EFFECTIVENESS OF FORMATIVE ASSESSMENT IN IMPROVING LEARNING QUALITY: *USING INTERACTIVE SCIENCE QUIZ TO ENHANCE STUDENT LEARNING OUTCOMES AT PRIMARY LEVEL*

Rida Imran

NUST201464141MSEECS61414F

Supervisor

Ms. Farzana Ahmad

A thesis submitted in partial fulfillment of the requirements for the degree of Masters of Science in Innovative Technologies in Education (MS ITE)

In

School of Electrical Engineering and Computer Science,

National University of Sciences and Technology (NUST),

Islamabad, Pakistan.

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by **Ms. Rida Imran**, **(**Registration No. **NUST201464141MSEECS61414F),** of **SEECS** has been vetted by undersigned, found complete in all respects as per NUST Statuses/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor: _Farzana Ahmad _____

Date: _____

Signature (HOD): _____

Date: _____

Signature Dean/Principal _____

Date: _____

Approval

It is certified that the contents and form of the thesis entitled "**Effectiveness of Formative Assessment in Improving Learning Quality: Using Interactive Science Quiz to Enhance Student Learning Outcomes at Primary Level**" submitted by **Rida Imran** have been found satisfactory for the requirement of the degree.

Advisor: Farzana Ahmad

Signature: _____

Date: _____

Committee Member 1: Dr. Asad Anwar Butt

Signature_____

Date: _____

Committee Member 2: Dr. Muddassir Malik

Signature _____

Date: _____

Committee Member 3: Ms.Erum Afzal.

Signature _____

Date: _____

Dedication

To my parents and brother for making me who I am today and supporting me all the way.

Certificate of Originality

I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by any other person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis.

I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged.

Author Name: Rida Imran

Signature: _____

Acknowledgement

This journey would not have been possible without my teachers, friends and family. I express my gratitude to my supervisor Ma'am Farzana Ahmad, who has supported me with patience and knowledge and provided me with her generous guidance and encouragement throughout the process of research and thesis writing. One could not simply wish for a better and friendly supervisor. I must be thankful to Dr.Asad Anwar Butt and Dr. Mudassir Malik for furthering my thinking. I would like to thank Erum Afzal, Saba Khateeb and Umer Nasim for being supportive and providing me significant suggestions throughout the application development. This accomplishment could not have been possible without their guidance and assistance.

I am grateful to my friends for listening, offering me sincere advices and supporting me through this entire process. Special thanks to Maria Tauqir, Ramsha Asrar and Mariam Sanaullah who never stopped helping me in developing ideas.

Finally, I would like to thank my parents and brother for their unconditional love and for providing me with unfailing support and continuous encouragement. They always believe in me and back me up with complete support. They have given me the hope and motivation to finish my MS.

Contents

Abbreviations

NEAS	National Education Assessment System
EFA	Education for all
NAT	National Achievement Test
ICT	Information and Communication Technologies
OECD	Organization for economic cooperation and development
CAA	Computer Aided Assessment
CBA	Computer Based Assessment
CBAAM	Computer Based Assessment Acceptance Model
UN	United Nations
SPSS	Statistical Package for the Social Sciences

List of Figures

List of Tables

Abstract

The main purpose of this study was to determine if the use of formative assessment had an impact on learning outcomes of grade four students, in science subject. In this study the research was conducted using an interactive science quiz application, developed in accordance with the latest Pakistan National Curriculum for General Science. The purpose of developing this formative assessment quiz application was to move students' learning forward by providing timely feedback in the form of additional information and to engage them with the content by adding game elements. The game elements used in this quiz application are story, characters, sounds, feedback, rewards and scores.

This research was carried out on fourth grade students in two federal government schools, for boys and girls, located in Islamabad. An experimental research design was utilized to gather quantitative data through the use of pre and post-test. Participants (N=120) were divided into two groups: control and experiment group. The students (N=60) in experimental group used a formative assessment interactive science quiz application for a week. The pre-test and post-test scores of control and experiment group were compared to determine if the use of formative assessment had a statistically significant impact on students' performance. In order to compare the scores within control group the pre-test/post-test scores of control group were compared. While the pre-test/post-test scores of experiment group were compared to determine the effectiveness of formative assessment quiz application. Moreover, to determine the impact of quiz application on gender differences in scores, pre and post-test of the experimental group were compared.

Results of this study reveal that the students (N=60), including boys and girls, in the control group did not show a significant change in scores. Their scores before and after the intervention remained same or decreased. The results from the experimental group (N=60), including boys and girls showed a statistically significant increase in scores after using the formative quiz application. This exhibits that descriptive feedback and game elements enhanced student outcome and developed their understanding. Students were found to be more engaged and attentive during the learning process. Additional research to determine the effectiveness of formative assessment at primary level, by making interactive quizzes for all the units of science textbook can make this research more valuable.

Chapter 1: Introduction

1.1. Introduction

Education has been an integral part of every society in the world and its due importance in shaping the present and the future progression of nation's calls for continuous improvement. Initiatives are being undertaken globally towards its betterment to ensure a steady evolution in student learning and pedagogy. Researchers and academics have been conducting evaluations to assess the educational process. Assessments using empirical data collection are conducted to understand student learning, teaching methodology and curriculum programs with an aim to augment them further (Allen, 2004). Assessments conducted at state level range from kindergarten to universities are summative in nature to audit the success of institutions and initiate reforms (Ruland, 2011). Summative assessment's aggregative nature is helpful in determining the strengths and weaknesses in the education system and gives an average of how a particular school, institute or a region is adhesive to an anticipated educational standard that is set aside by the educationist and evaluators. Such a collective assessment provides information about attainment in relation to the performance goals but lacks the characteristics of entailing student's learning and pedagogy. Consequently, the policies driven and measures initiated to improve the educational experience being generic are not effective as they do not cater for the diverse educational needs of different students. In contrast, assessment serves a formative function when evidence about student achievement is elicited, interpreted and used by teachers to make decisions that result in better instructional set (Wiliam P. B., 1998). Formative assessment involves an interactive learning environment that encourages discussion, self-assessment and peer review. All these activities help student develop a thought process and guide them towards quality learning.

1.2. Problem Statement

In Pakistan education sector is striving for catering the learning needs of students and teachers alike and meeting the minimum standards that are widely recognized by international community (Pakistan, 2014). Students' learning is not being positively impacted by internal examination and assessment system (Khattak, 2012). Students in our schools are forced to rote memorize in order

to get good grades, without being provided with an effective feedback. The feedback provided by teachers are in binary (yes or no) form or as ‘good’, ‘bad’, ‘fair’, ‘poor’ and ‘needs improvement’, without suggesting the ways by which they can improve their learning (Hafiz Muhammad Iqbal, 2009). All important self assessment activities and questions are missing (Sarwar et al, 2013).

Summative assessment done by National Education Assessment System (NEAS) for primary education in science subject reveals an overall poor performance with only 21% students crossing the mean score. The same report indicates students below mean for English reading and English writing. The most alarming aspect of report is that 98% of the teachers reported students’ difficulty in lesson during instructions, which raises serious questions regarding the education process that is failing to develop the needed understanding in the relevant subjects and not elevating the levels of student success.

One of the major contributing factors is the summative exam centric education system where emphasis is not on student learning and skill set development, rather the core focus is on getting marks, pass the predefined tests and get certified. Learning in public schools is even worse as it is mostly curriculum based and students are totally dependent on teachers. Exam questions and format is repeated, rote learning is widely practiced and it is difficult for the students to do basic comprehension and develop deep understanding (Rehmani, 2003). As a consequence, the students flounder when they have to address new problems using the concepts they should have learned during the education process. The exams conducted are based on 12-month academic year and often require student to reproduce what they memorized many months ago. Then the exam results are used to acclaim the performers while they can only discourage students that have not succeeded due to the lack of feedback that can help them improve and overcome their shortcomings. This clearly shows that feedback quality is not good enough to strengthen the education process and make it successful. In order to improve this situation steps are needed to reform the examination system such that it supports student learning, provides timely feedback and encourages group assignments rather than competition among students.

Teaching techniques used in Pakistan’s educational institutions are another factor that contributes to the poor academic results and deficiencies in student’s educational development. Teacher student discussions and group discussions among students are rarely employed to

enhance the brainstorming skills of students; instead the students are forced to grasp what they are told and asked to reproduce in the exams. Classroom instructions lack the needed examples, relative comparisons and multiple aspects of what is being taught. Students have different habits and backgrounds, which requires different teaching strategies to keep them engaged in the classroom to ensure active participation and deep learning. Often it is seen that teachers engage the students that are performers and ignore the ones that lag. This attitude results in demotivation and aggravates further disengagement of already poor performing students. Such students then avoid teachers or schools that instigate demoralization, which results in casual absence from the schools and even permanent detachment. Teaching techniques that stimulate and uphold students' interest have shown considerable improvement in understanding of underlying concepts (Aziz, 2003).

Our social culture being authoritative has its own implications on the lives of students, who once discouraged to ask questions, never try again. Such classroom instructions are not helping to nurture the brain of students and answering the needed concepts, which afterwards become ambiguities. This limits their thoughts, creativity and innovation capabilities, which otherwise when cherished can do wonders. Such arrogant classroom environment does not prompt in-depth knowledge and thorough understanding of course work, which is clearly visible from the results published by the NAT 2014 report.

Not all the students are alike in their educational needs, some are quick learners and others need more time to grasp the same material. The culture prevailing in our classrooms does not take into account the diverse needs of different students and it treats all of them alike, which results in inequity in student outcomes. It is therefore important to introduce teaching techniques that encourage thoughtfulness and student learning impacts are considerably improved.

Curriculum design is an additional factor that plays an imperious role in shaping the cognitive ability of students. Students studying in the public sector schools in Pakistan mainly use the textbook manuscripts from Provincial Textbook Boards that lack expertise in curriculum development (UNESCO, 1998). The textbooks are fixed course outlines and they lack examples, quizzes and group tasks, thus not cultivating a classroom culture of thought and collaboration. Due to this, students remain oblivious to life outside the classroom as there exists a complete disconnection between books and practical knowledge. Curriculum instead of being used as a

guideline is perceived as sole content that can fulfill the diverse learning needs of students. Can we assume that we have prepared the students for present and the future practical skills after they have gone through such learning experience? A comparison of Pakistan's primary education curriculum with Oxford curriculum, which prevails in private educational institutions in the country, can vividly show the shortcomings and a clear distinction in approach, confidence and attainment between learners of two educational systems. Therefore, it is imminent to introduce syllabus that takes into account learner's needs, learner's environmental experience and be objective.

1.3. Research Question

The purpose of this study was to determine the effectiveness of formative assessment on learning quality, using interactive science quiz application. This study sought to answer this research question:

What is the significance of formative assessment in improving learning quality?

Null Hypothesis One

H0: There is no significant impact of science quiz application on student learning outcomes using interactive science quiz application.

Alternative Hypothesis One

H1: There is a significant impact of science quiz application on student learning outcomes using interactive science quiz application.

1.4. Summary

If we summarize above, it is clear that summative assessment, traditional teaching techniques and imperious curriculum are not doing well to improve student learning. Students need timely feedback, adapted teaching assistance and manuscript guidelines that foster academic achievement and improve learning quality. Among the available assessment procedures, formative assessment is known to cater such learning needs of students and has shown improvements in student's attentiveness and involvement in the learning process (Ruland, 2011). Qualitative and quantitative researches conducted on formative assessment have established it as

an important intervention for promoting high performance and high equity. It builds student's learning to learn skills by placing emphasis on student involvement in the process of teaching and learning, self-efficacy and peer and self-assessment skills (Development, 2008). Formative assessment ensures teachers immediate individual feedback that is currently not possible in large scale exams (summative) employed in primary schools once a year. In this way the two most important stakeholders of assessment can use the information to adapt the learning process to student needs and impact the learning outcome. Assessments that inform students and teachers about an individual's standing on the learning curve during the learning process have a positive impact on student learning (Reeves, 2007). Formative assessment also incorporates collaboration and interactive sessions where students can indulge in dialogue with teachers and fellow peers, reinforcing analytical thinking that will help them apply concepts in larger context. Additionally, this can help students learn things that are outside the predefined lines of their curriculum, in this way overcoming the curriculum's shortcomings.

In this study we add on to the stock of research on formative assessment by investigating its efficacy on student learning outcomes at primary level, using an interactive science quiz. Since, information and communication technologies have never been so enriched in our daily lives then they are today. This has paved the way for technological innovations in education and research interests have grown in finding the potentials of using technology for the betterment of student learning. In the context of formative assessment, researcher and academics use the term formative e-assessment for formative assessments where technology is employed during assessment. Continuing in this direction, in our study we have developed the interactive science quiz using digital game development.

Chapter 2: Literature Review

2.1. Introduction

Educators and researchers have been studying the assessment methods utilized in academia for the qualification and quantification of the impact they have on the students learning, understanding and educational development. The areas of research covered by learner and instructors include student evaluation, educational milieu, role of ICT in education, privatizing education, critical thinking, educational interventions, standardized testing, teacher professional development and classroom interaction, where exploration has been ongoing. Below we have presented a comprehensive review of the work that has been done in the area of assessment in general and formative assessment in particular.

2.2. Assessment Trends and Practices in Pakistani Schools

The existing literature focuses on the current assessment practices and trends in Pakistani schools. In 2003, Rehmani discussed the impact of public examination system on teaching and learning in Pakistan (Rehmani, 2003). The study was conducted in four schools of karachi. It was found that the purpose of examinations in our country is merely to promote, select and certify learners. Because of the summative nature of assessments, teachers teach for testing rather than learning. Students are supposed to rote memorize the knowledge and reproduce in exams. Our current examination system doesn't play a positive role in improving education quality. It was suggested that making formative assessment an integral part of learning would help in improving the system that would in return improve learning and enhance quality of teaching.

The examination and assessment system prevailing in the primary education system of Pakistan was studied by Shamaas (Khattak, 2012). The study shows that internal examinations and assessments in educational institutes do not have a positive impact on student's learning due to teachers lacking the needed training in assessment. Secondly the assessments are summative in nature and are held annually to promote the students to next class or retain them in the current. Such assessments are subjective and do not provide enough feedback to pupils for improvement. Self-assessment and peer assessment are not practiced and therefore student lack the knowledge about learning goals and where they stand on the learning curve. Learners and learning outcomes

are not considered in designing the assessments, instead rote learning and memorization of manuscripts is encouraged. In this context the author suggests the use of multiple assessment tools to ensure a balance between formative assessment approach and summative approach. This will ensure inclusion of cognitive domain in the learning process and will improve student performance and understanding.

2.3. Impact of Formative Assessments on Students' Learning

Various researches have been conducted to determine the effect of formative assessment on student achievement. In 1998, Black and William in their article about classroom evaluation answered if formative assessment raises standards in education, leads to improvement and how formative assessment itself can be further enhanced (Wiliam P. B., 1998). In their work they studied books and articles from more than 160 journals and prepared a length review. They introduced effect size as a measure of ratio of average improvement in the scores of students involved in innovation to the average scores of a typical student group on the same test and showed that effect sizes of formative assessment experiments were between 0.4 and 0.7. Their work highlights that formative assessment results in equity in learning outcomes by particularly helping out low achievers and students with learning disabilities. Due to consistent assessment feedback pupils who would otherwise cease to take school seriously got better engaged. Enhancing feedback between students and teachers, actively involving students in the learning process, adjusting learning and teaching using assessment results are ways in which assessment can affect motivation and self-esteem of students. It is evident from their research work that current classroom assessment methods in many educational milieus are quantitative and lack guidance on improvement, encourage superficial learning rather than effective learning, overemphasize marks, inspire competition rather than collaboration and neglect learning functions.

In their work they maintain that a culture of self-esteem and success backed by the instinct of belief in yourself and your capabilities can be developed using formative assessment practices in the classroom particularly for low achievers, by concentrating on specific learning needs and providing a clear understanding, such that competition and comparison are not done among students. They suggest self-assessment and peer assessment complement formative assessment

and are inevitable as a successful innovation in the learning process. Self-assessment skills help students develop further understanding in terms of learning targets and where they stand on the learning continuum. Students become more involved and effective learners by discussing their assessment with teachers and other students, a thought process initiation that helps them establish concepts and resolve disparities. Peer assessment provides relevant feedback to students in their own language and makes them work in a group, which is good in terms of collaboration and group discussion, thus giving the pupils an opportunity to communicate their evolving understanding. To make teaching effecting, students should be given classwork and homework assignments as a means of promoting feedback such that they display students understanding and promote consistent dialogue regarding them to reorient pupils thinking. Black and William insist that when indulging in such dialogue teachers should not ask students direct questions, provide enough time to think and respond, give hints to direct them, initiate small group discussions and help them understand with examples in case of incorrect answers. In this way a formative process starts, where dialogue evokes a thoughtful reflection in which all pupils should be stimulated to participate. Feedback involving guidance on strengths and weakness should be elicited and opportunities should be provided to work on the shortcomings.

