining tasks. It contains tools for data preparation, classification, regression, clustering, association rules mining, and visualization.

4.2 Students and access to prototype systems

The prototype system was developed where the paring of students with low average marks with a tutor who obtained the highest pass rate in a specific topic was achieved. The students accessed the platform via their school network or their gadgets which could be connected to the internet any time of the day. Once accepted into the system, each student was given a unique number to promote easy identification of a specific student.

This unique number was different from student numbers linked to their schools and national identity numbers. Whenever the student logs onto the system, time would be allocated to each virtual class. The use of a prototype system to review how far the student was from obtaining good marks when it came to mathematics. Our aim was to evaluate the level of knowledge acquired, as this study intended to improve the level of the pass rate in mathematics.

Figure 13: Students accessing our system via gadgets

The picture in Figure 13 shows students in a class environment able to access our platform through different kinds of gadgets. Students register on the platform using a form that requires the student to provide the following details as shown in Figure 14, taking into consideration that their ID numbers connect them to the school that they attend for easy verification.

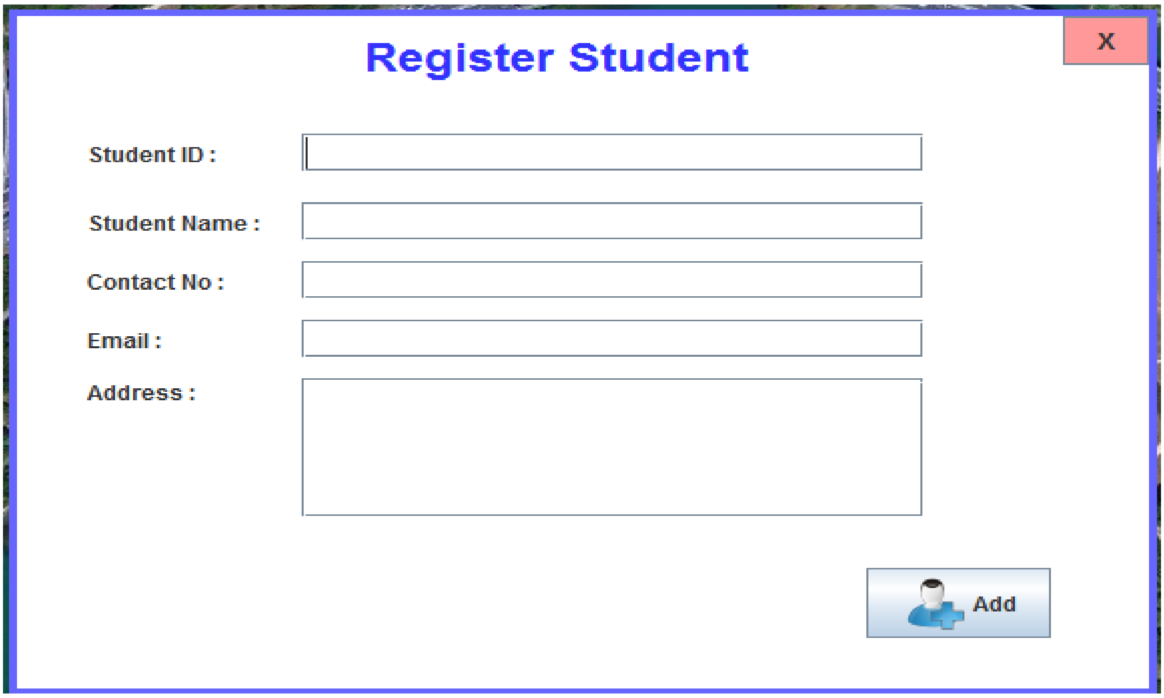

Figure 14: Form students use to register

STUDENT REGISTRATION AND AUTHENTICATION

A two-way authentication process was applicable to student registration and authentication. One being the traditional use of credential matching, where students input a login and password. Once a student was authenticated past this first level, it was possible to see their profile, marks and related metrics. Using this personalized account, our system presented the student with a list of mathematics tests that form part of our virtual class.

The student could be recommended to any of the virtual classes dependent on which topic the algorithm found the student obtained or got less than 60%. Once the student submitted their school marks and had been accepted into the system, second level authentication was activated where upon being accepted, the student knew which virtual class to attend to improve student marks.

If the details entered onto the form were incorrect the system would reject the student access as all students, in higher education in South Africa, are required to have ID numbers that connect them to their home, usually known as a physical address. Student authentication was a very significant part of our framework as it was important to distinguish each student from the public for them to be linked to their exact profile. Mysterious or unauthorized access was not allowed as it was critical to stop malicious users and avoid boot attacks (Balasubramanian and Estrada, 2016).

Once everything was verified, the student was then granted access and prompted further as to what needed to be done for the student to achieve the frameworks' aim to improve marks.

To obtain access to the virtual class, a level 2 authentication was required whereby the framework system generated a random 6-digit one-time password (OTP) to grant the execution of that specific task by the specific student. An OTP is a password that is valid for just one login session and created a strong advantage as OTPs' were temporary.

Contrary to static passwords, OTPs are not vulnerable to replay attacks and can effectively contain boot attacks (Balasubramanian and Estrada, 2016). The inclusion of OTPs, in this study, was vital as one had to ensure that unauthorized access to the virtual class, was avoided.

Students were recommended to our virtual class and once they received proof to attend, the information reflected on the virtual class time-table then the student was given the one-time access pin to that virtual class. This system of access also prevented students from providing access to unauthorized colleagues. More than often, students would simply give these colleagues their usual username and password, for easy access. This OTP allowed, only the specified student at that specified point in time.

1.	What is 2300 when written in standard form?					
		A 23×10^2	B 2.3×10^{-3}	C 2.3×10^3	D 2×10^3	E 230×10^1
2.	What is 0.00045 when written in standard form?					
		A 4.5×10^{-4}	B 4.5×10^{-5}	C 4.5×10^4	D 4×10^{-4}	E 45×10^{-5}
3.	What is 2×10^{-3} when written as a normal number?					
		A 2000	B 2003	C 0.0023	D 0.002	E 0.0002

Figure 15: Multiple choice sample questions

Figure 15 shows the multiple-choice questions that were written by the students after being recommended to virtual class as a re-evaluation process. The re-evaluation process was used to gauge the progress made to fulfill the frameworks' main aim which was to improve pupils' marks. The system also generated the timetable when the students would be re-evaluated, and the student only required their username and password to write the re-evaluation. An algorithm was used immediately after the re-evaluation test to check the progress made in relation to those students who excelled in the topic and students who had been attending virtual class to improve their marks.

The system that contains the historical test papers per topic, was found inside the crowdsourcing system and was a self-marking system. The number of tests inside the system was relative to the topics offered in the grade 11 and 12 mathematics syllabus. The test papers were loaded into the system by the system administrator, who in this case was the one in the process of implementing the proposed framework.

The question papers covered only one chapter as schools made student write a test after each chapter covered. The reason why students attempted a test after each chapter was because the system wanted to check on the chapters they were going to struggle with. The students would then concentrate on improving their performance based on that part. For example, if the student's marks on functions were lower than the benchmark set for this framework, the platform initiated an algorithm to review how much work was required in the virtual class for the student to become closer to the set benchmark. This assisted this study to indicate what would happen next with that specific student.