A qualitative study by Aziz (2003) described the impact of online quizzes on student engagement and learning. The study used University of South Australia's online environment:'UniSAnet' to develop online quizes for Principles of Computer Systems. The quizzes contained multiple choice questions and fill in the blanks. the quizzes could be attempted at anytime from anywhere. The flexibilty and usefulness of quizzes made them an intresting tool for learning. The results indicated that students were quite engaged with the quizzes and benefited form them. About 72% students found the quizzes helpful in finding out their strengths and weaknesses in the course. It could be stated that online quizzes can be an effective way to keep the students engaged with the course content. Quizzes can stimulate and maintain the interest of students in the course, hence moving the learning forward.

In a similar study by Balter, Enstrom and Klingenberg the effect of short formative diagnostic web quizzes with minimal feedback (correct/incorrect) was determined. They conducted this study in two universities: KTH Royal Institute of Technology in Stockholm, Sweden and Williams College in Williamstown, Massachusetts. The purposes of tests were to encourage

students to learn before coming to the class, to have an overview of their knowledge and to help teachers in identifying areas where they need extra support. Their self esteem would be strengthened by knowing that they know more than they thought and finding out weak areas would help them to study more. The results of study showed that formative assessment, as quizzes, improved students' learning.

Johnson and Kiviniemi studied the effectiveness of weekly reading quizzes on exam performance for Social Psychology Course (2009). This study was carried out at Mid-Wester State University on undergraduate students. The students were assigned a chapter for each week. The quiz was posted a week before the class and students had to read that chapter before taking quiz. At completion, the quiz provided them the feedback mentioning the correct or incorrect options along with the page number containing relevant information of chapter. The relation between completion of quizzes, exam performance and class performance was examined. Results of the study showed that the students who took quizzes showed a significant improvement in their performance. The feedback helped students to realize that they chose wrong answer and from where they could find the additional information to master the content. In order to keep the students motivated and maintain the effectiveness of reading quizzes, the authors suggested that the weight age of quizzes might be increased.

Several studies examined the use of technology in formative assessment. In a study by Jacoby, Heugh, Bax and Branford-White (2014) they determined the use of formative assessment in student engagement using virtual learning environment that was developed using ICT. The study devised formative assessments with learning modules such that each contained 10 multiple choice questions and 60 questions in total. The formative assessments were planned with unlimited time for attempts and unlimited number of attempts, so the students have enough time to complete the exercises. In order to attempt the exercises student were to study the core text and quiz related material; the goal was to enhance student learning skills by acquiring and researching the answers to questions. In addition to this assessment also included an opportunity to motivate their answer by incorporating comments and thoughts regarding their answers. The comments were categorized as positive, negative and neutral towards the formative assessment. After each assessment attempt, students were provided instant descriptive feedback that included marks and suitable corrections. The data showed an increase in the uptake of formative assessment and higher grades, resulting in student engagement in the learning process. This can

be attributed to the fact that students got enough time to investigate and answer their questions, got multiple assessment attempts and were provide instant feedback with not only marks but also further directions to improve learning. In their study they established that their virtual learning process designed using formative assessment framework has enhanced student learning experience by improving their comprehension skills and engaged them in to the learning process, resulting in higher student grades.

In 2012, Rehmani carried out a research in which the role of assessment for learning in improving the student learning through feedback was focused (2012). It presented how the change in assessment practice at middle school level can be initiated and what are the implementation challenges. The work introduced the use of portfolio assessment of project work as an alternative to the conventional assessment based on marks and grades in Pakistan. Portfolio assessment is considered to entail self-assessment, motivate students towards learning and enhance cooperative learning. It can help teachers how to assess students and provide remarks and feedback to improve student learning. The board at the institution designed and developed tasks for students with one core question in regard to the learning goals and content based objectives. The individual and group tasks were designed to provoke critical thinking, problem solving skills, decision making skills and ethical awareness. The teachers at the facility were provided the needed professional development to conduct the execution and assessment of tasks. Students were asked to do self-assessment and peer assessment using a checklist provided by the board. Each student was to submit a reflective note on the project work. There were no grades and instead remarks and descriptive feedback as an essence of formative assessment were used to address learning disabilities and support students in improving their learning. The board envisions that such an engaging learning process will gives students learning ownership, develop self-efficacy and boost confidence to succeed in life.

In one of the studies presented by Nobert, Caroline, Yishay and Harvey (2010), the role of technology in learning practices that use formative assessment are identified. The study presents an inclusive view of five case studies that were conducted on a group of students using projector, audio files, mobile phones, web tool and a string comparator as technological instruments to aid the presentation and organization of student thinking, support offline communication between learners/teachers, encourage individual reflection, provide graphical feedback to teachers on the

student response frequency and give systematic feedback on writing items in multiple languages respectively. From this study they establish a key role of technology in formative assessment as it opened up the window of student's thinking for teachers. Students had the ability to represent their ideas and designs using multiple ways and be more creative and innovative. Technological intervention stimulated their thought process and allowed them to synthesize what otherwise would not have been possible. Teachers also were able to provide better and immediate feedback that was found to be more effective, emphatic and personalized than usual written feedback practices. They found it to be consistent with Black and William view that pedagogical change is itself a core aim of the formative assessment. Dual contingency is found to be working with learners promoted to engage in the learning process and teachers responded to students work using their pedagogical knowledge. The authors conclude that formative e-assessment is a set of processes that include technological and social resources to engage learners and teachers to enhance learning and improve learning. It is not the technology alone but how learners and teachers utilize it to shape formative assessment makes it an important tool, which should be optimized as part of the learning progression.

In another study, Thiru, Weena and Allan evaluated the use of self-assessment quizzes made available as formative assessment on the performance of students (Thiru et al,2011). In total 14 quizzes were designed during the course period using attributes of formative assessment and were release to students iteratively. Students were given unlimited attempts for the quizzes and the marks were not counted towards the final grade. The quiz assignments were designed to include core concepts and were both qualitative and quantitative in nature with instant feedback to students. In case of incorrect answer, students were provided specific feedback and if students fail repeatedly on questions that belong to a certain area of study a detailed feedback was provided that included examples to improve the understanding and develop the fundamental learning concepts. Teachers were given the possibility to remotely monitor the student progress and generate statistics regarding student engagement and learning outcomes per subject area. This provided instructors the opportunity to focus on lessons and topics where students need more improvement and support to grasp the underlying concepts to utmost. They found online formative quizzes to be an effective learning tool that can foster student knowledge, understanding, critical thinking and comprehension skills. A correlation between the students

that completed most of the quizzes and their overall performance was very strong and they were able to score better on summative assessments than those who attempted less.

Findings from Ron, Jillian & Angela's study (2011) shows that a voluntary intervention designed using formative assessment attributes, was beneficial to student learning outcome. The study was conducted on newly inducted students in the graduate program who were offered a research methodology and scientific writing assignment (literature review of 5 journals) on voluntarily basis. The intervention was designed to be formative in nature such that students were provided topics, word limits, submission deadlines and regular engagement occasions with the staff to identify issues and address them during the course of intervention. Students were given written descriptive feedback that highlighted the areas of strengths and the areas that needed improvement in the context of analysis, research, academic referencing and writing of literature review. The results of this intervention were positive and showed that students that submitted a draft voluntarily and participated in the learning activity obtained higher marks for the literature review in the semester's summative assessment in comparison to the students who did not submit assignment. The feedback from the students who participated in the learning process identified the intervention as an enabler of improving their critical analysis and writing skills. The engagement opportunities with the staff enhanced their understanding of the learning process and fostered the link between learners and instructors. Both believed that the intervention fulfilled the formative function and lead to a new level of understanding that would be helpful student in the next courses. The intervention supported students' transition to the universities educational environment by providing an encouraging learning environment characterized by formative assessment features.

Sharmin investigated the inclusion of formative assessment and productive pedagogy in secondary schools of Finland by evaluating teaching practices, classroom environment, curriculum design and teacher guides (Ahmed, 2015). The research data constituted two teaching guides, core manuscripts and seven teachers from seven different schools in the subject of mathematics for grade seven. An open-ended questionnaire was used to gather teachers understanding of formative assessment in line with the research questions such that OECD (Organization for economic cooperation and development) guidelines regarding formative assessment were used as a framework for analysis. Similarly, an open-ended questionnaire was

used to gather teachers experience regarding productive pedagogy and analyzed using pedagogy dimension given by Queensland School Reform Longitudinal Study. The core curriculum document in mathematics from Finish ministry of education was analyzed for policy regarding the use of formative assessment and productive pedagogy directives in the educational milieu. Their analysis shows that the most powerful ways of improving learning is emphasizing the day to day use of formative assessment. Classroom practices that involve peer assessment, feedback and stimulate student engagement are prevailing formative assessment attributes in the Finish educational context, which can be further elevated by adjusting assessment as per abilities of the students and more focus on peer assessment. The work also shows that productive pedagogy is widely practiced in Finish educational system in terms of curriculum and teaching. Teachers have been successful in scoping the lessons and teaching material as per the goals outlined in the curriculum. In order to improve further, the study reveals more focus on formative feedback and peer assessment so the students have co-ownership in the learning and development. The study reveals strong interaction among formative assessment, productive pedagogy and curriculum and reiterates new modes of pedagogy where pupils are actively involved in the assessment process.

Pearson research report gives an overview of the use of digital games in learning and assessment (Katie Larsen McClarty, 2012). The use of digital games in education is evaluated on the basis of five claims 1) sound learning 2) engagement for learner 3) personalized learning 4) teach 21st century skills and provide an environment for relevant assessment. Their research shows that sound learning can be improved using digital games as they provide opportunity to think, understand, plan and take actions. There are possibilities for continued practice and encouragement as games allow replay and advancement to next levels. As games have milestones and provide immediate feedback, players can change their strategy and reach their goals. Immediate descriptive feedback is core to formative assessment and its reflection in the gaming domain is a promising learning aspect for educational system. Their findings show that digital games enable personalized learning as gaming adapts after inferring the abilities and the skills of the players. Traditional lessons and assessments lack this possibility of tailored learning and if a student lags behind or fails to grasp a concept, he/she will have difficulty to understand future concepts build upon the earlier concepts. Such gaps are found to be vividly covered when digital games are employed in education as students can have several attempts until they can fully develop the needed understanding. In addition to this, games provide hints and suggestions

to players in case they are not on the right trajectory, guiding them from level to level and allowing them to learn the knowledge and skills as the progress further. Student can try their skills, take risks and quickly learn from faults. The research shows that digital games provide more engagement to learners in comparison to the traditional classroom setting, which is found to be boring for students. The uniqueness in digital games comes from their ability to sustain engagement and motivation over the period of time, which is specifically useful when the learning tasks are challenging. The research in the study also shows that games play a relevant role in assessment as assessment is already an inherent part of gaming world where abilities are quantified. Being interactive in nature, games can provide real time data on student's actions and decisions that can be used for formative and summative assessment of student. This embedded gathering can help to examine the knowledge and skills of students. This research shows that digital games incorporate motivation, adaptivity, engagement, simulation, collaboration and data collection, with a potential to facilitate learning.

Formative assessment is defined as the process by which students recognize and respond to their learning in order to improve that learning (Beverley Bell, 2001). The findings in this study are about investigative research on formative assessment practices in science classrooms in New Zealand. Their research work was qualitative, interpretative, collaborative and utilized different data collection techniques including surveys, interviews and participants' observations. Data from dozens of teachers and their students about assessment ideas, classroom studies and teacher development was collected to evaluate the existing formative assessment practices and development of formative assessment by teachers. Their findings reveal that both teachers and students sense that formative assessment is flexible and responsive as it can be practiced in planned or unplanned, formal or informal and proactive or reactive way. Their work showed that formative assessment is a good source of actionable information and it provides ample evidence about student learning. Teachers utilized the information elicited to feedback the students and helped them improve their knowledge, thus enhancing learning and social development of students. Self-assessment and peer-assessment for both teachers and students were found to be important traits of formative assessment. Formative assessment is really a collaboration involving both teachers and students who learn to learn and learn to improve learning as they move forward in their education. In summary they maintain that formative assessment is characterized by contextualization as the information gathered, interpreted and actions taken

depend on learning situation (e.g., individual or whole class etc.), learning activities (e.g., brainstorming, self-assessment etc.), lesson goals, teacher's knowledge and the amount of information students disclose to their teachers.

In another study, Roy investigated the use of computer aided assessment (CAA) as a formative assessment tool on student performance and learning outcomes (Lowry, 2005). The study was conducted on students who were familiar with computer aided summative assessment. The methodology employed was to divide the summative assessment in multiple formative assessments based on the course work covered during that period of study. Conforming with the legacy multiple choice summative assessments, the multiple formative assessments were also kept as multiple choice questions. This enabled the students to develop their understanding of the subject and practice the future summative assessment module. In this manner, they could implicitly gain familiarity with the final assessment. CAA system used comprised of multiple programs to generate questions based on course work, do assessment compilation and monitor the student progression towards the learning goals. For each lecture 5 questions were made available via web and student could access them whenever they like and as many times as they need. After each assessment attempt student was given overall descriptive feedback that included not only scores to show how good student has performed but also feedback regarding their answers to each question, so they learn the intended lesson. The results of this research show that student showed considerable interest in these self-assessment tests that were formative and voluntary. The usage statistics from the CAA system showed that majority of the students who used this assessment started to use it within the first two weeks. Due to the fact that the CAA system was not designed to impact the final grades of students, it's uptake was encouraging and showed that students interest in developing their learning grew as the lessons continued. The results of this study show that students who actively participated in these voluntary computer aided self-assessments performed better in the final summative assessment than the group of students who participated to a lesser extent and the group of students who didn't participate. The study results also reveal that the student feedback gathered regarding the computer aided self-assessment showed that it promoted their confidence, ability to think and identify the areas of weakness.

2.3. Formative Assessment and Gender Disparity

Numerous studies have attempted to explain the difference of gender in formative assessment. Louis and Danielle studied the synergies and tensions between research and practice of formative assessment in the contemporary classroom setting in Canada's two school districts (Beckett, 2011). The researchers maintain that teacher's perception of assessment is crucial to teaching and learning as it directly influences their teaching and the insight students develop. Similarly, research suggests that changes in teacher's attitude influences the changes in formative assessment practices, which implicates teacher development providers to design and introduce teacher education reforms after judging their perception. In this context, Louis and Danielle conducted their study by interviewing twenty teachers about their understanding and use of formative assessment attributes in the classroom settings. The participants were selected using purposive and convenience sampling across the two school districts. The interview focused on questioning techniques, feedback without marks, self-assessment and peer assessment attributes of formative assessment in their study. Also highlighted in their work are the main factors that contributed to the gap in the classroom practice. Their results show that regular feedback without grades to students had a high value. The non-evaluative nature of feedback allowed them to focus on making the results better and improve students cognitive thinking and motivation. However, they do realize that there exists a tension between feedback driven teaching milieu and the pragmatics of evaluation. The results further show that involving the students in the assessment process is vital to student learning, as assessment itself is learning. Teachers shared that student are more keen and involved when they assess themselves in regard to their learning goals. In this way, they reflect upon their achievement and take ownership of their learning. Despite self-assessment's fruitful nature, teachers admitted that they need to do better in promoting it in their classrooms. In contrast to other research, this study noted difficulties in the use of peer assessment and its application in the classroom. The teachers reported that students were not truly objective when doing peer assessment and their qualitative feedback was fine but the quantitative data was biased by friendship and inclination with their fellow students. To summarize, their work shows the effectiveness of formative assessment attributes when practiced in an educational environment and also highlights the barriers that have resulted in the gap between research and practice of formative assessment.

Vasileios and Anastasios presented a comprehensive view on the constructs that affect male and female students' behavior to use a computer based assessment (A.Economides, 2011). Prior research on this topic has been based on different computer based assessment (CBA) acceptance models and in their work they have used already available research and build on it to develop their own computer based assessment acceptance model (CBAAM). Further, the study explored the gender differences in the perception of computer based assessment acceptance model (CBAAM) constructs namely perceived playfulness, perceived usefulness, perceived ease of use, social influence, computer self-efficacy and facilitating conditions. In their research they have defined these constructs as: Perceived playfulness is defined in terms of user's concentration, curiosity and joyful interaction with the system. Perceived ease of use is defined as how easy it is to use a system effortlessly. Computer self-efficacy is defined as the ability/capability of a person to use computer. Perceived usefulness is calculated as the extent to which a person believes that using a particular system will enhance his/her job performance. Social influence is defined in terms of the effect of society on a person. The experiment data was collected from 173 students who volunteered for the research study of which 33% were male and 67% were female students. The CBA consisted of 45 multiple choice questions that student need to respond in 45 min time period. At the end, each participating student had to answer a survey about the experiment. The result analysis in their study is helpful for teachers and examiners to understand how different information technology and learning management system constructs influence the gender, so they can develop better computer based assessments. Their research concludes that when male students use a CBA system they are influenced by its playfulness, usefulness, content and social influence. On the other hand, when female students use a CBA system they are influenced by ease of use, content and playfulness. Although content and playfulness were found to be common traits that influence male and female students, in male students the degree was more compared to their female peers. Secondly, the females were more instigated when the CBA interface was simple and everything was in a flow and pointing to a logical destination.