4.3 Experiment for algorithm selection

The experiment was done in relation to time performance analysis for *simpleKmeans* and *Hierarchical* clustering algorithms. It is performed, based on the amount of data that is considered for clustering and how much time the entire process took. This study used different data to test from both student and tutor data. In Figure 16, the x-axis indicates the amount of data and the y-axis indicates the time in seconds.

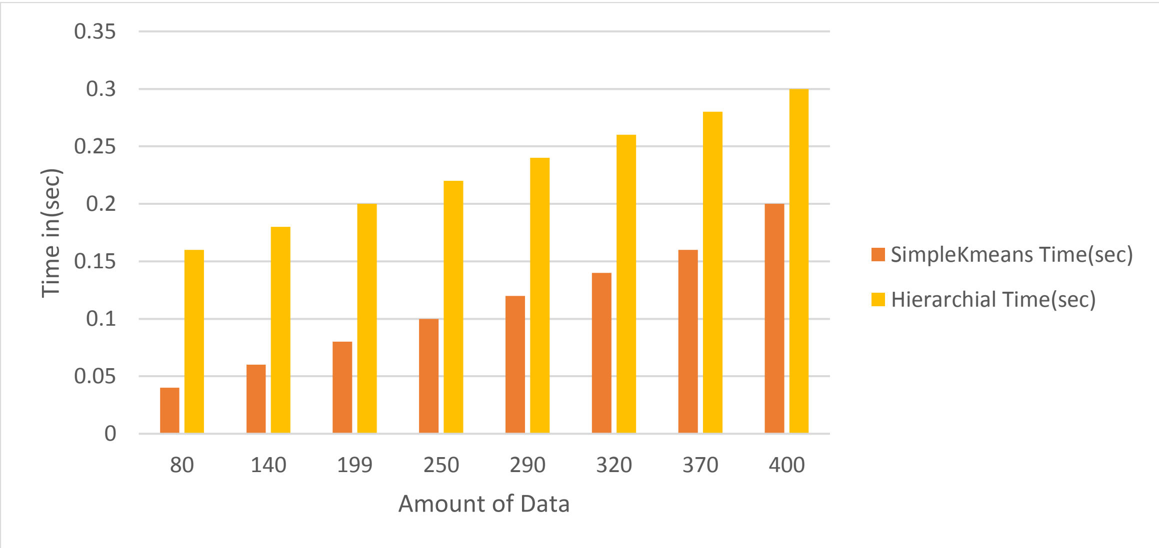

Figure 16: Comparison of performance time for SimpleKmeans and Hierarchical algorithms

Figure 16 displays various time executions using different strings of data and *Hierarchical* algorithm taking longer as compared to *SimpleKmeans* algorithm.

4.4 Experiment for student selection

This experiment was conducted using the weka tool. The data was collected from students who wrote tests for each topic covered in grade 11 or grade 12 in the Gauteng Region of South Africa.

An average of approximately 400 students participated from schools around the Gauteng province for this specific experiment. This experiment only focused on the mathematics syllabus for grade 11 and grade 12. From the list below, it is evident that a test is written after each topic. The topics are:

- T1 Probability
- T2 Statistics
- T3 Euclidean Geometry
- T4 Analytical
- T5 Differential Calculus
- T6 Polynomials
- T7 Trigonometry
- T8 Finance
- T9 Functions
- T10 Sequences and Series.

This *simpleKmeans* algorithm is within the weka tool. Weka is a collection of machine learning algorithms for data mining tasks and is a freely accessible software that could be used for data mining (Aher and Lobo, 2011a). Data Mining could be used to extract knowledge from e-learning systems such as Moodle through the analysis of the information available in the form of data generated by their users (Aher and Lobo, 2011a). Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses or other information repositories. It includes various tasks such as classification, clustering, association rule etc. (Aher and Lobo, 2011a).

SimpleKmeans classified between best performing students, moderate performing students and those that were truly having difficulties. It also helped to determine how to assist students to improve their marks in each topic covered, from what they received previously. After collecting the data from the 10 schools around Gauteng province, the first task was to allocate each student their unique number, which will assist in easy identification logging on to the system.

The loading of the data file student data.arff into the weka tool, was perfumed and this study could see if the correct data had been loaded through the pre-process tab. From this tab again, this study could view the columns, the attribute data, distribution of the columns and was able also to clean the data from the pre-process tab.

From this pre-process tab this study can choose which attributes were needed and which ones were not, for results to be meaningful and desirable. During the experiment, it was observed that the results needed refinement to deliver correct and interpretable results, since inside the weka tool there were lot of processes that provide options of what was intended with the loaded data. In our case this study chose the clustering process as it intended to group students performing superbly and those not to recommend them to our virtual class. Clustering proved be the best method to classify data and group according to similarities.

After this, the study was satisfied that our data was ready to be processed and the decision in choosing of the algorithm was done from among the other clustering algorithms, inside the weka tool. *SimpleKmeans* was chosen, as explained on the methodology, as the algorithm to be used for this study prototype systems. The data was ready as a training set to be processed to produce the expected results. The mining of the data was performed where clusters were generated in relation to student performance in each topic for grade 11 and 12 mathematics syllabus. Here clustering was done which was considered the best when using numeric data, especially the *simpleKmeans* algorithm.

Then after noticing that our data was around 400 students to be used for this clustering, this study decided to cluster our data into five clusters; considering that when this study tried to use two clusters our results did not provide the desired results. The five clusters generated proved to be what this study desired. Bear in mind that 90 rows of data, with five data clusters, may likely take more minutes or even hours of computation with a spreadsheet, whereas weka can generate the results in a few seconds.

The process was completed, and students' results were displayed, those with high average marks and those with low average marks as per tests written in different topics in the mathematics syllabus for grade 11 and 12. There are 10 topics in all covering the whole syllabus and from each topic, this study was telling how each student performed in each topic. This allowed us to be able to choose any of the topics and used it as our experiment in our virtual class. The pairing, of most suitable tutors and a student with low marks, occurred. Struggling students in probability were most likely from cluster 2 and cluster 3 and they were 142 in total.

This has given us another task of how this study could divide them into a manageable number for each virtual class, approximately 15 or 20 per virtual class. We established a reasonable number for our virtual class dependent on the number of tutors available. Students in Figure 17 were classified and some are failing certain topics as per the benchmark set for this framework. These failing students became the perfect candidates for our virtual class.

Students

Clusterer output

Attribute	Full Data	Cluster# 0	1	2	3	4
	(400)	(170)	(59)	(111)	(31)	(29)
T1 Proba	67.04	79.0588	67.2373	48.7207	44.871	90
T2 Stati	69.8175	85.1	60.5424	54.009	54.4194	76.069
T3 Geom	69.105	78.3235	86.339	51.6667	66.5484	49.4828
T4 Analyt	69.5575	83.5	82.5932	48.7477	49.7097	62.1724
T5 Calcu	68.6575	82.4059	69.8475	46.4414	86.871	51.2069
T6 Polyn	68.9025	77.7824	74.3559	49.4685	65.129	84.1724
T7 Trigon	69.84	83.8824	81.5424	50.955	49.9355	57.2759
T8 Finan	71.12	79.4235	84.9322	46.9369	88.7097	68.1034
T9 Funct	69.5125	81.7706	59.2203	52.8559	56.7097	96.0345
T10 Seque	66.6925	83.8882	70.1864	45.955	55.4516	50.1724

Figure 17: Classification of students with high average marks and low average marks from the different topics

The results in Figure 17 showed how each cluster came together in relation to what each student received from the tests marks resulting in everyone in one cluster sharing the same average marks, e.g. in probability 170 students shared the same percentage of 79% or 80% in cluster 0 which means they were similar. This made it easier for a new student to be seen if similar marks were shared with other student and to decide what could be done to help the student obtain better marks in probability or to be tested on another topic. This study concentrated on helping students who range around less than 60%. From 60% or above, those students were referred to another topic where they may obtain less or more. Each cluster presented with the performance of those students for this study to draw some conclusions as our aim was to find where a student was lacking and then recommend him or her to our virtual class in the crowdsourcing tutoring framework.