Historic research review shows that assessments and evaluations are intrinsic to raising the standard of education and is perceived as the only tool that can help qualify and quantify the effectiveness of what is being taught and learnt. Educational institutes, teachers, instructors, learner and students all are assessed in different educational settings in order to identify the areas of improvement and bring reforms to the prevailing practices. Policy makers and regulators in

the education field use the feedback they gain through such large scale assessments (summative assessment) to unify and bring equity in the educational standard across a state or region. There are global initiatives like education for all (EFA) program under the united nations (UN) millennium goals for education. This and other such programs are meant to facilitate the acquisition of education and expand its reach in order for it to be widely and easily available to masses. Such practices are duly important and absolutely necessary keeping in view the knowledge based economy needs of nations.

2.4. Summary

Literature review founds that Summative assessments prevailing in the established educational institutions put so much focus on marks and grades that the whole system is merely a statistics machine that puts students in the order of achievement and in competition with each other. This practice is used for promoting students from one grade to next grade but it does not help them develop interpersonally, socially and intuitively. During the course of their study, students go through lessons, readings and exercises and afterwards they are required to give an assessment that quantifies their learning and understanding. Thus the system uses assessment as a means to measure the student performance instead of helping them do continuous learning enhancement. Summative assessments being in the end of the learning cycle therefore do not really influence how student learning can be elevated and how to cater diverse student learning needs.

According to the review of the literature, formative assessment is found to be imperative to student learning and effective enhancement in their achievement. Black and William showed that formative assessment results in equity in learning outcomes by particularly helping students with learning disabilities (Wiliam P. B., 1998). Motivation and self-esteem of student is elevated when they are actively involved in the learning process and teaching and learning are tailored to students learning needs. Formative assessment is an engaging learning process that gives students the ownership of their learning and authority to develop themselves, thus improving their performance and understanding. Research shows that descriptive feedback is more valuable and effective than regular feedback as it includes comments and suggestions to students so they can work on their deficiencies and develop better understanding.

The use of technology and social resources along with the formative practices have led to the introduction of formative e-assessment (Norbert Pachler C. D., 2009) practices, which are relevant in today's information and communication technology (ICT) dominated world. Use of technological innovations in education in the context of formative assessment studies during the literature review establish such innovations as effective learning tools that are more powerful than traditional interventions and plausibly more flexible to optimize as part of the learning progression. In this context, web based online quizzes based on formative assessment characteristics are found to be effective in engaging students in the learning process and improving their performance (Thiru Aravinthan, 2011). Digital games are engaging and inherently evaluative and research shows that they have the potential to provide real time data about student's thinking, understanding and decision making that can be efficiently used for formative and summative assessment to facilitate student learning (Katie Larsen McClarty, 2012). Similarly, research shows that computer aided assessments based on formative assessment attributes like self-assessment, engagement, interaction and descriptive feedback promoted student's confidence and ability to think (Lowry, 2005). Our research review further shows that the use of technology as an innovation in education does require its intended users to be familiar with the technology and the behavior intentions different between males and females in this regard (A.Economides, 2011).

The literature review done as part of this study reveals that much work has been done in evaluating the effectiveness and importance of the use of formative assessment in education as a tool to improve student learning, engagement, attainment and performance. However, there are yet no conclusive remarks regarding its efficacy when innovations in education are used as formative assessment tool generally and particularly in Pakistan's educational milieu. In Pakistan, majority of the assessments are summative in nature and as highlighted in the problem statement student learning outcomes are not the focus. From research, we have seen that in order to truly impact the student knowledge, skills and performance it is imperative to perceive student learning in terms of their quest to learn, engage and be part of the learning process. We need to examine the education environment to investigate how it fosters self-esteem, motivation and a positive attitude towards assessment. Literature review shows that such assessment practices that results in student engagement and high student efficacy by using technological innovations in education are lacking.

Chapter 3: Methodology

3.1. Overview of the application and targeted outcomes

The science quiz application was based on first unit, named "understanding ourselves", of grade four general sciences' textbook. It was developed according to the National Curriculum for General Science 2006. Formative assessment is defined as a process that enhances student learning, during the learning (Cowie, 2001). According to Burr, assessments should be designed in a way that they meet the intended learning outcomes (Burr, 2009). The purpose of developing this formative assessment quiz application was to move students' learning forward by providing timely feedback in the form of additional information and to engage them with the content by adding game elements. The students' learning outcomes covered in the application, according to the National Curriculum are given below:

Grade –IV **Learning Contents and Students' Learning Outcomes**	
Contents	**Students' Learning Outcomes**
UNDERSTANDING OURSELVES • Introduction to Human Body • Major Body Parts and their Functions (Teeth, Brain, Lung, Heart, Stomach, Skin, Eye and Ear)	All the students will be able to: • Identify major parts of the human body, • State functions of major parts of the body.

Table 1: Learning Outcomes from National Curriculum for General Science 2006

Formative assessment when combined with technology enables students to take actions by which they can bring changes in their knowledge, skills and understanding (Norbert Pachler C. D., 2009). In our study we have used gamification as technology enabler to explore the formative assessment's effectiveness. Game is defined as "physical or mental activity or contest that has rules and people do it for pleasure" (Dictionary and Thesaurus, 2016). According to Wikipedia

game is defined as “structured form of play, usually undertaken for enjoyment and sometimes used as an educational tool”. In gamification game design elements are used in non-game context to increase user engagement (Arieli-Attali, 2015). Gamification applies characteristics of video games to non-game applications (Doyle, 2014). Gamification can improve the quality of learning by absorbing students in a better way with learning activities (Christopher Cheong, 2013). Unlike the traditional education manuscript, games are natural learning machines as they have potential to deliver on demand and within context information (Gee, 2007). Gamified learning has the potential to support people in improving their skills and sustain the need engagement level. Based on these learning potentials of games, Microsoft has introduced Ribbon Hero, a video game with the purpose to educate the users of Microsoft office. Similarly, GamiCAD is a gamified tutorial of AutoCAD for beginners. Inspired by gaming Mozilla introduced OpenBadges, which are virtual signs of accomplishments for the skills people learn and can be shared and verified as records of learning across the web.

Digital games are games that require an electronic hardware for execution and constitute high resolution images, sounds and simulations to create a fantasy where players compete, challenge, and interact. Digital game based learning is an important means of altering learning process by appealing and exciting students. It is about fun and engagement (Prensky, 2001). According to the Gamification wiki game mechanics like achievements, infinite play, levels and bonuses help increase student engagement during the game based learning. Games consist of a number of key elements called game elements such as goals, outcomes, feedback, competition, interaction, story and representation (Prensky, 2001). For creating a gamification environment to increase student motivation, the focus should be on fundamental elements that make games appealing to their players (Adrián Domínguez, 2013).

In our work we have gamified the science quiz application with the purpose to enhance students’ engagement with the content and motivate them to progress further in their learning quest rather than assessing them in terms of fail or pass grade. The game elements used in this quiz application are story, characters, sounds, feedback, rewards and scores. The game application employs narrative scenario for storytelling to keep the students engaged and instigates thinking. Use of characters and sounds in the game application is to bring aesthetic pleasure to the players. Jon Ingold, creative director at inkle Studios said “For me, games are the most fascinating

medium for storytelling available right now". Dave Gilbert, founder of Wadjet Eye Games said "Video games achieve something that other forms of storytelling just can't". The questions following the narrative scenario are designed to enhance the student learning about main human body parts and their biological functions. The application design supports self-assessment so students can find out what they know and what they need to learn in order to augment their knowledge. Oxford dictionary defines self-assessment as "Assessment or evaluation of oneself or one's actions and attitudes, in particular, of one's performance at a job or learning task considered in relation to an objective standard" (Press, 2017). Immediate descriptive feedback supports the self-assessment by giving a hint or an indication in relation to the learning goal, so student can address the lack of understanding and do self-correction. Effective feedback is one that initiates thinking (Dylan, 1998). Continued descriptive feedback related to learning goals is employed in the application design to improve student knowledge and build a culture of learning in the classroom. Below we present a summary of the targeted outcomes of the gamified quiz application:

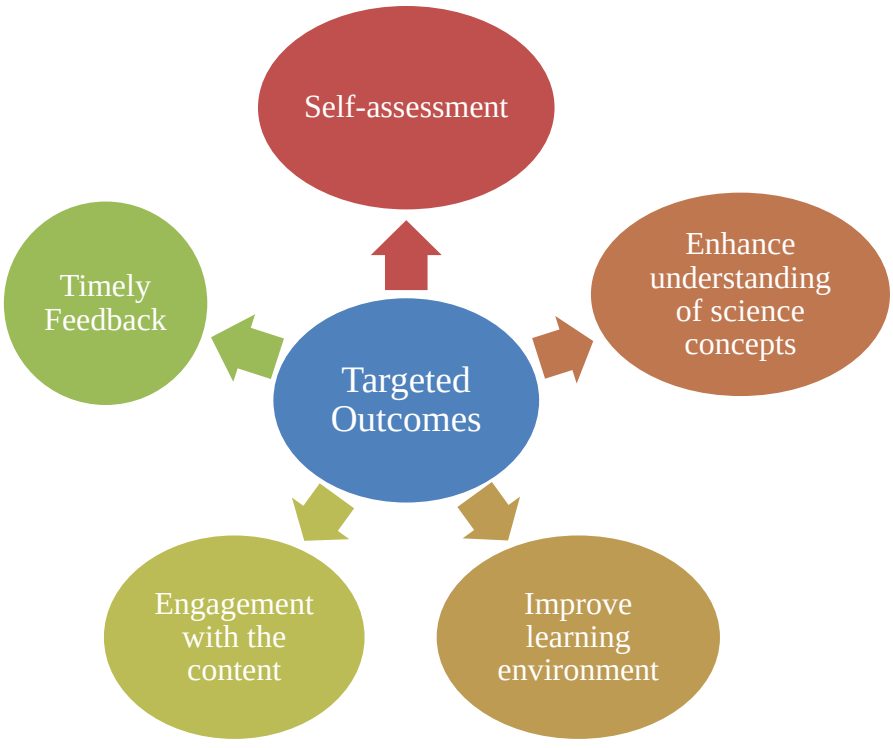


Figure 1: Targeted Outcomes of the Science Quiz application

3.2. Application Design

The application was designed for Android OS, which is ready made open source mobile operating system from Google Inc. The motivation for designing the application for Android OS comes from its vast share in the smart devices market globally. Being open source it is not only easy to be customized for different devices and usability needs but also becomes affordable enough to be found on low cost devices. When it comes to producing affordable devices for emerging markets, or providing access to thousands of educational and health applications, it helps people access information and opportunity than any other mobile platform (android.com).

The game development for the science quiz application was done using Unity, which is a multi-platform game engine used to develop video games for PC, consoles and handheld devices. The selection of Unity for game design and development was based on its powerful capabilities.

The main screen of the quiz application is shown below:

Figure 2: Main Screen of the application

Science quiz application starts with a scenario based story with the purpose of activating prior knowledge of students. (a) There is a boy sitting in his study room. (b) He feels hungry and asks his mother to bring something to eat. (c) His mother brings apple and milk for him. (d) After eating that apple, he questions himself which organs help him in chewing the apple? (e) The boy

says: let's answer some questions. From here the formative assessment based quiz starts. The screenshots of this story are given below:

Figure 3: (a) The boy is feeling hungry

Figure 4: (b) Boy is asking his mother to bring something to eat

Figure 5: (c) His mother brings apple and milk for him

Figure 6: (d) The boy is thinking which body parts help him in chewing the apple

Figure 7: (e) The boy is saying: let's answer some questions

3.2.1 Features of the Application

The quiz questions are divided into two activities, after completing eight questions, the application moves to the activity 2. The quiz contains drag and drop and multiple choice questions. These questions were made keeping in mind the learning outcomes defined in National Curriculum. The features of the quiz application are:

- The quiz consists of 16 questions; 13 multiple choice questions and 3 drag and drop questions.
- Each *multiple choice question* has three options. On clicking the *correct option,* the application moves to the next question. The student gets scores and rewards upon choosing the correct option. When a correct option is selected the empty ball is replaced by a blue ball. The screenshot of a multiple choice question is shown below:

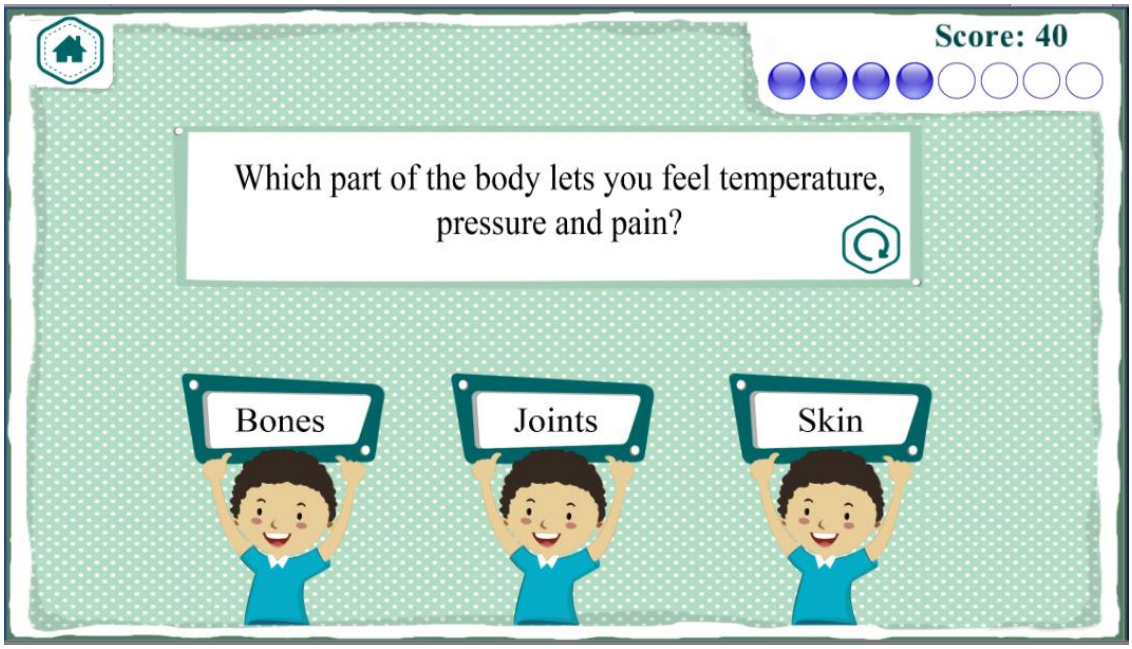


Figure 8: Sample Multiple Choice Question

When a *wrong option* is selected, the student hears a buzzer sound, a red cross is displayed on that particular option to show that the selected option is incorrect and the empty ball is filled by a red ball. A screenshot given below depicts that the student has chosen wrong option:

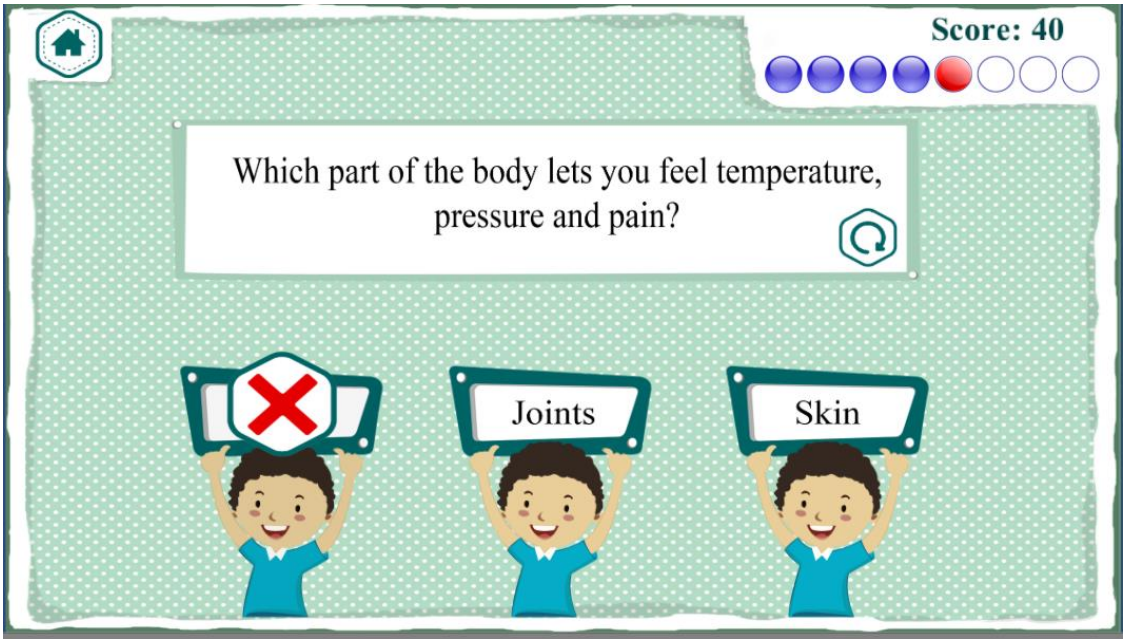


Figure 9: Selecting a wrong option displays a red ball and a cross with a buzz sound

Moreover, an *immediate feedback* dialogue box pops up, enabling the student to think again. An example of feedback is given below:

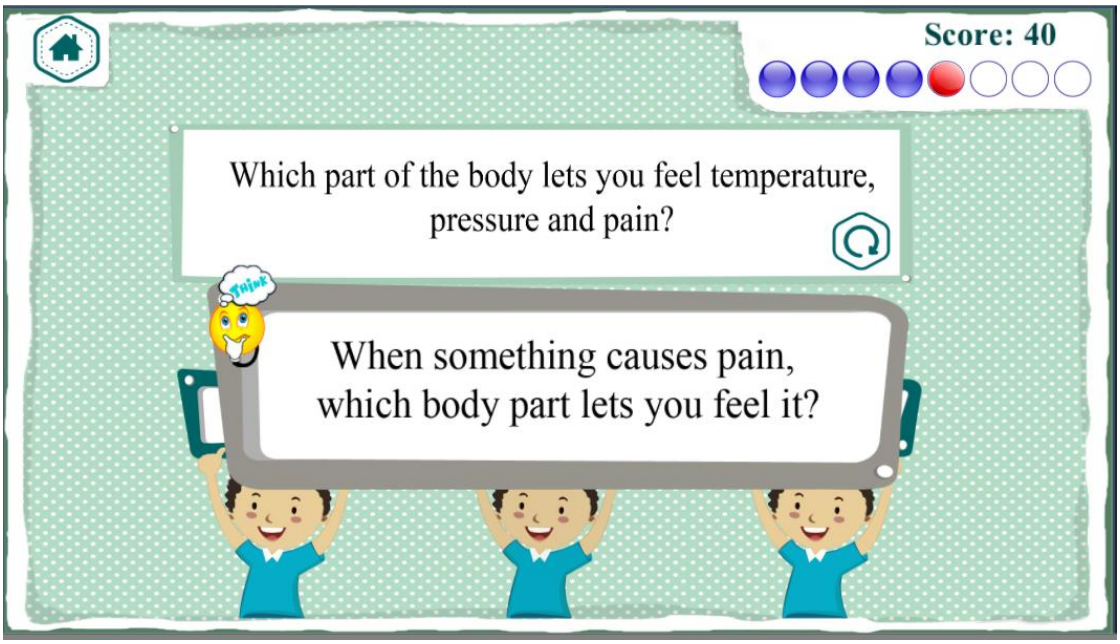


Figure 10: Feedback's dialogue box pops up, enabling the student to think again

- In *drag and drop questions*, the students have to select a correct picture by grabbing it and dropping it to the correct location. An example of drag and drop question is as follows:

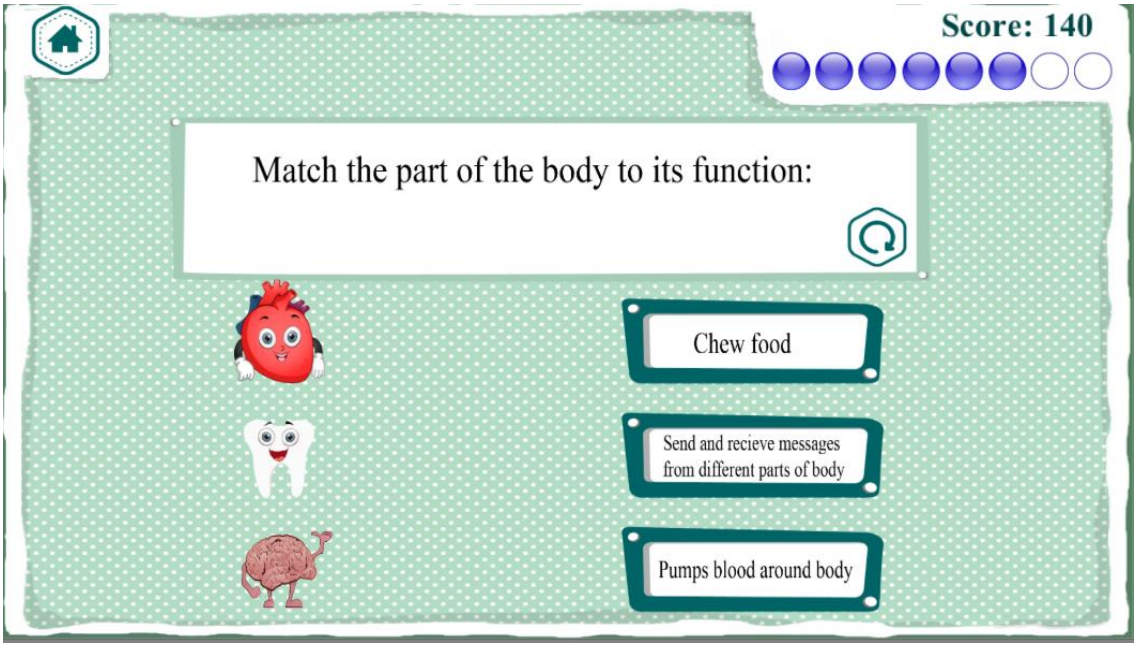


Figure 11: Drag and drop question; a human organ is to be grabbed and dropped to a correct function

- The last screen of the application displays total score:

Figure 12: When a student completes the quiz, the total score is displayed. The screen has a home button to go to the main screen and an exit button to close the application.

3.2.2. User-friendly Interface

The interface of an application has a direct relation to the success of the application. Keeping in mind the guidelines of a user friendly interface (Ben Shneiderman, 2010), following guidelines were used to get the students' attention:

- Commonness:
  - ✓ All the multiple choice questions have same interface; a home button to go to the main screen, score text box, balls as rewards, a question, a replay button and three options. In this way the students do not need to learn the interface from question to question. A sample interface is shown below:

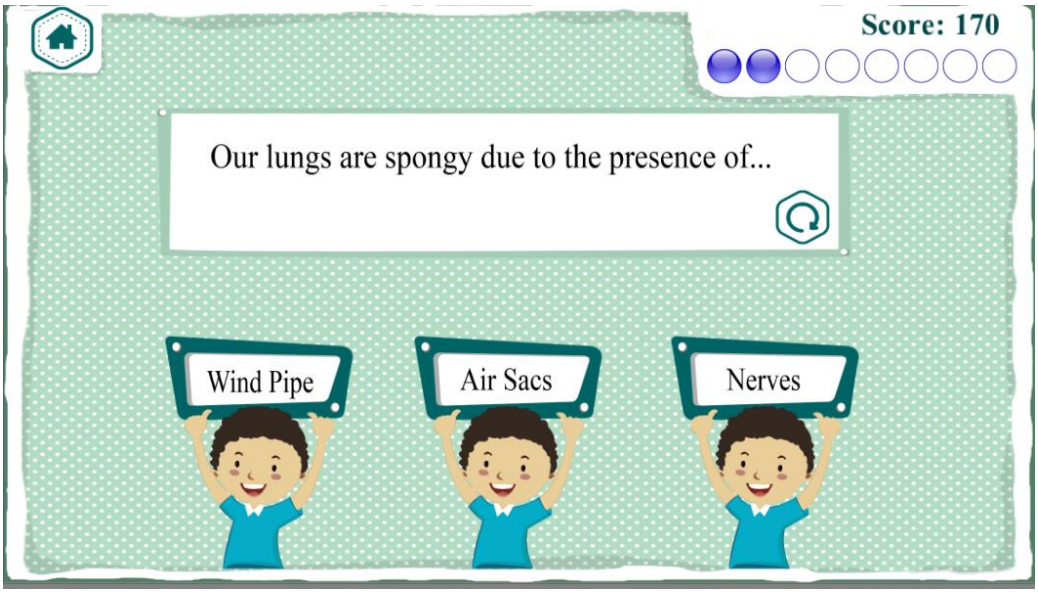


Figure 13: Multiple choice questions interface has a home button, replay button, score text box

  - ✓ All the three drag and drop questions work alike. The student has to grab a correct picture to the correct location. The sample interface is given below:

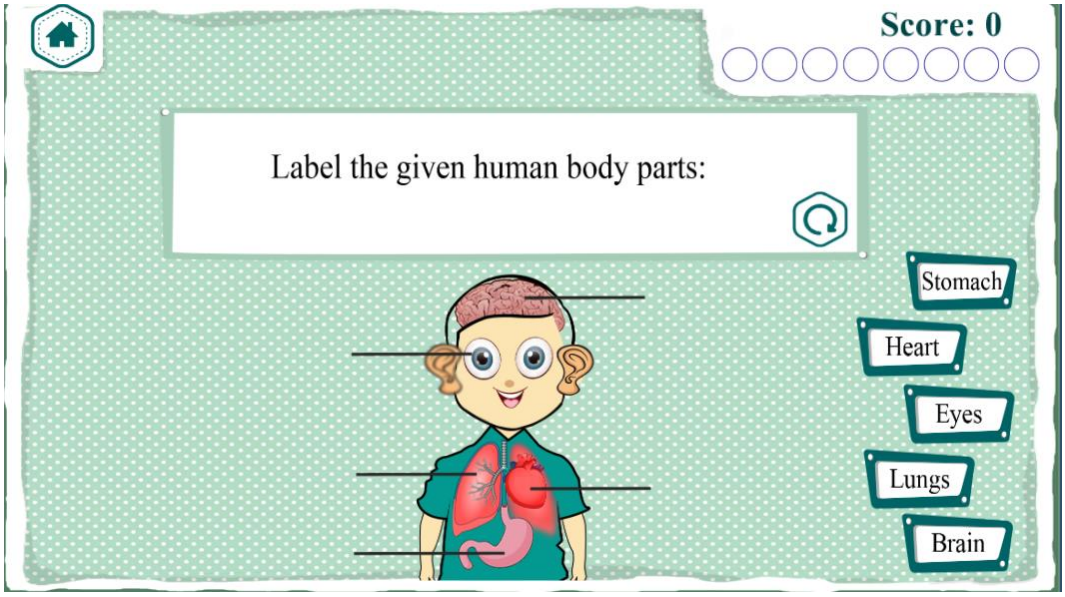


Figure 14: A correct option has to be grabbed and dropped to the correct location, otherwise it will come back.

- Consistency: There are no unpleasant surprises, the entire screens work alike.
- Feedback: There is feedback that keeps students informed about their actions. If a wrong option is selected student hears a buzz sound, sees a red cross and gets a red ball instead of a blue ball.

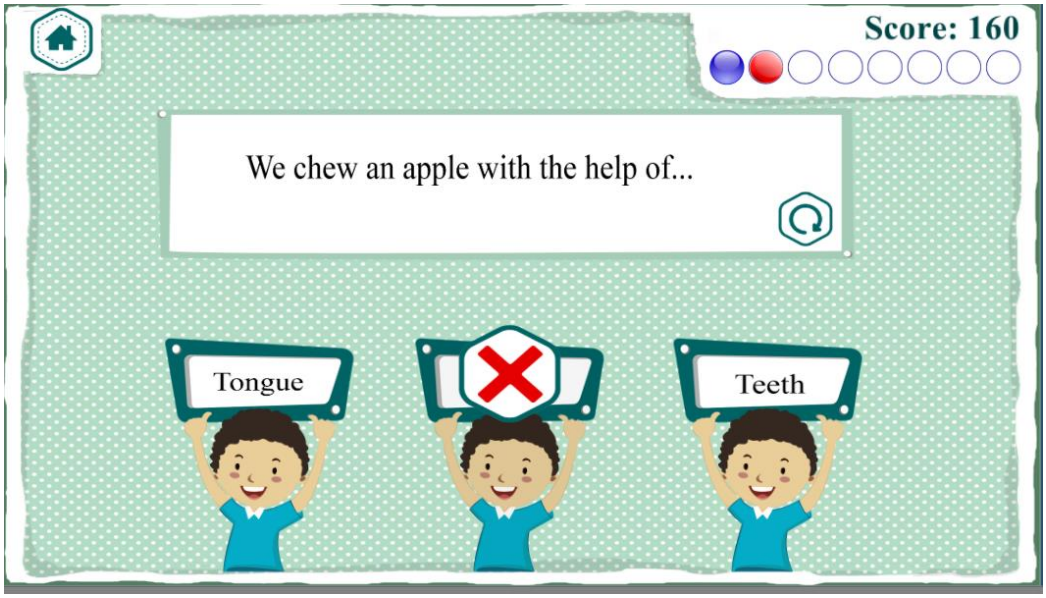


Figure 15: Choosing a wrong option is being informed by showing a red ball and a red cross

- Audio: Soft tones are used for dialogues, rewards and feedback, and a buzz sound is used for wrong option.

- Colors: The quiz application has up to four colors, to avoid over-stimulation. The overall theme is with green color. The main screen looks like:

Figure 16: The main screen of application

3.3. Usability testing of the application

Usability testing is a way by which actual users trying a product are observed and information is collected about the ease and difficulty of that product (Redish, 1999). The science quiz application was developed with usability in mind, tested iteratively and was improved on the basis of usability results.

The purpose of evaluating the application for testing was to identify the usability problems that were experienced by students. The testing comprised of 15 fourth grade students, including boys and girls. The usability assessment methods used were: observation, user feedback and questionnaire (Nielsen, 1993). The participants were briefed about the application and were asked not to share it with other sections, as they were selected as control and experiment groups. Android tablets were provided to these participants. The participants were observed silently against a usability checklist (see Appendix A).

After using the application, each participant was given a questionnaire regarding quiz application (see Appendix B). The questions were about:

- satisfaction with application
- clarity of the content
- ease of use
- ease of drag and drop
- understandability of the content
- icon clarity

Likert scale was used to collect their responses about the design and functionality of application.

3.3.1. Results of Usability testing

The information about the efficiency of quiz application was collected on a likert scale. The usability problems found in the application are shown in table:

Sr. No.	Usability Problem
a.	The *exit button* was missing.
b.	There was no *back button* to go to the main screen.
c.	There was no button to replay the question
d.	In drag and drop questions, the pictures of human organs had shadow that gave an impression to the participants as they are buttons.
e.	The main screen has images of organs with their labels. Each image had a shadow that made them look like a clickable button. Participants clicked them rather than clicking the start button.

Table 2: Usability Problems

Screenshots of the application before and after usability testing

a. There was no exit button. It was added later.

Figure 17: The exit button was missing, it was added later.

b. After watching the character based story the next screen of the quiz application consisting of a start button appeared. If a participant wants to watch the story again, there was no button to go back to the main screen. The 'back' button was added.

Figure 18: The back button was missing, it was added later.

c. Each quiz question had no replay button. If the participant had missed the sound, he/she could not listen to it again. Hence a ‘replay’ button was added.