Cluster 0 indicated that these students were all doing well as they performed very well in all the tests that were offered. This indicated that these students need little assistance for them to improve their marks. The student could voluntarily opt to attend that virtual class to see whether it could help them to improve their marks from e.g. 80% to 95%. When this study reviewed at all the topics offered, it indicated a range between 77% and 84% which meant was up to the student to maintain the high marks by attending any of the virtual classes the need be to improve their marks e.g. from 77% to 88% or to 98%. In polynomials, finance and probability most of them range around 80% which was incredible although they could do better if they are willing to attend more lessons via the virtual class.

Cluster 1 indicated the students were also doing well in all the topics but there was a serious problem when it came to the functions as it indicated a 59% pass rate. This study does not have a choice but recommends them to our virtual class for functions where they will find a tutor who have an excellent passing record in functions. After they attend the virtual class they would need to be re-evaluated and see whether there was improvement or not from 59% and if there was improvement then this study would have achieved its main aim.

By reviewing the statistics, the results were approximately 60% and from the way it looked, these students do need a virtual class though it could only be up to if they were meeting the frameworks' set passing mark of 60%. These students ranged between 59% and 84%. In comparison 84% is also the highest mark in the previous cluster which is cluster 0.

Cluster 2 indicated these students were not surpassing the pass mark at all when it came to mathematics in grade 11 or 12, all ranging between 45% to almost 54%. This meant the students needed to attend the virtual class starting with probability and become re-evaluated and move on to the next virtual class until all marks improved. These students required a tutor who achieved the highest pass rate in each topic offered, to tutor these student in such a manner by the time they were re-evaluated, most of them would obtain better marks.

Cluster 3 indicated these students were having extreme problems with almost all the topics except for finance, calculus, polynomials and geometry. It was recommended that they attend the virtual classes except for finance, calculus, polynomials and geometry. Timetables were available to them to consult up until re-evaluation. In all the virtual classes there was a specialist tutor or educator because of this study having a good way of recruiting volunteering tutors.

We used an algorithm that grouped best performing tutors together, in each topic offered, for this study to be able to choose from a pool of tutors in each topic. This study relied on the fact that with the best tutor on the delivery site surely the students' passing marks would improve as well, e.g. from 44% to maybe 74%. *SimpleKmeans* algorithm made it possible for this study to recruit the best tutor or educator by being considered the best when it came to the grouping of numeric values.

Cluster 4 indicated that the group of 29 students were doing well, in almost half of the topics offered and not doing well in the other half of topics offered. Their pass rate ranged from between 49% and 96%. With the right tutor delivering to them, they would have potential for doing much better in all the topics. This study recommend that they attend the following virtual classes: geometry, calculus, trigonometry and sequences. The timetable was given to them on how to attend each class and on when re-evaluation would take place to see if progress has been achieved. They were doing very well on probability, polynomials and functions ranging between 84% and 96%. This indicated that their tutor was the most suitable on delivery of the three topics. This study was confident that dependent on the recruitment system, tutors will be delivering in one of our virtual classes.

This framework's aim was to match the best tutor for each topic with a student whose marks ranged from average to below par and display positive results in improving pass rates. On the other topics such as statistics, analytical and finance; attending the virtual classes would assist them to improve their pass rate as per each student's choice. The virtual class helped as the students were having a speciality tutor allowing attention to be focused on one student.

4.5 Experiment for tutor selection

This experiment was conducted using the same weka tool that was used for the students' experiment. The data was collected from tutors who had an interest in helping improve students' marks or students' performance when it came to their mathematics pass rate. An average of approximately 182 tutors were selected based on their interest from the same schools as the students around the Gauteng Province, for this specific experiment.

After registration onto the system, they were given unique numbers to use for identification. During their time at school, they were referred to as teachers but once they were recruited to our system, they became tutors. The tutors completed an acceptance form which required of them to provide their achievements based on their past performances e.g. pass rate.

This study focused on only providing specialist tutors on the topics being offered on the framework. The marks, from the tests written after each topic by previous students, were recorded to indicate the average marks achieved by the tutors as part of reviewing their ability to maintain an excelled record year after year.

We had approximately 10 virtual classes that covered the topics in grade 11 or 12 mathematics. Each virtual class only offered one topic and thereafter an evaluation to see if improvement was reached as this study solely concentrated on improving the pass rate of students in South African schools around the Gauteng region. The topics were as follows:

- T1 Probability
- T2 Statistics
- T3 Euclidean Geometry
- T4 Analytical
- T5 Differential Calculus
- T6 Polynomials
- T7 Trigonometry
- T8 Finance
- T9 Functions

- T10 Sequences and Series.

We used the *simpleKmeans* algorithm and this experiment was conducted using the weka tool. *SimpleKmeans* algorithm classified the tutors according to pass rate in each topic offered in the mathematics syllabus.

The process was completed, and the tutors' results were displayed as shown in Figure 18, those with best results and those with undesired results from the different topics using the weka tool. These results were displayed, as per the different topics in the mathematics syllabus for grade 11 and 12. There were 10 topics that comprised the whole syllabus and the prototype was able to classify each tutor to a topic.

Best tutors in probability were mostly from cluster 2 and they were 40 in total as shown in Figure 18. This resulted in this study having approximately 40 tutors to select from, for the probability virtual class and made it easier for the provisions of results as opposed to one virtual class to accommodate the large number of students having trouble with probability.

A reasonable number of tutors this study can allocate to a virtual class will depend on the number of students having trouble on that specific topic at that point in time. This study only assigned the best tutors, in that specific topic, for each virtual class as the aim was to recruit the most suitable tutor for each virtual class.

		Cluster#	Tutors			
Attribute	Full Data	0	1	2	3	4
	(182)	(51)	(14)	(40)	(13)	(64)
T1 Proba	67.7802	76.5098	76.9286	87.225	57.1538	48.8281
T2 Stati	70.3022	80.098	51.9286	88.2	70	55.3906
T3 Geom	69.7802	72.8824	87.6429	76.325	84	56.4219
T4 Analyt	70.4011	80.0784	72.2857	80.875	92	51.3438
T5 Calcu	69.6264	72.6275	66.7143	85.8	75	56.6719
T6 Polyn	69.6319	79.7843	87.5714	77.35	62	54.3438
T7 Trigon	70.2802	72.7255	72.6429	89.1	90	52.0469
T8 Finan	70.9505	69.5686	85.0714	88.225	80	56.3281
T9 Funct	69.6319	80.6471	65.2857	86.85	53.3077	54.3594
T10 Seque	67.0659	74.3137	70.5714	85.175	70	48.6094

Figure 18: Classification of tutors with best results from the different topics

These results shown in Figure 18 display tutors' pass rates in different topics and how each cluster came together in relation to what each tutor obtained from the average mark pass rate per test written after each topic. Thus, everyone in that cluster shared the same percentage of average marks achieved. For example, Probability 51 tutors share the same percentage of 75% or 76% in cluster 0 which meant they were similar. This resulted in it being easier for a new tutor to be seen if the tutor shares similar marks with those tutors. This also indicated to which virtual class the tutor can be assigned to help students obtain better marks in probability.