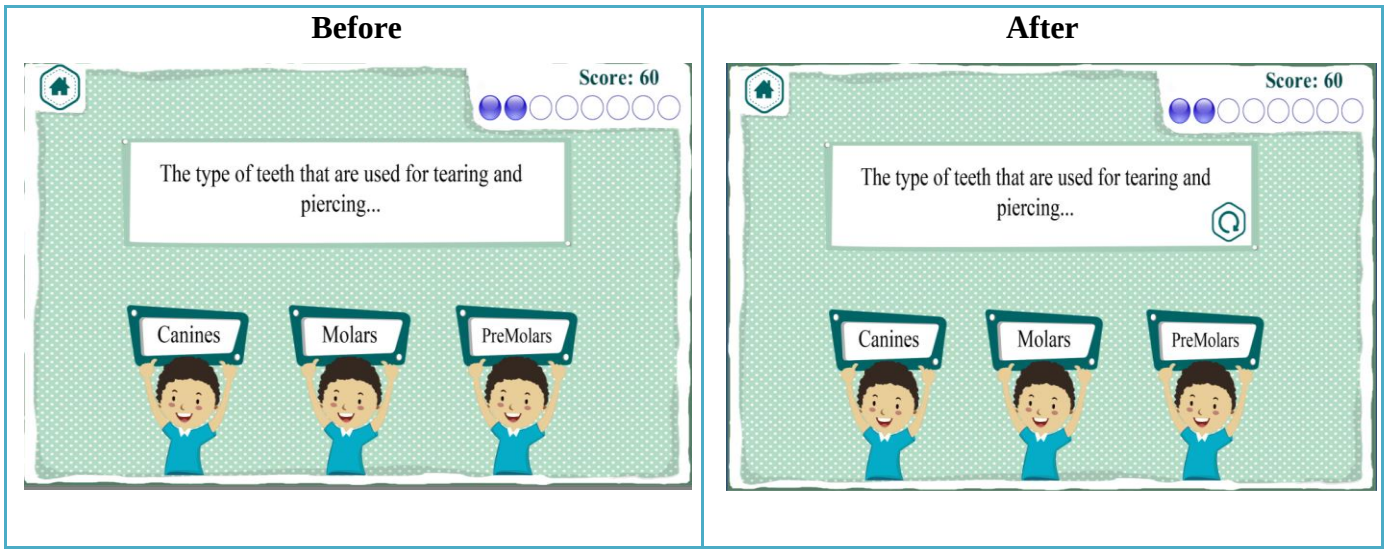


Figure 19: The replay button was missing; it was added after usability testing.

d. In drag and drop questions, the images had a shadow that visually make them look like buttons. The participants clicked them rather than grabbing the correct image to the correct location. The images are given below:

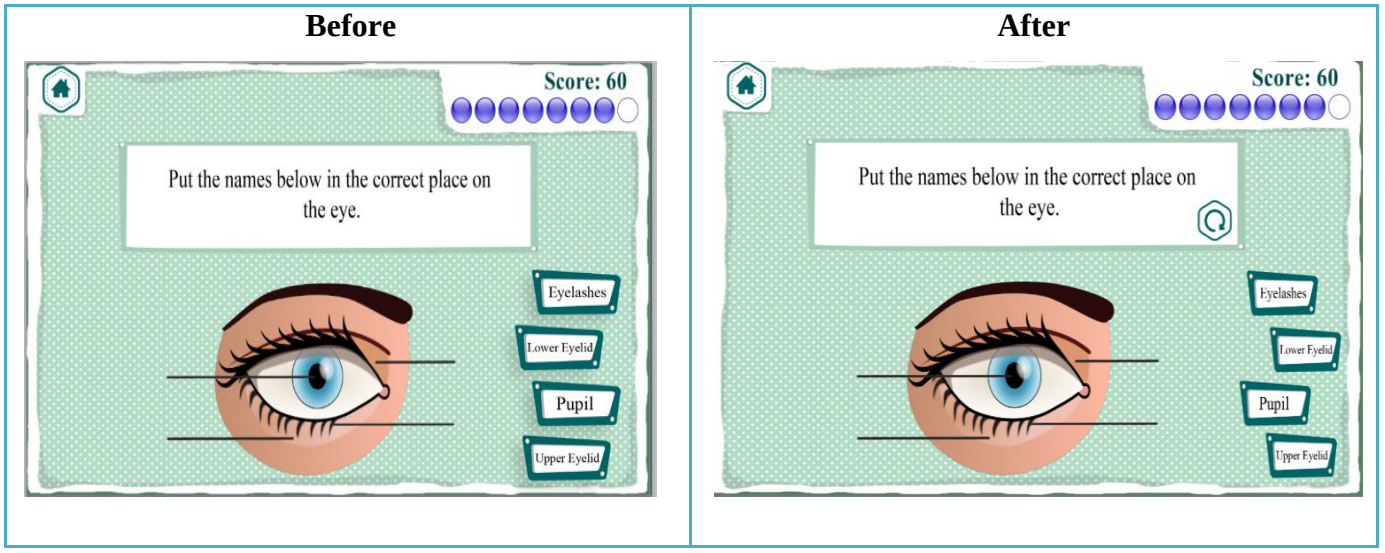


Figure 20: Shadow makes images look like buttons. The shadow was removed to give them look of images.

e. The organ images on the main screen had shadow, which made them look like buttons. The participants clicked them rather than clicking the start button. The screenshots are given below:

Before	After
	

Figure 21: The images of organs had shadow, which made them look like buttons. A few participants clicked them

f. If a student wants to skip the scenario based story, there was no skip button. It was added after usability testing. The screenshots are given below:

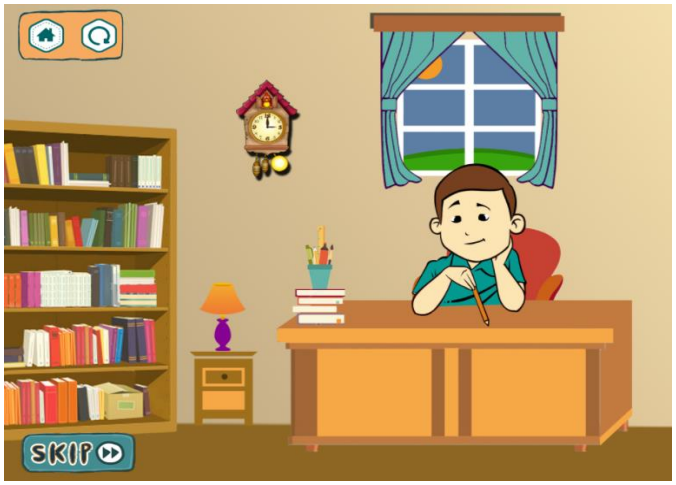
Before	After
	

Figure 22: The images of organs had shadow, which made them look like buttons. A few participants clicked them

3.4. Research Methodology

The study was carried out to determine if formative assessment had an impact on learning quality of grade four students, after utilizing an interactive science quiz application. The intervention took 4 days and continued two weeks in all for both control and experiment group. This chapter explains the: (a) research design (b) research question directing this study (c) research site and sample (d) research constraints (e) ethical concerns (f) instruments used to collect the data (g) procedures for data analysis.

3.4.1. Research Design

This study used an experimental research design. The experimental method is a method which can truly test a hypothesis concerning cause- and- effect relationship (Consuelo G.Sevilla, 1992). Assessment not only informs students about their achievements but is also a necessary condition for quality learning (Kerri-Lee Harris, 2006). The quiz was based on the first unit of science textbook for grade four. The students were chosen by random sampling and divided in two groups control (n = 60) and experiment group (n = 60). In random sampling each individual has an equal probability of being selected (Creswell, 2013). The minimum number of cases for statistical analysis is considered to be a sample size of thirty.

The control group was not given any treatment. The participants in experiment group received a treatment of science quiz application. Quantitative data of both groups was collected through pre and post-test and compared in SPSS. The independent variable was group: control and experiment and the dependent variable was score.

The research was conducted as follows:

1. Participants were tested before treatment through pre-test.
2. The control group learned by traditional method while experiment group was given a treatment in which they were given android tablets to play and learn the science quiz application.
3. Participants were tested after treatment through post-test.

The intervention took for 4 days in two weeks. The data was collected by following ways:

Tools	Type of data
Pre-Test	Quantitative
Post-Test	Quantitative
Student Questionnaire	Quantitative

Table 3: Tools used to collect the data

The data collected was coded in SPSS software and analyzed.

3.4.2. Research Question

The purpose of this study was to determine the effectiveness of formative assessment on learning quality, using interactive science quiz application. The study addressed the following research question:

What is the significance of formative assessment in improving learning quality using interactive science quiz application?

3.4.3. Research Site and Sample

A sample is a smaller and manageable version of a large data set. Samples are used in statistical data analysis when the observation population size is too large (Investopedia, 2016). The sample used for this study consisted of grade 4 students. The study was conducted in two federal government schools i.e. Islamabad Model College for Boys I-10/1 and Islamabad Model College for Girls I-10/4, located in Islamabad during the academic year 2016. The schools were chosen for convenience sampling due to ease of access. Convenience sampling is selecting available individuals as respondents continuously such that either the required sample size is achieved or those accessible respondents cease to be available (Louis Cohen, 2007). Two sections of the participants were chosen randomly as control (n = 60) and experiment group (n = 60). The gender distribution included 60 boys and 60 females. Total number of participants was 120. The particulars of participants are given in the table below:

Participants

Gender	Group	No. of participants	School
Male	Control	30	Islamabad Model College for Boys I-10/1, Islamabad
	Experiment	30	
	Sub-total	60	
Female	Control	30	Islamabad Model College for Girls I-10/1, Islamabad
	Experiment	30	
	Sub-total	60	
Total			120

Table 4: Participants of both schools

3.4.3.1. Control and Experiment Group

Control Group

The control group included 30 male participants and 30 female participants. Research was carried out separately for boys and girls in sister institutions. In both schools, two sections of fourth grade students were randomly selected as control and experiment group. The total number of control group participants was 60.

Participants in the control group were taught first unit of science through conventional method as teachers normally instruct in schools. After teaching the first unit of science textbook, post-test was conducted to measure their learning. There were no instructional changes nor were the participants provided with any treatment. The scores were recorded before and after the instruction, through pre-test (see Appendix C) and post-test (see Appendix D) respectively.

Experiment Group

The experiment group consisted of 30 males and 30 females. Research was conducted separately for boys and girls in sister institutions. In both schools, two sections of fourth grade students were randomly selected as control and experiment group. The number of participants in experiment group was 60.

A pre-test was taken by these 60 participants. In the next step, Android tablets were given to the experiment group participants for science quiz application. Before proceeding to the formative quiz assessment, students had to play a scenario based story regarding human organs in order to help them link the prior knowledge. The scenario based story was tried to be related with real life. Story was added as it is known to be a traditional way of transferring knowledge (Association, 2012).

There were otherwise no major differences between control and experiment groups.

3.4. 4. Research Constraints

There was limited time for the development of science quiz application. The development not merely included coding but also graphics of the application. The implementation had to be done before summer break, so the research had to be conducted by the end of May 2016. It was a bit difficult for the school management too, to take out time for research as the students were about to have first term exams. To save their time, the research was conducted in their science subject period.

3.5. Ethical Issues

The schools chosen were chosen by convenience sampling as they were easy to reach. The heads of both the schools were explained about the intervention and a suitable time was mutually agreed. They signed the consent letter in this regard (see Appendix E). Also, students were asked whether they were willing for the intervention or not and those who participated were volunteers. No personal information from participants was collected.

3.6. Data Collection

In order to collect data about the impact of formative assessment on students' learning quality, control and experiment group were monitored before the study through pre-test and at the end of the study by post-test. The data was recorded using numerical values in Microsoft Excel Spreadsheet. The excel sheet was then used in SPSS 20.

The responses of students about student learning, enjoyment and engagement were collected on likert scale, through a questionnaire (Christopher Cheong, 2013). The tools used for data collection are discussed below:

3.6.1. Tools for Data Collection

3.6.1.1. Pre-Test and Post-Test

To determine the effectiveness of formative assessment, pre-test and post-test (see Appendix C and Appendix D) were used to collect the data before and after the intervention. Both tests comprised of 15 questions from the first unit of 4th grade science text book, named "understanding ourselves".

To validate the pre-test and post-test, three science teachers were asked to fill a questionnaire; two from boys' school and one from girl's school, that confirmed that the tests were understandable to the fourth grade students and were made according to the National Curriculum 2006. The validation form is included in Appendix G.

3.6.1.2. Student Questionnaire

The experiment group was given a questionnaire to collect data about student learning, engagement and enjoyment. The questionnaire (Christopher Cheong, 2013) consisted of 6 engagement questions, 4 enjoyment questions and 5 learning questions (see Appendix F). The responses were recorded on a 5 point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, strongly agree).

Engagement

The purpose of the quiz application was to enhance the learning outcomes by keeping the student engaged with the content. The questions used to check whether the students were engaged or not are:

Sr. No	Questions
Q1	I wanted to complete the quiz.
Q2	I wanted to explore all the options available to me.
Q3	I found the quiz satisfying.
Q4	I felt absorbed in the quiz.
Q5	I felt that time passed quickly.
Q6	I felt excited during the quiz.

Table 5: Questions to evaluate Engagement

The responses for each of the question are shown below:

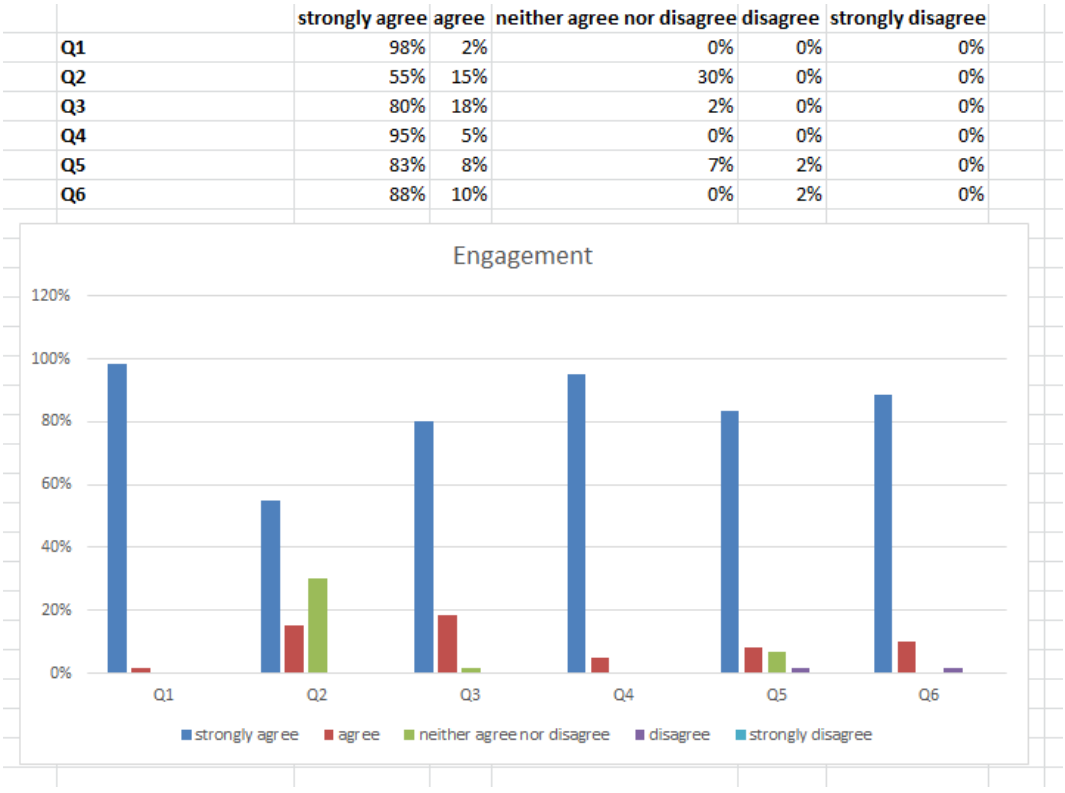


Figure 23: Responses of Engagement Questions

The results of the students' responses showed that almost all the students (98%) agreed that they wanted to complete the quiz. Only half (55%) of the students wanted to explore all the options, while rest of the students wanted to learn and chose correct option. Majority of the students

found the quiz satisfying and felt absorbed in the quiz (80% and 95% respectively). Students were so engaged and excited (88%) with the quiz application that they felt the time passed quickly (83%). Overall responses of the students showed that they were satisfied and absorbed in the quiz application.

Enjoyment

The purpose of adding game elements was to add the factor of enjoyment for students. By this students could have a game like feeling. To evaluate enjoyment the questions used are:

Sr. No	Questions
Q1	I felt exhausted when playing Science Quiz.
Q2	I felt unhappy when playing Science Quiz.
Q3	I felt worried when playing Science Quiz.
Q4	I felt happy when playing Science Quiz.