Considering these results, the 40 tutors from cluster 2 were the best when it came to teaching probability. My choice was thus to choose tutors from cluster 2 to teach in the probability virtual class. The reason was because our aim was to get the best tutor for each virtual class. As this study concentrate on helping students who range around less than 60%, this study made sure that it identified a tutor that would satisfy our main aim which was to improve the marks of students from 60% to above.

Each cluster presented the pass rate of the tutors as per their previous pass rate academic achievement. From there, this study drew more conclusions specifically in relation to finding the most suitable specialist tutor for each topic for the mathematics syllabus in grade 11 and 12 to help students through the crowdsourcing tutoring framework.

Cluster 0 indicated that these 51 tutors were all doing well as they were producing the desired marks in all the tests that were offered as per the expectations of this framework of a pass rate of 60% and above. Analytical, functions and statistics tutors could be recruited to teach in our virtual class because their passing mark was above 80%. The pass rate ranged from 69% to 81% which meant that should there be unavailability of a tutor, then a substitute could be chosen from this pool, especially from cluster 0.

What was observed from this cluster was that they were doing well, however, they weren’t exceeding others from the other clusters, such as cluster 1 and so on. This cluster provided us with an option to choose from a pool of excellent tutors or educators who were specialists in the said topic. This was a positive result for the students interested in improving their marks.

Cluster 1 indicated that the 14 tutors were also doing very well in all the topics but there was a serious problem when it came to statistics as it showed a 51% pass rate. This indicated that this study was left with no choice but to refuse these tutors to form part of our virtual class for statistics as they were below the set benchmark for this framework being 60%. The results ranged between 51% and 87% for only one topic as an undesired result. This study recommended these tutors to only teach geometry, polynomials and finance as they all range above 80% which was a very good passing rate percentage. In this framework, this study only matched the best tutor with a having trouble student thus only considering a tutor with the highest passing rate.

Cluster 2 indicated the 40 tutors were doing an excellent job in all the topics offered as they range between 76% and 89%. From this cluster, this study was spoiled for choice especially in sequences, functions, finance, trigonometry, calculus, analytical, statistics and probability as the results were all above 80%. When looking at the two topics that range around less than 80%, this study could still consider them, as more than 75% pass mark in general was considered a distinction and students who had achieved such marks should also easily accepted to institutes of higher learning. These tutors indicate precisely what type of tutors are required in the education sector and more of them are needed in this framework.

Cluster 3 indicated the 13 tutors were very good in most topics and very weak in a few topics. They ranged between 53% and 90%. This framework could make good use of them especially when it came to finance, trigonometry, calculus, analytical, geometry, and statistics. To note in this cluster, there were tutors that excelled above all the other tutors from the previous clusters and, without a doubt, the tutors were the priority to be recommended to teach in the virtual class for trigonometry and analytical. These tutors ranged above 90% on the two topics when compared with other tutors from other clusters. They weren't a large number in terms of capacity, but they were excellent in geometry and analytical. These tutors would not be recommended to teach functions and probability as they had less than 60% pass rate. On polynomials, they were far better tutors than those from the other clusters, however, and this study couldn't take any of them to teach in virtual class.

Cluster 4 indicated that these 64 tutors did not meet our criteria to form part of the tutoring framework as their results ranged from 48% to 57% which is below the set standard of 60%. By reviewing this cluster, one can draw the conclusion that students that were taught by these tutors, were bound to fail and subsequently would not be able to enroll in institutes of higher learning especially when it came to mathematics. It was either the students that are being taught that are not gifted in mathematics or there was a serious lack of delivery resources for the tutors. A conclusion drawn from this cluster indicates that the educational officials who run the education system in the Gauteng region, especially when it comes to mathematics, have a great challenge in helping empower these tutors to produce better results.

4.6 Book selection

Books that were previously used by the teachers, at school before they were recruited for our created virtual class, were used to create content sufficient enough to produce excellent results and thus recommended for this study. These were the books that were used by the tutors or educators to easily deliver to students for better understanding.

The content extracted from these books have proved to be sufficient enough to serve the purpose of this study. For this study, these books were used by the recruited tutor to serve the purpose intended by providing content that would improve the marks of the students. The content from these books were clearly defined and therefore the tutors found it easier to source from the books and provide meaningful content.

4.7 Virtual class

This was a platform where the students whom required assistance, was tutored by the right match of tutor who had the most knowledge on that topic. One platform used was that of a web platform where the specialist tutor would interact with the student in relation to the topic that the student was referring to.

Figure 19: Displays the virtual class comes into action once the students have been recommended to a suitable tutor

Figure 19 displays a picture showing the kind of topic offered in that virtual class, which in this case is probability. The reason this study was showing probability was because when analyzing our output from both the specialist tutor and the student, this study noticed that students from cluster 2 and cluster 3 were truly having trouble with the topic. The troubled students were 142 in total and this study decided to group them into a manageable group of 15 per probability virtual class. The study also had an extensive list of tutors to choose from and therefore could create the manageable groups. This study had 40 tutors to choose from and decided to allocate only three tutors per virtual class with one in the morning, another during midday and the last one allocated in the afternoon.

This large number of students having trouble in probability led to many virtual classes created to help as many students as possible to improve their marks.

In public schools, mathematics teachers are required to teach for the most part of an entire day. In this study, teaching for the entire day was discouraged to deliver the best tutoring achieved from a fresh and energetic tutor. By allocating three tutors, the tutors had enough time to rest and prepare sufficiently for their next virtual class. The 15 students were indeed referred to this virtual class where they found three tutors selected from the 40, as shown in cluster 2, in the classification of the students' results from the weka tool. When reviewing the results, it can be concluded that the 40 tutors were the best tutors when it came to teaching probability as their average marks indicated 87% or greater.

The virtual class clearly showed how all the components came together after tutors and students were classified using the weka tool. The books recommended by the system were those that the tutors were using in obtaining the best results in the probability topic. Three books were selected and recommended for this virtual class. The books used by the assigned tutor, to the probability virtual class, were mathematics for higher grade national mathematics, higher grade algebra, higher grade calculus and lastly mathematics for higher grade finance.

These books produced content that the tutor was able to use and produced excellent results, especially in probability. The study assigned the best tutors and low performing students to meet our framework's aim. Reviewing the interface in Figure 20, it can be easily interpreted that approximately 15 students from cluster 2 were not performing well in the topic probability hence this study had no choice but to send them to this virtual class. This study also had the three most successful tutors allocated. This system was made more viable in the fact that this study didn't rely on only one tutor but rather had an entire pool in cluster 2 from where one could choose.

From the 40 tutors this study decided to use only 3 tutors. If one tutor was not available, another tutor could be assigned during this time and work continued. Tutor 5, tutor 7 and tutor 9 were the three tutors allocated for this probability virtual class chosen from the pool of 40 tutors classified as the best tutors in probability.

By reviewing Figure 20, this study indicated which student attended the virtual class and if the student was required to be re-evaluated to review how much had been invested into improving the students' previous marks. Furthermore, Figure 20 indicates that the objectives of this study were fulfilled as this study was able to recommend a student with low marks to a virtual class, recommend the most suitable tutor and able to identify the books that were the most practical to be used for each virtual class.

With the books producing the right content for the virtual class probability, the best tutor allocated went and taught in that probability virtual class with a guarantee that the student having trouble with probability, would be helped and result in improvement of the students' probability tests.

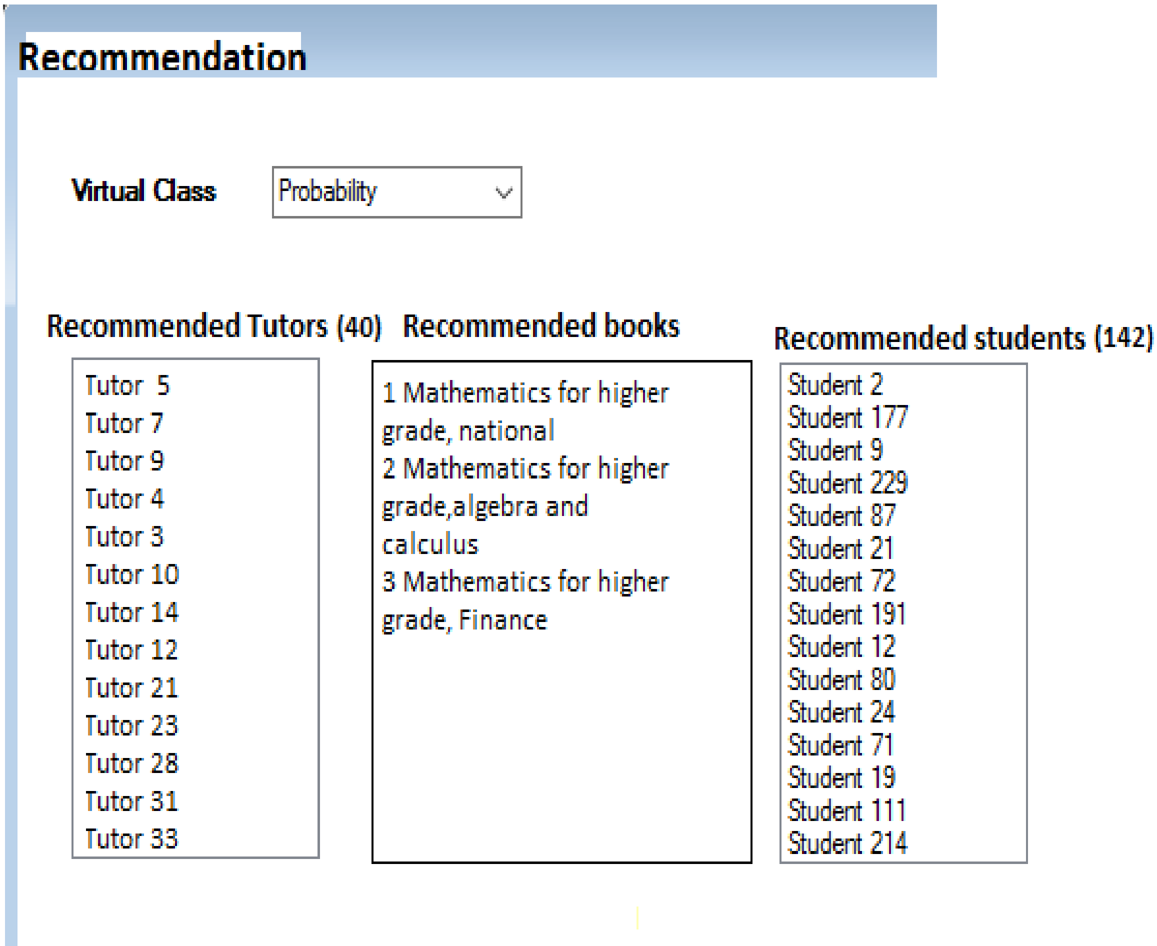

Figure 20: Results classified indicating all the components placed together

This study created the virtual class displayed in Figure 20 consisting of students, books and tutors as components for this newly developed crowdsourcing tutoring framework. This study crowdsources the volunteering tutors and retired educational experts who were willing to contribute to our education system. This crowdsourced expertise will go a long way in providing solutions to many students who might not have a clue on how to better their marks especially in relation to mathematics.

SimpleKmeans algorithm has proven to be very effective because it was able to cluster student and tutor data successfully. This study confirmed that, through the classification of the best tutor, and the classification of the student with a low average mark, it became easy to create a virtual class. It was also able to classify all students with low average marks in probability and all tutors with the highest pass rate in probability. This in a sense, tells us that it can classify with 100% accuracy.

4.8 The crowdsourcing prototype provided the following:

- The office center where all the components were monitored

- The interface that allows the students to register, so they can attend the virtual class

- It provides timetables for both the virtual class tutoring and for the re-evaluation resulting in an easier process for the student to know when he or she was tested again and for the period of the learning. Furthermore, it stated clearly which topic one had to attend and the assigned tutor by stating the starting and ending date. Time will also be specified

- Another role was to ensure the platform adheres to quality assurance

- It assigns a positive match for the tutor and student to the virtual class for those students who did not obtain 60%

- It prepares a timetable for both the tutor and the student and allocates time according to tutor availability and based on the topic to be taught

- It has collection of previous tutoring sessions, improved with annotations and links created by this study

The results from this study were generated from 10 schools around Gauteng and from each school approximately an average of 40 students were used to conduct the experiment which shows a fair representation of the Gauteng population. The purpose was to help students improve their marks in mathematics between grade 11 and 12 using the proposed crowdsourcing tutoring framework.

COMPARISON

Chhabra et al. (2015) used only 60 students to conduct his experiment where he was improving a book that would be used to improve educational content using a framework for textbook enhancement. This study used 400 students to obtain results and showed that students with a low passing rate were able to improve their marks through the virtual class successfully.

Chhabra et al. (2015) also used the leaners feedback to review the effectiveness of the framework and his results were that the collection of annotations was the right means to improve educational content. This study used a prototype to review the effectiveness of our framework as the prototype was able to classify students and tutors, in order for them to be recommended to our created virtual class. Hills (2015) used 47 students to conduct his experiment where he was enhancing content creation in the classroom to avoid giving students what they already knew but to also offer new content to save time. This study used virtual class to improve the students' marks. Students were offered a specific method of teaching mathematics, concentrating on a specific topic in relation to a student's need to enhance better understanding of the mathematics syllabus.

4.9 Chapter summary

In summary, this chapter introduced the experiment conducted as was explained in the methodology. It illustrated how the proposed framework would work and how the algorithm was used to classify tutors and students. Data was collected from schools around the Gauteng Province. An average of approximately 400 students were chosen for this experiment from 10 different schools around the Gauteng province. Also, an average of approximately 182 tutors from the 10 schools around the Gauteng province were chosen for this experiment. This chapter also briefly explained how the tutors and students would access and register on the system. The registration form was displayed and the process on how to go approximately register was explained. A graph showed the selection of the appropriate algorithm in relation to its response time against another algorithm that worked more like the one selected.

The sample for the questions to be used for re-evaluation is displayed. The experiment for the student to classify them was conducted using *simpleKmeans* which was the chosen algorithm to be used for this experiment. It classified them according to which student was having trouble in a certain topic and which student was doing well in a certain topic. The topics used for this experiment were those offered in grade 11 and 12 mathematics syllabus. This algorithm was tested using the weka tool.

The process on how the results were obtained was explained. Results indicated that the classification of students with high average marks and low average marks, from the different topics using the weka tool, were displayed and discussed. A conclusion was then derived from these results where a student whom required tutoring was recommended to our created virtual class.