Table 6: Questions to evaluate Enjoyment

The responses for each of the question are shown below:

	strongly agre	agree	neither agree nor disagre	disagree	strongly disagree
Q1	2%	0%	5%	5%	88%
Q2	0%	2%	0%	5%	93%
Q3	0%	2%	2%	10%	87%
Q4	93%	5%	0%	0%	2%

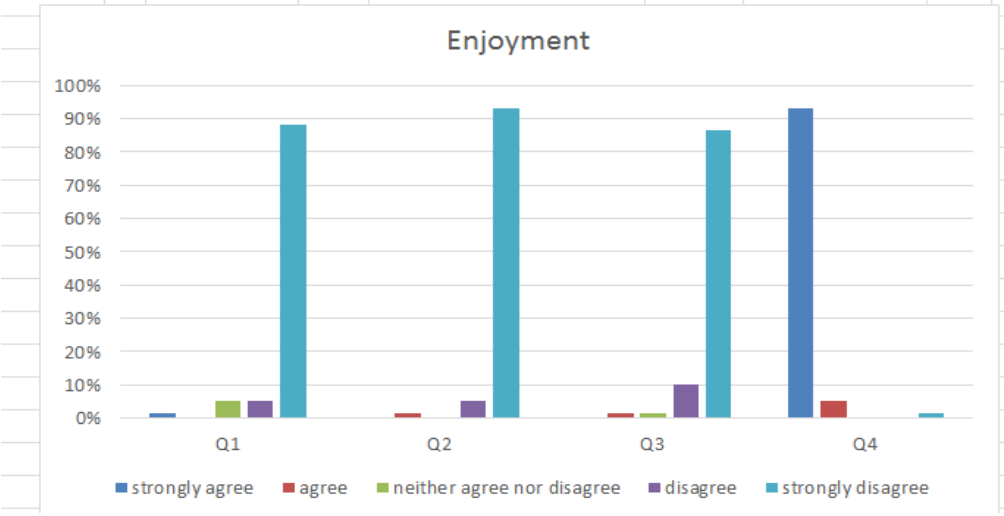


Figure 24: Responses of Enjoyment Questions

The results of the students' responses showed that only a very few (2%) of the students felt exhausted, unhappy and worried while playing the science quiz. Majority (93%) of the students were happy while playing the quiz. The overall responses were positive

Learning

The questions for the engagement are:

Sr. No	Questions
Q1	Playing Science Quiz improves my learning quality.
Q2	Playing Science Quiz enhances my learning effectiveness.
Q3	Playing Science Quiz improves my learning productivity
Q4	Playing Science Quiz helps to get better grades.
Q5	Playing Science Quiz improves my learning performance.

Table 7: Questions to evaluate Learning

The responses for each of the question to evaluate learning are shown below:

	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
Q1	85%	12%	3%	0%	0%
Q2	92%	8%	0%	0%	0%
Q3	75%	23%	2%	0%	0%
Q4	92%	5%	3%	0%	0%
Q5	95%	3%	2%	0%	0%

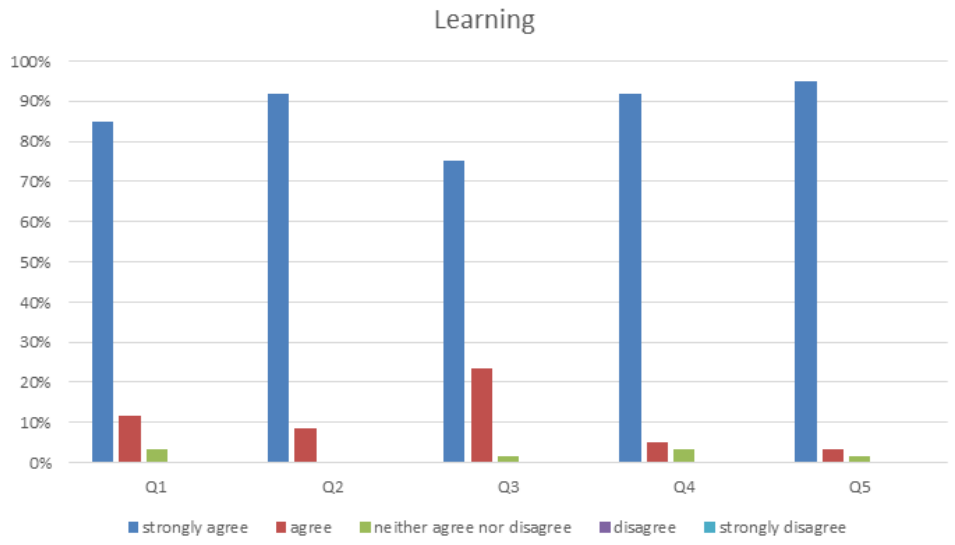


Figure 25: Responses of Learning Questions

The overall results were positive. The students' responses showed that majority of the students strongly agreed that their quality and effectiveness (85% and 92% respectively) of learning was improved. 75% students agreed that the interactive science quiz application has improved their learning productivity. 92% and 95% students agreed that formative assessment based science quiz would help them in improving their performance in class and eventually help them in getting better grades.

It can be concluded from the results that formative assessment based interactive science quiz application could be used as a learning tool that could keep the students engaged and focused towards the content. They could enjoy learning rather than being forced for assessments.

Data Collection Tools

The below table gives an overview of the data collection tools, their purpose and the participants involved in the research:

Data Collection Tool	Purpose	Participants
Pre-Test	To determine what students already know	Control Group Experiment Group
Post-Test	To determine if there was improvement in learning	Control Group Experiment Group
Student Questionnaire	To explore whether participants enjoyed, were engaged with the content and learned anything.	Experiment Group

Table 8: An overview of data collection tools, their purpose and the participants involved

This chapter gave an overview of the application, described usability testing, research design, research question, research site, participants, constraints, ethical issues, and data collection. The following chapter presents the results of this study.

Chapter 4: Results

4.1. Introduction

In order to find out the statistical significance of formative assessment using interactive science quiz, the data was coded and analyzed using IBM's SPSS 20 (Statistical Package for the Social Sciences) statistical software. Statistical significance means that results from a research or experiment are not likely to occur randomly or by chance and instead have a dependent variable that influences them (Investopedia, 2016). Since statistical analysis of data depends upon the number of samples collected, it has inherent degree of error which is negligibly small if the data set is quite large, for instance from the entire population. The researcher needs to define the probability of sampling error in advance for any such tests. The level that is used to assert if an event is statistically significant or not is known as the significance level (Investopedia, 2016). In our statistical data analysis, for testing the statistical significance of research hypothesis we have assumed a significance value or alpha value of 0.05. This means that we have 5% chance of concluding a false positive. According to Wikipedia, an alpha equal to 0.05 means that probability of wrongly rejecting the null hypothesis is 5%. The null hypothesis is usually a hypothesis of no difference (StatsDirect). Null hypothesis or conjecture assumes that any difference or variance between data sets is due to chance and no statistical significance exists in the observation data collected (Investopedia, 2016). The opposite of null hypothesis is alternative hypothesis and is accepted when the null hypothesis is rejected. Calculated probability or P-value is the statistical term that is used in the data analysis to check if the null hypothesis is accepted or rejected. P-value is the probability under which a statistical result occurred by chance or by sampling error (Investopedia, 2016). In order for the result to be statistically significant the p-value should be below the significance level, this means that the chance of the observation being random is less than 5%.

In SPSS the data of control and experiment ground was analyzed for normality to find out how much data set matches the normal distribution. According to the MathWorld wiki, normally distributed data is one with a central value (mean) and most of the values are symmetrically centered around the mean, with the spread determined by its standard deviation. The mean and standard deviation of theoretical distributions are referred as parameters of the distribution. In

statistical data analysis we estimate the parameters of distribution assumed for the data set, such methods that use distributional assumptions are called parametric methods (Altman). In statistical analysis, normal data set is a requirement for parametric tests (Laerd, 2013). Parametric methods assume that data being analyzed follows normal distribution, which means that the population from which the data samples are collected is normally distributed with uniform variance between groups. Populations that are not normally distributed are analyzed using non-parametric methods. Non-parametric methods are commonly used in data analysis when data collected does not meet the distributional requirements of parametric methods (BMJ 2009;338:a3167).

In order to determine the impact of formative assessment, the gathered data was compared:

1. Between groups: (a) pre-test of control and experiment group (b) post-test of control and experiment group were compared.
2. Within groups: (a) pre-test/post-test of control group and (b) pre-test/post-test of experiment group were compared.
3. Between boys and girls (after invention)

Based on the theoretical review on statistical data analysis with non-parametric methods we used Mann-Whitney U Test and Wilcoxon Signed-Rank on population samples for comparison between groups and comparison within the group respectively.

Mann-Whitney U Test

It is a non-parametric test of null hypothesis that is used to check how likely it is that a random value from one sample set is less than or greater than a random value from another sample set. Our data set passed the following assumptions, as listed in Mann-Whitney U test in SPSS (Laerd, 2013).

- Sample scores in control and experiment group are independent of each other
- Distribution of scores for both groups of independent variables (control and experiment) are symmetrical in shape
- Samples scores are measured at continuous level (0-10)

Wilcoxon Signed-Rank

It is a non-parametric hypothesis test that is used to check how much the mean rank differs for two related samples or two matched samples. In our work we used it for statistical analysis of samples within the group while satisfying the assumption of use as listed in Wilcoxon Signed-Rank Test using SPSS (Laerd, 2013).

- Sample score are measured at continuous level (0-10)
- Independent variable consists of two related groups (pre-test experiment and post-test experiment) to indicate same subjects are present in both groups before and after intervention.
- Distribution of scores between independent variables needs to be symmetrical in shape

Comparison between Groups

Pre-test

Data of pre-test was examined to determine if there was statistical significant difference between control and experiment group. The non-parametric method, Mann-Whitney U Test was used to compare the mean rank of pre-test results. The significance level or p-value was set at < 0.05.

Post-test

In order to determine the impact of formative assessment using interactive science quiz, the data of post-test was examined to determine if there was statistical significant difference between control and experiment group. The non-parametric method, Mann-Whitney U Test was utilized to compare the mean rank of post-test results. The significance level or p-value was set at < 0.05.

Comparison within Group

Control Group

The control group was taught in a traditional manner. The data collected from control group was examined to determine if there was statistical significant difference between pre-test and post-test. The non-parametric method, Wilcoxon Signed-Rank Test was used to compare the median difference between pre-test and post-test scores. The significance level or p-value was set at < 0.05.

Experiment Group

In order to examine the impact of formative assessment using interactive science quiz, the results of pre-test and post-test of experiment group were analyzed using Wilcoxon Signed-Rank Test, to determine if there was statistical significant difference between pre-test and post-test. The significance level or p-value was set at < 0.05.

Gender Based Comparison of Scores

In order to determine if there was a statistical significant difference between pre-test and post-test of males and females of experimental group, a non-parametric test, Mann-Whitney U Test, was utilized. The significance level or p-value was set at < 0.05. The effect of independent variable on dependent variable (scores) was determined in these comparisons.

4.2. Purpose of the study

The purpose of this study was to determine the impact of formative assessment in enhancing learning quality, using interactive science quiz application at primary level. This study addressed the following questions:

4.3. Research Question

What is the significance of formative assessment in improving learning quality using interactive science quiz application?

Null Hypothesis

H0: There is no significant impact of science quiz application on student learning outcomes using interactive science quiz application.

Alternative Hypothesis

H1: There is a significant impact of science quiz application on student learning outcomes using interactive science quiz application.

4.4. Test of Normality

In order to check whether the data was normal or not, a test of normality was conducted, using SPSS 20 software. There are two tests in SPSS which determine if the data is normally distributed or not i.e Kolmogorov-Smirnov Test and Shapiro-Wilk Test. Both tests compare the data being analyzed with the quantiles of normal distribution in order to assess that data is likely from normal distribution or not. Shapiro-Wilk test however is considered to be powerful than Kolmogorov-Smirnov Test (Zahediasl, 2012). It can handle sample sizes up to 2000. In this study the sample size was of 120 students, hence Shapiro-Wilk test was used for normality check.

From SPSS data the p-values of the Shapiro-Wilk test for control group's pre-test and post-test were less than 0.05. Hence it was concluded that control group data was not normally distributed. Therefore, we had to use non-parametric methods of data analysis. The normality results of control group's pre-test and post-test are shown in the table below:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Control_Pre_Test_Score	.165	60	.000	.947	60	.012
Control_Post_Test_Score	.168	60	.000	.940	60	.005

a. Lilliefors Significance Correction

Table 9: Test of normality for control group's pre-test and post-test

Similarly, from the SPSS data the p-values of the Shapiro-Wilk test for experiment group's pre-test and post-test were less than 0.05. Hence it was concluded that data was not normally distributed. Therefore, we used non-parametric methods of data analysis. The normality results of experiment group's pre-test and post-test are given in the table below:

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Experiment_Pre_Test_Score	.134	60	.009	.948	60	.013
Experiment_Post_Test_Scores	.171	60	.000	.897	60	.000

a. Lilliefors Significance Correction

Table 10: Test of normality for Experiment Group's Pre-test and Post-test

4.5. Null Hypothesis

To determine if formative assessment based science quiz application had an impact on student learning outcomes, the pre-test scores of control group were compared with the pre-test scores of experiment group while the post-test scores of control group were compared with the post-test scores of experiment group. The data collected through pre-test and post-test was validated by science teachers. This data was analyzed in SPSS 20 statistical software. A non-parametric test, Mann Whitney U Test, for the comparison of scores between groups was used. The significance level was set at 0.05.

The null hypothesis: "there is no significant impact of science quiz application on student learning outcomes using interactive science quiz application", was rejected with a significance level set at < 0.05. The results of Mann Whitney U Test concluded that there was a significant impact of formative assessment in improving scores in science at primary level.

The null hypothesis was rejected in following ways:

4.5.1. Comparison between Groups

4.5.1.1. Demographics of the participants

To reject the null hypothesis, the data of 60 students of control group and 60 students of experiment group was analyzed. The table below shows the demographics of the participants:

Group	No. of participants	Gender	Grade
Control	60	Male Female	4th Grade
Experiment	60	Male Female	
Total	120		

Table 11: Demographics of Participants of research question

4.5.1.2. Test to compare scores between Control and Experiment Group

Pre-Test

A Mann-Whitney U Test was run to determine if there were differences in pre-test score between control and experiment group. The effect of independent variable (group) was measured on dependent variable (score). The results showed a significance value of 0.31 with U= 1,395.000 and z = -2.153. The results are shown in the table below:

Independent-Samples Mann-Whitney U Test

Total N	120
Mann-Whitney U	1,395.000
Wilcoxon W	3,225.000
Test Statistic	1,395.000
Standard Error	188.068
Standardized Test Statistic	-2.153
Asymptotic Sig. (2-sided test)	.031

Table 12:Mann Whitney U Test results for pre-test of control and experiment group

It was assessed by visual inspection that distributions of the pre-test score for control and experiment group were symmetrical. There was less difference in the median of both groups. The median of pre-test score for control group was 6.00 and that of experiment group was 5.50, with a significance of 0.31.

Median

Group	Pre_Test
Control	6.00
Experiment	5.50
Total	6.00

Table 13: Median of pre-test for control and experiment group

The hypothesis test summary as illustrated below, shows that the distribution of pre-test score was not same in both groups, there was a slight difference between pre-test scores of control and experiment group.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Pre_Test is the same across categories of Group.	Independent-Samples Mann-Whitney U Test	.031	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 14: Hypothesis Test Summary of pre-test

The figure below compares the mean rank of control and experiment group. The pre-test mean rank for control group was 67.25, which was higher than experiment group which had a mean rank of 53.75.

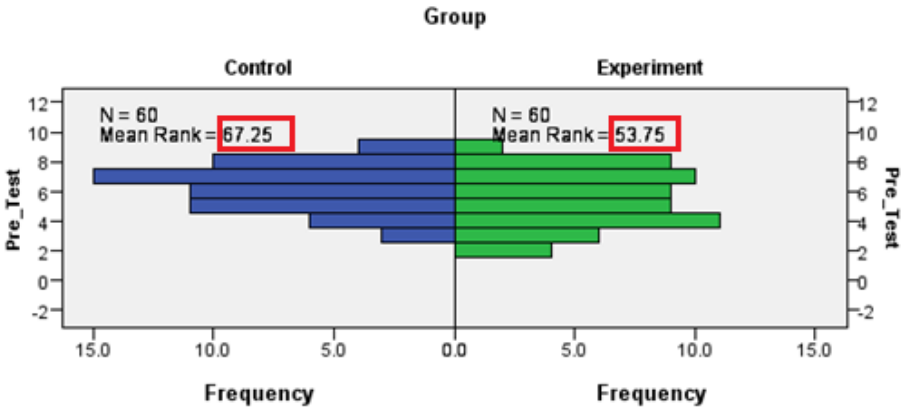


Figure 26: Comparison of control and experiment group's pre-test

Post-Test

In order to reject the null hypothesis "there is no significant impact of science quiz application on student learning outcomes using interactive science quiz application", a non-parametric test, Mann-Whitney U Test was run on post-test data. The effect of independent variable (group) was measured on dependent variable (score). The results of Mann Whitney U Test are given in the table below. The results showed a significance of 0.001 that is less than 0.05, with U= 3599.500 and z = 9.503. The significance value rejected the null hypothesis and accepted the alternative hypothesis.