The experiment to classify tutors was conducted in this chapter using a similar algorithm for the student experiment. The results from the tutor experiment showed how the classification of tutors, with best results from the different topics using the weka tool, were achieved as displayed. These results helped us to identify the best tutor and recommend the tutor to our virtual class. The virtual class interface also provided the means to illustrate how the components of the proposed crowdsourcing tutoring framework came together. The creation of the virtual class was achieved in this chapter and briefly explained why it was so important for this crowdsourcing tutoring framework.

The effectiveness of the *simpleKmeans* algorithm was tested and it was in this chapter where this study discovered that it accurately classified students and tutors and thus successfully created a virtual class. The crowdsourcing middleware was explained as the center where all the components were monitored. The books used, by the tutors for this experiment, were displayed on the virtual class interface. A comparison between authors who used crowdsourcing for the same aim as this study, was presented towards the end of this chapter.

Chapter 5: Conclusion and Recommendations.

5.1 Introduction

This chapter provides the conclusion and future work of the research. In the first chapter the main aim of the research is detailed as follows: "This research has been intended towards a mission of improving mathematics performance by designing a crowdsourcing tutoring framework to improve the pass rate in mathematics for South African schools." There was a need to evaluate whether the goal was achieved or not and to discuss future work. In this study, the aim was to design a crowdsourcing tutoring framework that will be used to improve the pass rate in mathematics for South African schools. The framework was successfully designed and developed.

5.2 Conclusion

The studying of crowdsourcing in education, through literature exploration, was conducted and proved to be of importance as an early step in responding to the research questions. The literature exploration involved finding and studying literature for information on the existing crowdsourcing frameworks that were used in any field to provide a solution, especially in education.

The literature exploration was also used to study various crowdsourcing tutoring frameworks that were specifically used in academic institutions to add value or contribute towards the improvement of results. A proposed crowdsourcing framework, for an effective mathematics tutoring system was developed and the results indicated that the main objective was achieved through this study. The framework was able to pair a student with low marks with the most suitable tutor in a specific topic, through probability, in the grade 11 and grade 12 mathematics syllabus.

This study concentrated on each topic offered and once a weakness was identified on the specific topic among the students, a method through the application of an algorithm *simpleKmeans* was used to classify that a student's average marks warrant him or her to attend a virtual class. A re-evaluation was done, after the student attended the virtual class, to establish the level of progress. This method proved to be a success as the student, prior to attending the virtual class, didn't understand that topic very well. After attending, the end results showed an improvement in performance by obtaining instructions from an expert tutor in that topic.

The recruitment of the expert tutor was actioned through an algorithm *simpleKmeans* that classified best tutors to make sure that the best tutor was allocated to the right virtual class. With the best tutor allocated to the right virtual class, the failing students was recommended to the most suitable tutor, via the virtual class, and created the expectation of an improved pass rate from students doing mathematics in South African schools.

The development of an effective crowdsourcing mathematics tutoring prototype system for South African schools for the mathematics syllabus has been done to model and to measure if it was designed to correspond to the accomplishment of mission objectives and achievement of desired results to measure the effectiveness of the proposed prototype system.

The main aim of this study was achieved as all the objectives were fulfilled in the sense that the algorithm *simpleKmeans* was able, through the weka tool, to classify the tutor with the highest pass rate and was the best tutor to be recommended for the created virtual class. Furthermore, *simpleKmeans* could classify students with low average marks that required tutoring to be recommended to our created virtual class. This study created a virtual class consisting of students, tutors and books which were the components for the developed proposed crowdsourcing tutoring framework for mathematics in grade 11 and 12 for South African schools.

5.3 Future work

The books that were used by the tutors or educators to create content, were considered as the books with the correct content and was already used for the mathematics syllabus for our framework. The tutors using these books to create educational content to be used on our virtual class have proven that the books do serve the purpose for this study as the improvement of marks for students' was achieved. The content from these books was clearly defined and the tutors thus found it easier to source from them and come up with a meaningful content.

An algorithm, as a prototype system, was used to classify between students to determine in which topic some students were weak as well as to classify between tutors to determine which tutor was the best in which topic. This enabled this study to match the tutors with our virtual class and the books proved to have the right content. Books should, also in future, be checked for similarities to create a large database as a library of books.

This study used the books that provided sufficient content to assist tutors, as a tool, to improve the marks of the students. Checking similarity within books, in future, will contribute to a large database or library of books to select from and therefore this future work deserves serious consideration. In addition to future work, perhaps another study should be conducted on how to develop an additional predictive framework that would be able to predict if the students would pass or fail, after attending the virtual class.

5.4 Chapter summary

This chapter started by looking whether this study did meet its main aim which was intended towards improving mathematics performance by designing a crowdsourcing tutoring framework, improve the pass rate in mathematics for South African schools. This chapter evaluated whether the goal was achieved or not and then discussed future work. A proposed crowdsourcing framework for effective mathematics tutoring system has been developed and results indicated that the main objective was achieved through this study. The proposed framework was able to pair a student with low marks to a suitable tutor in a specific topic (probability) in the grade 11 and grade 12 mathematics syllabus.

The study also revealed that the development of an effective crowdsourcing mathematics tutoring prototype system, for the South African school mathematics syllabus, was constructed into a model that measured the accomplishments of the mission objectives and to measure the achievement of desired results resulting in the effectiveness of the proposed prototype. This chapter presented that the main aim of this study was achieved because all objectives were fulfilled as the algorithm *simpleKmeans* was able, through the weka tool, to classify the tutor with the highest average mark, recommended the most suitable tutor for the created virtual class.

The algorithm *simpleKmeans* was able to classify students with low average marks that required tutoring to our created virtual class. This chapter tells us that this study was able also to create a virtual class consisting of students, tutors and books that were the components for the developed crowdsourcing tutoring framework for mathematics in grade 11 and 12 for South African schools. The books that were used by the tutors or educators to create content were also being used for the mathematics syllabus for our framework and considered as books with the right content. As discussed, a future work would be to carefully consider creating a library or database, where applicable books, could be added as a source of quality content. Furthermore, on future work, a proposal to search for a method on how one can create a predictive framework to predict if students would pass or fail after attending the virtual class.

REFERENCES

ABERNETHY, M. 2010. Data mining with WEKA, Part 2: Classification and clustering. *IBM Dev*.

AHER, S. B. & LOBO, L. Data mining in educational system using weka. International Conference on Emerging Technology Trends (ICETT), 2011a. 20-25.

AHER, S. B. & LOBO, L. 2011b. A Framework for Recommendation of courses in E-learning System. *International Journal of Computer Applications,* 35**,** 21-28.

AMROLLAHI & ALIREZA 2016. A process model for crowdsourcing: Insights from the literature on implementation. *arxiv preprint arxiv: 1605. 04695*.

ANDERSON, M. 2011. Crowdsourcing Higher Education: A Design Proposal for Distributed Learning. *MERLOT Journal of Online Learning and Teaching,* 7**,** 576-590.

ASMOLOV, G. Crowdsourcing as an Activity System: Online Platforms as Mediating Artifacts. Sintelnet WG5 workshop on crowd intelligence: Foundations, methods and practices, 2014.

BAILARD, C. S. & LIVINGSTON, S. 2014. Crowdsourcing accountability in a Nigerian election. *Journal of Information Technology & Politics,* 11**,** 349-367.

BALASUBRAMANIAN, T. R. & ESTRADA, T. Crowdlearning: A framework for collaborative and personalized learning. Frontiers in Education Conference (FIE), 2016 IEEE, 2016. IEEE, 1-9.

BARBIER, G., ZAFARANI, R., GAO, H., FUNG, G. & LIU, H. 2012. Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory,* 18**,** 257-279.

BERGVALL-KÅREBORN, B., ERIKSSON, C. I., STÅHLBRÖST, A. & SVENSSON, J. A milieu for innovation: defining living labs. ISPIM Innovation Symposium: 06/12/2009-09/12/2009, 2009.

BERGVALL-KÅREBORN, B. & HOWCROFT, D. The Apple business model: Crowdsourcing mobile applications. Accounting Forum, 2013. Elsevier, 280-289.

BERNSTEIN, M. S., KARGER, D. R., MILLER, R. C. & BRANDT, J. 2012. Analytic methods for optimizing realtime crowdsourcing. *arXiv preprint arXiv:1204.2995*.

BIJURAJ, L. Clustering and its Applications. Proceedings of National Conference on New Horizons in IT-NCNHIT, 2013. 169.

BOTT, M. & YOUNG, G. 2012. The Role of Crowdsourcing for Better Governance in International Development. *PRAXIS: The Fletcher Journal of Human Security,* 27**,** 47-70.

BOUDREAU, K. J. & JEPPESEN, L. B. 2015. Unpaid crowd complementors: The platform network effect mirage. *Strategic Management Journal,* 36**,** 1761-1777.

BUECHELER, T., SIEG, J. H., FÜCHSLIN, R. M. & PFEIFER, R. Crowdsourcing, Open Innovation and Collective Intelligence in the Scientific Method-A Research Agenda and Operational Framework. ALIFE, 2010. 679-686.

CHATZIMILIOUDIS, G., KONSTANTINIDIS, A., LAOUDIAS, C. & ZEINALIPOUR-YAZTI, D. 2012. Crowdsourcing with smartphones. *IEEE Internet Computing,* 16**,** 36-44.

CHERNER, T., DIX, J. & LEE, C. 2014. Cleaning up that mess: A framework for classifying educational apps. *Contemporary Issues in Technology and Teacher Education,* 14**,** 158-193.

CHHABRA, A., IYENGAR1, S. R. S., SAINI2, P. & SHREEDHAR, R. A Framework for Textbook Enhancement and Learning using Crowdsourced Annotations. *Dept of Computer Science and Engineering India***,** 1-11.

CHIU, C.-M., LIANG, T.-P. & TURBAN, E. 2014. What can crowdsourcing do for decision support? *Decision Support Systems,* 65**,** 40-49.

CORNELI, J. 2013. *Peer supported problem solving and mathematical knowledge.* PhD thesis, The Open University.

COVE, S. 2007. What does crowdsourcing really mean. Wired News.

CRESSWELL, S. 2016. Harvard-business-school-digital-initiative. *Harvard-business-school-digital-initiative* [Online]. Available: https://medium.com/harvard-business-school-digital-initiative/the-biggest-challenge-to-the-future-of-crowdsourcing-in-business-fada67196d34 [Accessed September 8].

DALVI, N., DASGUPTA, A., KUMAR, R. & RASTOGI, V. Aggregating crowdsourced binary ratings. Proceedings of the 22nd international conference on World Wide Web, 2013. ACM, 285-294.

DHIYA, A.-J., ABIR, H., MOHAMMED, A., CHELSEA, D. & JAN, L. 2015. Educational crowdsourcing to support of computer programing. *Research and practice in technology enhanced learning,* 10**,** 1.

EGELMAN, S., KANNAVARA, R. & CHOW, R. Is this thing on?: Crowdsourcing privacy indicators for ubiquitous sensing platforms. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015. ACM, 1669-1678.

ELLERO, A., FERRETTI, P. & FURLANETTO, G. 2015. Realtime crowdsourcing with payment of idle workers in the Retainer Model. *Procedia Economics and Finance,* 32**,** 20-26.

ESTELLÉS-AROLAS, E. & GONZÁLEZ-LADRÓN-DE-GUEVARA, F. 2012. Towards an integrated crowdsourcing definition. *Journal of Information science,* 38**,** 189-200.

ESTEVES, T., LEUENBERGER, D. & VAN LEUVEN, N. 2012. Reaching Citizen 2.0: How Government Uses Social Media to Send Public Messages during Times of Calm and Times of Crisis. *Citizen 2.0: Public and Governmental Interaction through Web 2.0 Technologies.* IGI Global.

FARASAT, A., NIKOLAEV, A., MILLER, S. & GOPALSAMY, R. CROWDLEARNING: Towards Collaborative Problem-Posing at Scale. Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, 2017. ACM, 221-224.

GIUDICE, K. D. 2010. Crowdsourcing credibility: The impact of audience feedback on Web page credibility. *Proceedings of the Association for Information Science and Technology,* 47**,** 1-9.

GUPTA, D. K. & SHARMA, V. 2013. Exploring crowdsourcing: a Viable solution towards achieving rapid and qualitative tasks. *library hi tech news,* 30**,** 14-20.

HATTEM, G., INDURKAR, D., ZHANG, N., EASTEP, J., LI, T., ZHANG, J. & SHAO, W. 2003. Adaptive Learning Class. *The zombies***,** 1-12.

HETMANK, L. Towards A Semantic Standard For Enterprise Crowdsourcing-A Scenario-Based Evaluation Of A Conceptual Prototype. ECIS, 2013. 118.

HILLS, T. T. 2015. Crowdsourcing content creation in the classroom. *Journal of Computing in Higher Education,* 27**,** 47-67.

HOFSTETTER, R., DAHL, D. W., ARYOBSEI, S. & HERRMANN, A. 2017. Creativity in Open Innovation Contests: How Seeing Others' Ideas Can Harm or Help Your Creative Performance.

HOFSTETTER, R., NAIR, H. & MISRA, S. 2018. A Copycat Penalty: Micro Evidence From an Online Crowdsourcing Platform.

HÖRLER, R. 2014. Crowdsourcing In The Humanitarian Network – An Analysis Of The Literature.

HYVÄRINEN, J. & VOS, M. 2015. Developing a Conceptual Framework for Investigating Communication Supporting Community Resilience. *Societies* 5**,** 583–597.

JAIN, S., AALAM, M. A. & DOJA, M. K-means clustering using weka interface. Proceedings of the 4th National Conference, 2010. 25-26.

KAUR, M. & KAUR, U. 2013. Comparison between k-means and hierarchical algorithm using query redirection. *International Journal of Advanced Research in Computer Science and Software Engineering,* 3.

KEATINGA, M. & FURBERG, R. D. A Methodological Framework for Crowdsourcing in Research. Proc. of the Federal Committee on Statistical Methodology (FCSM) Research Conference, 2013.

KEVIN, B. & KARIM, L. 2013. Using crowd as an innovation partner. *Harvard business review,* 91**,** 60-90.

KITTUR, A., NICKERSON, J. V., BERNSTEIN, M. S., GERBER, E. M., SHAW, A., ZIMMERMAN, J., LEASE, M. & HORTON, J. J. The Future of Crowd Work. proceedings of the 2013 conference on computer supported work, 2013 New York, USA. 1301-1318.

KRIVOSHEEV, E., CASATI, F., CAFORIO, V. & BENATALLAH, B. 2017. Crowdsourcing Paper Screening in Systematic Literature Reviews. *arXiv preprint arXiv:1709.05168*.