Independent-Samples Mann-Whitney U Test

Total N	120
Mann-Whitney U	3,599.500
Wilcoxon W	5,429.500
Test Statistic	3,599.500
Standard Error	189.364
Standardized Test Statistic	9.503
Asymptotic Sig. (2-sided test)	0.001

Table 15: Mann Whitney U Test results for post-test of control and experiment group

It was assessed by visual inspection that distributions of the post-test score for control and experiment group were symmetrical. Median post-test score was statistically significantly higher in experiment group (13.00) than in control group (7.00).

Median

Group	Post_Test
Control	7.00
Experiment	13.00
Total	10.00

Table 16: Median of post-test for control and experiment group

The hypothesis test summary is illustrated below shows that the distribution of post-test score was not same in both groups, and there was a significant difference between post-test scores of control and experiment group. The significance value is less than 0.05 that rejected the null hypothesis.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
2	The distribution of Post_Test is the same across categories of Group.	Independent-Samples Mann-Whitney U Test	.001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 17: Hypothesis Test Summary of post-test

The graph below compares the mean rank of control and experiment group. The post-test mean rank for experiment group was 90.49, which was statistically significantly higher than control group that had a mean rank of 30. 51. This showed that there is a significant impact of formative assessment using interactive science quiz application. It was supported by results that there is a significant difference in students' score between post-test of control and experiment group

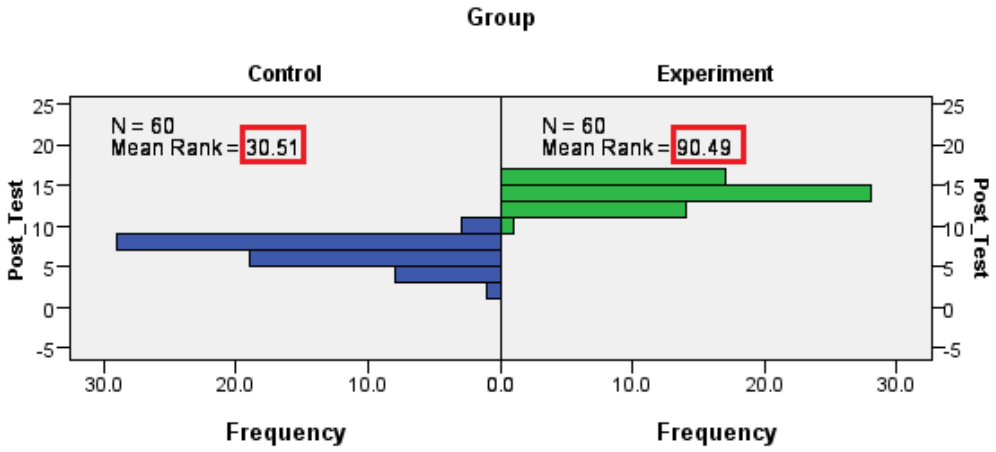


Table 18: Comparison of control and experiment group's post-test

The following figure gives an overview of the mean rank of pre-test and post-test of both control and experiment group. The mean rank of control and experiment group's pre-test was 67.25 and 53.75 respectively. While the mean rank of control and experiment group's post-test was 30.51 and 90.49 respectively. The collected data rejected the null hypothesis, which showed that there

was a strong impact of formative assessment, using interactive science quiz application, in enhancing learning outcomes of students.

	Group	N	Mean Rank	Sum of Ranks
Pre_Test	Control	60	67.25	4035.00
	Experiment	60	53.75	3225.00
	Total	120		
Post_Test	Control	60	30.51	1830.50
	Experiment	60	90.49	5429.50
	Total	120		

Table 19: Mean rank of pre-test and post-test of control and experiment group

4.5.2. Comparison within Groups

To answer the research question, and reject the null hypothesis, the difference of scores within control group and within experiment group was determined. The pre-test scores and post-test scores of control group were compared to prove that there was no significant improvement of scores after teaching with conventional method. While the pre-test scores of experiment group were compared with the post-test scores of experiment group to determine the impact of interactive science quiz application. The data collected through pre-test and post-test was validated by science teachers. The data was analyzed in SPSS 20 statistical software. The non-parametric test, Wilcoxon Signed-Rank test was used to compare the scores within groups. The significance level was set at 0.05. The data was compared (1) within control group and (2) within experiment group.

The null hypothesis: “there is no significant impact of science quiz application on student learning outcomes using interactive science quiz application”, was rejected with a significance value < 0.05. The results of Wilcoxon Signed-Rank test concluded that there was a significant increase in students’ scores after exposing them to formative assessment science quiz application.

4.5.2.1. Test to compare Scores within-groups

4.5.2.1.1. Within Control Group

Demographics of the participants

Group	Gender	No. of participants	Grade
Control	**Male**	30	4th Grade
	Female	30	
Total		60	

Table 20: Demographics of the Participants for research question

The table above shows the demographics of the participants. The control group consisted of 60 participants: 30 males and 30 females. To compare the control group's pre-test scores with post-test scores, data of 60 students of control group was analyzed.

To test the null hypothesis, "there is no significant impact of science quiz application on student learning outcomes using interactive science quiz application", this test analyzed a sample of 60 participants of control group. At the start of experiment, the dependent variable (score) was measured by taking a pre-test. The students were taught through conventional method, there was no intervention. The same dependent variable was measured again for these 60 participants by taking post-test. The effect of independent variable (group) was measured on the dependent variable (score). The group had two related groups:

Related Group # 1: immediately at the start of the manipulation (pre-test)

Related Group # 2: after teaching with the conventional method (post-test)

A non-parametric test, the Wilcoxon Signed-Rank test was used to determine whether there was a median difference between pre-test and post-test scores of control group. From SPSS data the statistical significance (p-value = 0.448), showed that the median difference between the two related groups was not statistically significantly different. The results of difference in pre-test scores and post-test scores of control group are shown in the table below:

Total N	60
Test Statistic	487.500
Standard Error	75.135
Standardized Test Statistic	.759
Asymptotic Sig. (2-sided test)	.448

Table 21: Wilcoxon Signed-Rank test's results of control group's pre-test and post-test

The significance value (p = 0.448) came out to be greater than the significance level of 0.05. The hypothesis test summary of control group generated by Related-samples Wilcoxon Signed-rank test is given in the table below:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between pre_test_score and post_test_score equals 0.	Related-Samples Wilcoxon Signed Rank Test	.448	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 22: Hypothesis Test Summary of comparison of pre-test & post-test within control group

The median difference of pre-test and post-test scores within control group came out to be 0.0000, as illustrated in the table below:

Median

pre_test	post_test	difference
6.00	7.00	.0000

Table 23: Median difference of control group's pre-test and post-test

The results of Wilcoxon Signed-Rank test showed that there was no improvement in scores of control group after teaching the participants with traditional method. The test generated a histogram which illustrated that out of 60 participants of control group, 18 participants obtained fewer score in post-test as compared to pre-test. 23 students showed positive results, with an increase of one or two score and 19 students had same scores in post-test, as in pre-test. Hence it

was concluded that there was no improvement in scores of control group after teaching them the first unit of science by conventional method. The histogram is illustrated below:

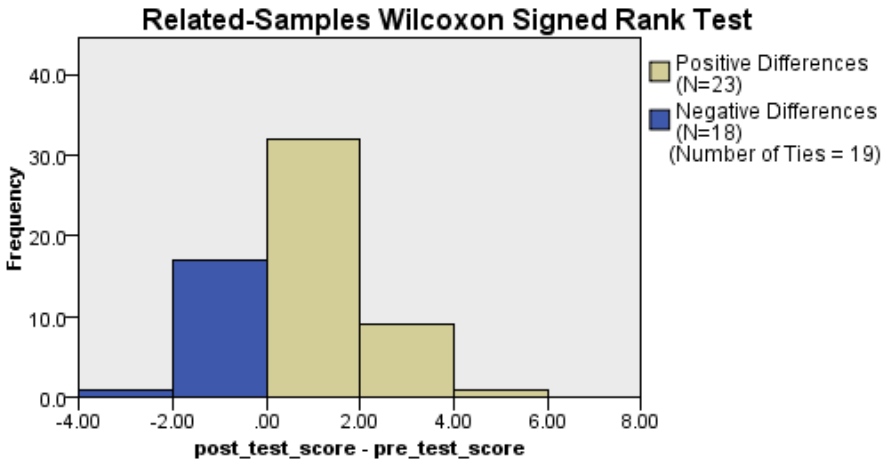


Figure 27: Histogram of difference between Pre-Test score and Post-test score of control group

The results obtained indicate that there was no significant difference in control group's pre-test and post-test score. The mean for pre-test (M = 6. 25) was nearly equal to post-test mean (M=6.42). Hence it was concluded that there was no significant difference in students' scores between pre and post-test of a control group, with a significance value of 0.448.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
pre_test_score	60	6.25	1.590	3	9
post_test_score	60	6.42	1.700	2	10

Table 24: Results of Wilcoxon Signed-Rank Test of Control Group

4.5.2.1.2. Within Experiment Group

Demographics of the participants

Group	Gender	No. of participants	Grade
Experiment	Male	30	4th Grade
	Female	30	
Total		60	

Table 25: Demographics of the Participants

The table above shows the demographics of the participants. The experiment group consisted of 60 participants: 30 boys and 30 girls. To compare the experiment group's pre-test scores with post-test scores, data of 60 students of experiment was analyzed.

To test the null hypothesis, "there is no significant impact of science quiz application on student learning outcomes using interactive science quiz application", the data of 60 participants of experiment group was analyzed. At the start of experiment, the dependent variable (score) was measured by taking a pre-test. The same dependent variable was measured again for these 60 participants after intervention by taking post-test. The effect of independent variable (group) was measured on the dependent variable (score). The group had two related groups:

Related Group # 1: before intervention (pre-test)

Related Group # 2: after intervention (post-test)

Total N	60
Test Statistic	1,830.000
Standard Error	135.452
Standardized Test Statistic	6.755
Asymptotic Sig. (2-sided test)	0.001

Table 26: Wilcoxon Signed-Rank test's result of experiment group's pre-test and post-test

A non-parametric test, the Wilcoxon Signed-Rank test was used to determine whether there was a median difference between pre-test and post-test scores of experiment group. The significance value came out to be 0.0005, with z = 6.755. From SPSS data, the statistical significance (p-value = 0.001), showed that the median difference between the two related groups was statistically significantly different. The results of difference in pre-test scores and post-test scores of experiment group are shown in the table above.

The significance value (p = 0.001) came out to be less than 0.05, the null hypothesis is rejected. The hypothesis test summary of experiment group generated by Related-samples Wilcoxon Signed-Rank test is given in the table below:

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between score_before_intervention and score_after_intervention equals 0.	Related-Samples Wilcoxon Signed Rank Test	0.001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 27: Hypothesis Test Summary of comparison of pre-test & post-test within experiment group

The median difference of pre-test and post-test score within experiment group was 8.0000, as illustrated in the table below. The median of score before intervention was 5.50 and the median of score after intervention was 13.00.

Median

score_after_intervention	score_before_intervention	difference
13.00	5.50	8.0000

Table 28: Median difference of experiment group's pre-test and post-test score

The results of Wilcoxon Signed-Rank test showed that there was a significant improvement in scores of experiment group after exposing them to formative assessment based interactive science quiz application. The test generated a histogram which illustrated that all the 60 participants of experiment group, showed a significant increase in scores.

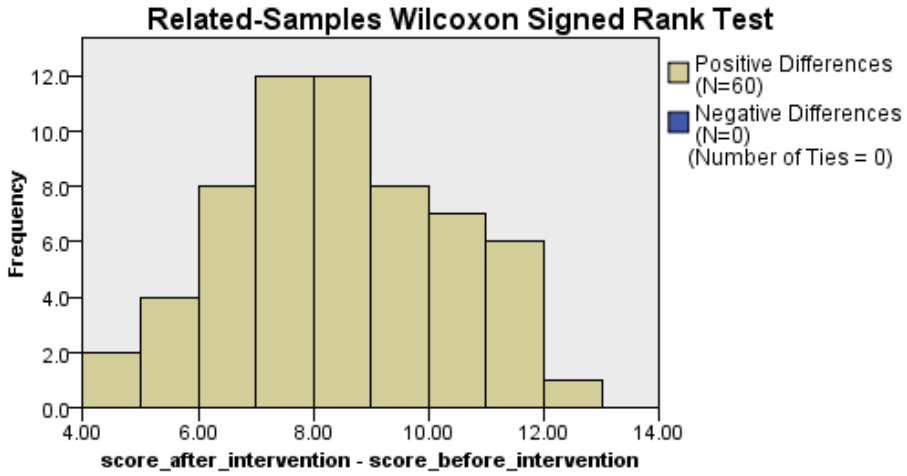


Figure 28: Histogram of difference between Pre-Test score and Post-test score of experiment group

There were no participants who showed negative results or gained the same marks as obtained in pre-test. Hence it was concluded that there was a significant impact of formative assessment in enhancing learning quality, by using interactive science game based quiz. The conclusion can be drawn that there was a statistical significant improvement in experiment group's pre-test and post-test score, after intervention.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
score_before_intervention	60	5.48	1.927	2	9
score_after_intervention	60	13.42	1.306	10	15

Table 29: Results of Wilcoxon Signed-Rank Test of experimental group

The mean for pre-test was 5.48 while post-test mean was 13.42. Hence there was a statistical significant difference between scores of pre-test (before intervention) and post-test (after intervention) of experiment group. The null hypothesis "there is no significant impact of science quiz application on student learning outcomes," was rejected with a significance value of 0.001.

4.6. Gender Disparity in student's score

To determine the gender difference in students' score, after being exposed to an interactive science quiz application, experiment group's pre-test and post-test scores were compared. The data collected through pre and post-test was validated by science teachers. The data was analyzed in SPSS 20 statistical software. A non-parametric test, Mann Whitney U Test, for comparison of scores between groups was used. The significance level was set at < 0.05.

It was found that "there is a significant gender difference on students' scores after being exposed to an interactive science quiz application".

Demographics of the participants

Group	**Gender**	**No. of participants**	**Grade**
Experiment	Male	30	4th Grade
	Female	30	
Total		60	

Table 30: Demographics of Participants of research question

The above table shows the demographics of the participants. The data of 60 students of experiment group, including 30 males and 30 females, was analyzed.

4.6.1. Gender Based Comparison of scores

A Mann-Whitney U Test was run to determine if there were differences in pre-test and post-test scores of males and females of the experiment group. The effect of independent variable

(gender) was measured on dependent variable (score). The scores of males were compared with the scores of females for pre-test and post-test respectively.

Pre-Test

In order to determine the differences in pre-test score between males and females, Mann Whitney U Test was utilized. The effect of gender was measured on pre-test scores. The statistical significance (p = 0.139) was greater than 0.05, which showed that the difference of scores between males and females before intervention was not statistically significantly different, U = 549, z =1.480.

Total N	60
Mann-Whitney U	549.000
Wilcoxon W	1,014.000
Test Statistic	549.000
Standard Error	66.891
Standardized Test Statistic	1.480
Asymptotic Sig. (2-sided test)	.139

Table 31: Mann Whitney U Test results of gender comparison for pre-test

The hypothesis test summary as illustrated below shows that the distribution of pre-test score before intervention was same for males and females. The significance value (p = 0.139) was > 0.05, hence null hypothesis is retained.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of score_before_intervention is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	.139	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 32: Hypothesis Test Summary of pre-test for males and females

The mean ranks of female before intervention was 33.80, which was higher than the mean rank of males, which were 27.20. From results it was concluded that there was no statistically significant difference of pre-test scores between males and females.