LINKRUUS, K., NILSSON, K. & WESTERBERG, A. 2012. Crowdsourcing-Take on Goliath: Motivating people to participate in Crowdsourcing.

LLORENTE, R., MORANT, M. & GARRIGOS-SIMON, F. J. Crowdlearning, Applying Social Collaboration in Higher Education. Conference Proceedings. New Perspectives in Science Education, 2015. libreriauniversitaria. it Edizioni, 256.

MAHESHWARI, D. & JANSSEN, M. 2014. Reconceptualizing measuring, benchmarking for improving interoperability in smart ecosystems: The effect of ubiquitous data and crowdsourcing. *Government Information Quarterly,* 31**,** S84-S92.

MAO, K., CAPRA, L., HARMAN, M. & JIA, Y. 2015. A survey of the use of crowdsourcing in software engineering. *RN,* 15.

MCCARTHY, J. & OLIPHANT, R. 2013. MATHEMATICS OUTCOMES IN SOUTH AFRICAN SCHOOLS What are the facts? What should be done? *CDE Insight*.

MCINTOSH, J. C. 2015. Assessing Social Media Use in Entrepreneurship Education: The Role of Crowdsourcing in Advancing Innovation Quality and Student Satisfaction.

OLIVER, U., OLIVER, P., DANIEL, S. & KAI, F. 2015. Internal crowdsourcing: Conceptual framework, Structured review and research agenda. *Forthcoming journal article, working paper version*.

OYELADE, O., OLADIPUPO, O. & OBAGBUWA, I. 2010. Application of k Means Clustering algorithm for prediction of Students Academic Performance. *arXiv preprint arXiv:1002.2425*.

PAPADOPOULOU, C.-A. & GIAOUTZI, M. 2014. Crowdsourcing as a tool for knowledge acquisition in spatial planning. *Future internet,* 6**,** 109-125.

RANI[1], Y. & ROHIL, H. 2013. A study of hierarchical clustering algorithm. *ter S & on Te SIT-2***,** 113.

SALMINEN, J. 2015. The role of collective intelligence in crowdsourcing innovation. *Acta Universitatis Lappeenrantaensis*.

SAXTON, G. D., OH, O. & KISHORE, R. 2010. Rules of Crowdsourcing: Models, Issues, and Systems of Control. *Information systems management,* 30**,** 2-20.

SINGLA, A., BOGUNOVIC, I., BART´OK, G. A., KARBASI, A. & KRAUSE, A. On Actively Teaching the Crowd to Classify. NIPS workshop on data driven education, 2013.

SISILIA, K., PERANGINANGIN, Y., SETYORINI, R. & MOELIONO, N. 2015. A Framework of Affiliation Partnership between Univesity, SMEs, and Business Sector: A Case Study of PKBL Telkom, Indonesia. *Procedia-Social and Behavioral Sciences,* 169**,** 2-12.

SKARZAUSKAITE, M. 2012. The application of crowdsourcing in educational activities. *SOCIALines TECHNOLOGIES,* 2**,** 67-76.

SOLEMON, B., ARIFFIN, I., DIN, M. M. & ANWAR, R. M. 2013. A review of the uses of crowdsourcing in higher education. *International Journal of Asian Social Science,* 3**,** 2066-2073.

SPAULL, N. 2013. South Africa's Education Crisis: The quality of education in South Africa 1994-2011. *Report commissioned by CDE***,** 1-65.

STARBIRD, K. 2012. *CROWDWORK, CRISIS AND CONVERGENCE: HOW THE CONNECTED CROWD ORGANIZES INFORMATION DURING MASS DISRUPTION EVENTS.* university of Colorado.

STEFANOWSKI, J. 2009. Data Mining-Clustering. *Institute of Computing Sciences Poznan University of Technology, Poznan, Poland, Lecture,* 7.

SULLIVAN, J. 2016. *Terrific, You Did It!: The Influence of Feedback on a Crowdsourcing Participant's Satisfaction.* Pennsylvania State University.

SVINICKI, M. D. 2010. A Guidebook On Conceptual Frameworks For Research In Engineering Education. *Rigorous Research in Engineering Education NSF DUE-0341127, DUE-0817461*.

TAN, E., XIA, H., JI, C., JOSHI, R. V. & HUANG, Y. 2015. Designing a Mobile Crowdsourcing System for Campus Safety. *iConference 2015 Proceedings*.

TARASOWA, D., AUER, S., KHALILI, A. & UNBEHAUEN, J. 2014. Crowd-sourcing (semantically) structured multilingual educational content (CoSMEC). *Open Praxis,* 6**,** 159-170.

THEBAULT-SPIEKER, J., TERVEEN, L. G. & HECHT, B. Avoiding the south side and the suburbs: The geography of mobile crowdsourcing markets. Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015. ACM, 265-275.

VIRMANI, D., TANEJA, S. & MALHOTRA, G. 2015. Normalization based K means Clustering Algorithm. *arXiv preprint arXiv:1503.00900*.

VISSER, M., JUAN, A. & FEZA, N. 2015. Home and school resources as predictors of mathematics performance in South Africa. *South African Journal of Education,* 35**,** 1-10.

VORNANEN, V.-J., SIVULA, A. & TAKALA, J. 2016. Hybrid Management in Preparedness: Utilizing Cooperation and Crowdsourcing to Create Joint Performance in the Logistic Society. *Management (18544223),* 11.

WANG, A., HOANG, C. D. V. & KAN, M.-Y. 2013. Perspectives on crowdsourcing annotations for natural language processing. *Language resources and evaluation,* 47**,** 9-31.

WAY, K. A., OTTENBACHER, M. C. & HARRINGTON, R. J. 2011. Is crowdsourcing useful for enhancing innovation and learning outcomes in culinary and hospitality education? *Journal of culinary science & technology,* 9**,** 261-281.

WELD, D. S., ADAR, E., CHILTON, L., HOFFMANN, R., HORVITZ, E., KOCH, M., LANDAY, J., LIN, C. H. & MAUSAM, M. Personalized Online Education—A Crowdsourcing Challenge. Workshops at the twenty-sixth AAAI conference on artificial intelligence, 2012. 1-31.

WHITLA, P. 2009. Crowdsourcing and its application in marketing activities. *Contemporary Management Research,* 5.

WILLIAMS, J. J., WHITEHILL, J., MITROS, P., KRAUSE, M., REICH, J., HEFFERNAN, N., PARITOSH, P., KIM, J. & KEEGAN, B. C. Connecting Collaborative & Crowd Work with Online Education. Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing, 2015. ACM, 313-318.

YANNIS, C., EURIPIDIS, N. L., AGGELIKI, A., VANGELIS, K. & ANNA, T. 2014. Passive crowdsourcing in government using social media. *Transforming government:people, process and policy,* 8**,** 283-308.

YUEN, M.-C., KING, I. & LEUNG, K.-S. A survey of crowdsourcing systems. Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, 2011. IEEE, 766-773.

YUEN, M.-C., KING, I. & LEUNG, K.-S. 2015. Taskrec: A task recommendation framework in crowdsourcing systems. *Neural Processing Letters,* 41**,** 223-238.

ZAMBRANO, G. R. & VELIZ, R. N. H. 2016. Vehicular Swept Path Analysis Using K-Means Algorithm. *International Journal of Applied Engineering Research,* 11**,** 9536-9538.

ZHAO, Y. & ZHU, Q. 2014. Evaluation on crowdsourcing research: Current status and future direction. *Information Systems Frontiers,* 16**,** 417-434.