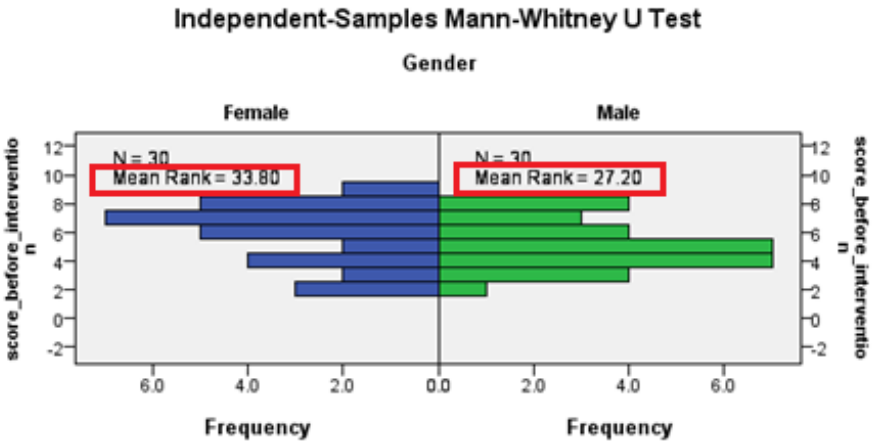


Figure 29: Comparison of pre-test scores for males and females

Post-Test

To determine the gender differences in post-test scores, Mann Whitney U Test was used. The effect of independent variable (group) was measured on dependent variable (score). The statistical significance (p = 0.001) shows that the difference of scores between males and females after intervention was statistically significantly different, with U = 689, z = 3.640. The results are shown in the table below:

Total N	60
Mann-Whitney U	689.000
Wilcoxon W	1,154.000
Test Statistic	689.000
Standard Error	65.659
Standardized Test Statistic	3.640
Asymptotic Sig. (2-sided test)	0.001

Table 33: Mann Whitney U Test results for post-test of control and experiment group

The hypothesis test summary as illustrated below shows that the distribution of post-test score after intervention was statistically significantly different. The significance value (p = 0.001) was < 0.05, hence null hypothesis was rejected.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
2	The distribution of score_after_intervention is the same across categories of Gender.	Independent-Samples Mann-Whitney U Test	0.001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Table 34: Hypothesis Test Summary of post-test for males and females

The histogram given below shows the mean rank of females after intervention was 38.47, which was statistically significantly higher than the mean rank of males, that was 22.53. The results of Mann Whitney U Test indicated that overall females performed better than males.

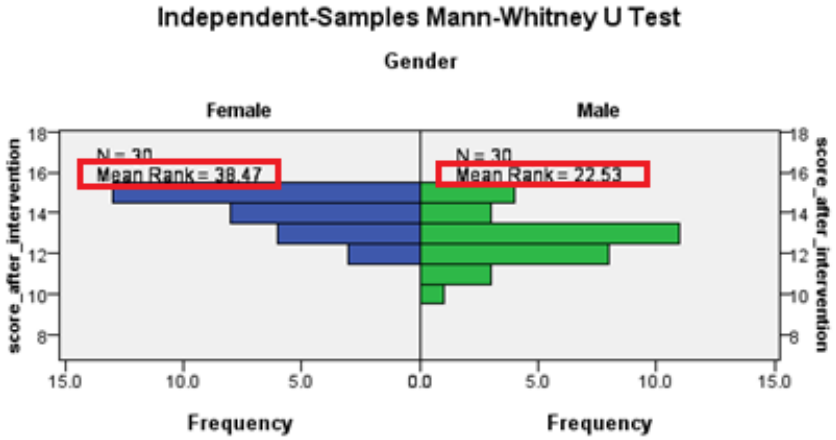


Figure 30: Comparison of post-test scores for males and females

The resulted data showed that there was a significant difference in scores of males and females of experiment group after being exposed to formative assessment quiz application.

The purpose of this study was to determine the impact that formative assessment had in enhancing learning outcomes of grade four students using an interactive science quiz application, developed in accordance with the National Curriculum for General Science. The study results showed that formative assessment ameliorates student's learning, understanding, motivation, engagement and performance.

Chapter 5: Discussion

5.1. Discussion

Assessment is an essential component of education for improved learning. Research done in this regarding acclaims the benefits of formative assessment when used as standalone assessment or when used to support the summative assessments. Black and William were pioneers in the literature review when they brought to light the key attributes of formative assessment that were found to enhance learning: effective feedback, self-assessment, students' involvement in learning, modifying ways of teaching and recognizing the effect of assessment on students' self-esteem (Dylan, 1998). Formative assessment is an ongoing assessment that improves learning through efficient feedback and giving ownership of their work to students (Abiodullah, 2009). Literature review done as part of the current study shows similar notions about the use of formative assessment practices in the classroom environment. In addition to this, use of computer based assessments as an enabler has also proved its significance in today's ICT dominated world. Academia can use more powerful tools for effective learning than the traditional methods. Gamification has added another dimension to learning that is found to be more engaging and encouraging for students learning progression. While the past research work has indeed established formative assessment and its associated attributes to be really effective in improving the learning quality and learning philosophy, we have examined its efficacy in the context of Pakistan's primary education milieu in conjunction with gamification of science quiz assessment.

5.2. Purpose of this study

The purpose of this study was to determine the impact that formative assessment had in enhancing learning outcomes of grade four students using an interactive science quiz application, developed in accordance with the National Curriculum for General Science. The study presented an overview of Pakistan's educational system current state of affairs and how it lacks in fulfilling the learning needs of students. The research design and methodology used formative assessment attributes to shape the core manuscript from science book in order to investigate the research question. The study results showed that formative assessment ameliorates student's learning, understanding, motivation, engagement and performance.

5.3. Findings

Our research question “What is the significance of formative assessment in improving learning quality using interactive science quiz application?” was established into one null hypothesis.

5.3.1. Null Hypothesis

The null hypothesis “there is no significant impact of science quiz application on student learning outcomes” was analyzed by applying the non-parametric Mann-Whitney U at significance level < 0.05.

The null hypothesis was rejected by comparing scores between groups and within groups:

Between Group Comparison

The results of Mann-Whitney U test showed that post-test data of control group and post-test data of experiment group not only differed in score distribution but also the high mean score values of experiment group lead to the decision that the null hypothesis is rejected. The result established that the use of formative assessment using game based science quiz application indeed impacted student academic efficacy.

Within Group Comparison

The null hypothesis “there is no significant impact of science quiz application on student learning outcomes” was analyzed using the non-parametric Wilcoxon Signed-Rank test at significance level < 0.05. The results of Wilcoxon Signed-Rank test showed that post-test scores of experiment group were statistically significantly different and large improvement in experiment group’s scores established that the null hypothesis is rejected. From the results, we can recognize the effectiveness of using interactive science quiz based on formative assessment attributes.

The findings from the above hypothesis suggest that student performance significantly improved in comparison to the traditional classroom setting. The use of formative attributes in learning as underlined in literature review of earlier research work proved to be an important catalyst in increasing the understanding and cognitive thinking of students. A culture of success can be developed using formative assessment practices in the classroom by concentrating on specific learning needs (Wiliam P. B., 1998). The gamification of the science quiz application that was

used as an implementation of formative attributes, i.e. engagement, motivation and multiple attempts to grasp the concepts proved to be an effective tool. Games have the ability to sustain engagement and motivation over the period of time, which is specifically useful when the learning tasks are challenging (Katie Larsen McClarty, 2012). The use of rewards to gauge the learning progress in the application design supported self-assessment. Self-assessment helps students develop an understanding of learning goals and where they stand on the learning continuum (Wiliam P. B., 1998). The use of immediate descriptive feedback in the assessment application design helped the students to identify the areas that needed more understanding and attention in order to well develop their thought process and resolve disparities as Black & William highlighted in their findings. Formative feedback and peer assessment results in co-ownership in the learning and development of students (Ahmed, 2015). Further the quiz application designed didn't limit the assessment time for students, which enabled the students to think and then answer the questions without the stress of being timed out. There are possibilities for continued practice and encouragement as games allow replay and advancement to next levels (Katie Larsen McClarty, 2012). The use of multiple choice questions and drag and drop options in the assessment also invalidated rote learning that is in common practice in the schools in Pakistan. Learners and learning outcomes are not considered in designing the assessments, instead rote learning and memorization of manuscripts is encouraged in Pakistan's educational system (Khattak, 2012). The application design goal was instead to instigate thinking and reasoning that proved to enhance learning and student performance. It is not the technology alone but how teachers and students utilize it makes it an important learning tool (Norbert Pachler C. D., 2009). Our study further proves that the use of technology in the assessment has the potential to diversify the assessment methods being used in Pakistan, which are mainly summative in nature. The positive results from this digital assessment show that learning methods and teaching strategy can be enhanced by using new innovations in education.

5.3.1. Gender Disparity

By analyzing the data for null hypothesis, it was found that "there was a significant gender difference in students' scores after being exposed to an interactive science quiz application". The non-parametric Whitney U Test was applied on experiment group's male and female student scores at significance level < 0.05. The results showed that there was no statistically significant difference between male and female scores before intervention but there did exist a statistically

significant difference between male and female scores after the intervention. Female students were found to be more effective learners than their male counter parts. The reason for this disparity between male and female scores after invention is due to the fact that male students were not as objective in their assessment as their female peers. The male students were more focused on the playful aspect of the game application and android tablet rather than the assessment content, while the female students were more influenced by the content of the assessment application. When male students use a computer based assessment system they are influenced by its playfulness, usefulness and social influence. On the other hand, when female students use a computer based assessment system they are influenced by ease of use, content and playfulness (A.Economides, 2011).

From the research design, methodology and corresponding results of this study we can emphasis the use of formative assessment in primary classroom setting. Our work establishes that in the current learning milieu of primary schools in Pakistan, use of formative assessment is indeed valuable and can improve learning outcomes of students. The student performance can foster in comparison to the traditional ways of teaching and learning that prevail presently. In an interactive learning environment students felt more motivated and involved in the learning process and their ownership towards their learning goals enhanced. Teachers and instructors in Pakistan can utilize these attributes of formative assessment in contrast to summative assessment, in their classroom to foster learning and develop an inherent interest in education. As pointed out in the problem statement the educational milieu in Pakistan is considerably lacking compared to the internationally set standards of minimum education. Summative assessment is the commonly practiced mode of assessment in which learning targets and achievements of students are in clear disparity. Thus there is a need to introduce reforms and make education friendly policies on the national level to addresses the current deficiencies. The work done in this study is not only an eye opener but brings a lot of possibilities and opportunities for educators and learners to improve their learning skills, attain achievement, accelerate knowledge and develop true understanding of their learning targets. The literature review of the formative ways of learning and the formative characteristics used in the research design of this study clearly establish formative assessment and its attributes as a gate keeper in education. Engagement, interaction, collaboration, self-assessment, descriptive feedback, motivation and self-esteem are the colors of formative assessment that are vivid in research and practice of education. Our study

added to the existing findings and augmented it further with the use of gamified quiz application for assessment in the primary schools of Pakistan. Specifically, for Pakistan where technology is not as widely used as in the developed countries, the use of game based application for assessment is one example of how much potential the information and communication technology possesses to improve the standard of education, which is in complete agreement with earlier research work. The use of tablets and handheld devices is getting popular day by day and so is their availability and affordability. From our results, we believe that if game based assessments become readily available for different courses, students will actively participate in achieving their learning objectives. Also, teachers will be able to monitor the student progress and optimize the course work and assessment design accordingly. This can bring a shift in the pedagogy and assessment practices, which today are dominated by summative assessment. Primary education is compulsory in Pakistan and it is therefore really important to engage students well into the educational philosophy instead of demotivating them or proving them as failure. From our study, we establish that formative assessment is imperative to the learning improvement, self-efficacy and motivation of students. It considerably fosters learning outcomes and understanding of students, therefore it should be practiced in the education sector of Pakistan.

Chapter 6: Conclusion

6.1. Conclusion

The current assessment methods are mainly summative and have not contributed competently towards the learning goals and achievement of students. According to the National Assessment Report, 98% of the teachers reported student learning disabilities during lessons. The reasons behind these deficiencies are lack of interest and motivation, which have aggravated the student achievement. From the literature study it is known that student motivation and engagement are neglected when only summative assessments are used. This research conclusion is equally true in the context of our country, where summative assessment is the main and most practiced form of assessment. The results are devastating as students in Pakistan quit their quest for learning when they establish that they do not have the ability to learn and perform. The review of literature on education challenges and improvements asserts formative assessment as an effective teaching and learning method that can positively impact abilities of students. The purpose of this study was to investigate the effectiveness of formative assessment in primary classroom setting using interactive science quiz application. The results of this study suggests that academic efficacy and learning outcomes of grade 4 students considerably improved after the introduction of formative ways of learning. The data collected during this research has added to the already established belief that formative assessment positively impacts the student learning by enhancing self-esteem, motivation, confidence and cognitive thinking of students. The use of formative method in contrast to the traditional learning method proved really effective as student learning quality improved. Continuous descriptive feedback and multiple attempts as an essence of formative assessment provided the opportunity to grasp the challenging concepts, which are otherwise difficult to develop in a traditional classroom setting. Additionally, technology as a means to introduce formative intervention in education was further augmented by this study.

6.2. Limitations of the research

A limitation of this research was the time limit to develop and implement the quiz application in schools. The research had to take place before the first term exams and summer break in primary schools. The interactive science quiz development did not merely include coding but also

incorporated storyboarding, multimedia (sound and graphics) and content development, which took plenty of time. This study was conducted on primary students (male and female) of grade 4. Extension of study to different grade levels would have required additional time and resources.

Secondly, the interactive science quiz application could not cover all the units of science textbooks for grade 4. There are total ten units in the textbook. But the formative assessment quiz application was developed according to the first unit of general science textbook.

Thirdly, there were not enough Android tablets available to conduct the assessment on large population. The data collected in this study utilized 20 tablets and used them iteratively to record assessment response from 120 primary school students.

In addition to this, not all the students were alike in the use of Android tablets. The non-uniformity in the ease of use raises the question whether this study was able to effectively measure the qualitative and quantitative data that may change over time.

Last but not the least, not all the primary schools in Islamabad were available to conduct this study due to managerial issues. So, we are really thankful to the Islamabad Model College for Boys (IMCB) I-10/1 and Islamabad Model College for Girls (IMCG) I-10/4 for helping us with our research work.

Chapter 7: Recommendations

7.1. Recommendations

In this study we have investigated the impact of formative assessment on primary school students' (grade 4) learning outcomes using an interactive science quiz application. Due to the time limitation, the study focused on primary school and specifically on the grade four students. Some recommendations are given below:

7.1.1. Recommendation 1: Replicate the study across other subjects and multiple units

We recommend that this study should be replicated in secondary school and other lower and higher grades. Additionally, since the study was done on science subject and only one unit of science textbook was used as the core text for content development, we recommend to replicate the study across other subjects and multiple units in order to further assess the use of formative assessment on learning outcomes.

7.1.2. Recommendation 2: Extend the study into more schools

The study was done by collecting sampling data from two schools only, due to managerial issues and limited time frame, so we recommend to extend the study into more schools. This will increase the population size for the statistical data analysis and further refine the results to develop better understanding regarding the implementation of formative methods in Pakistan's education milieu.

7.1.3. Recommendation 3: Continue the study longer

The duration of intervention during this study was four days in two weeks. We recommend to continue the study longer so both teachers and learners have enough time to get familiarized with the formative learning process and their perception about its efficacy and usefulness becomes mature.

7.1.4. Recommendation 4: Professional development of Teachers and Educators

Another aspect that should be explored in future research is the professional development of teachers and educators in Pakistan regarding the formative assessment. As we have discussed in

the literature review, it is a pre-requisite to develop the philosophy of formative practices in the classroom education as teacher are the facilitators and enablers of this change. In our authoritarian culture, the classroom environment is inherently dominated by the attitude of teachers and there is a need to evolve and improve it in order to orchestrate productive and effective learning. The study should review the current deficiencies and conclude what professional development scope for teachers is vital to the establishment of an efficient learning setting.

7.1.5. Recommendation 5: Assess the use of video films and multimedia technologies to improve the understanding and broadening the vision of the students

From technological innovations in education point of view, we recommend to assess the use of video films and other multimedia technologies in improving the understanding and broadening the vision of the students. Similarly, the use of smart-boards in the classroom for lecture creation and delivery can be studied in nurturing the engagement and imagination of students, which can translate to better performance. In this context, the teachers and learners can collaborate in creating learning activities that are interesting and also suit the curriculum plans.

7.1.5. Recommendation 6: Use of computer based formative assessment in an educational setting

Computers are present in almost all schools today, so computer based learning and assessments are another area that should be studied. The use of computer based formative assessment in an educational setting has shown potential in the earlier research, so it's implementation and execution can be studied in Pakistan's context. Similarly, the use of computer based educational games and puzzles can be also be studied. Such studies can reveal significant information regarding effective learning practices and foster the knowledge of pupils.

7.1.6. Recommendation 7: Use of project based formative assessment

Last but not the least, project based formative assessment can also be studied such that students are provided with a project grounded on their learning goals and curriculum and designed on themes of formative attributes i.e. peer assessment, collaboration, and cognitive thinking. The study can show the impact of group tasks on learning and student outcomes.

7.2. Conclusion

To conclude, we think that students and learners in Pakistan deserve better learning and education. We should ensure that they are prepared for future and can compete globally as present economies are knowledge base and survival with dignity in this high paced knowledge world requires good investment in education. Research in education is imperative to investigate and propose ways of learning that can foster cognitive domain, creativity and develop learner's interest in education by taking into account the diverse learning needs. Based on research findings, policies should be made and driven to ensure the implementation and accountability for the sake of quality learning and better future of our nation.

Bibliography

A.Economides, V. T. (2011). Computer based assessment: Gender differences in perceptions and acceptance. *Elsevier: Computers in Human Behavio* , 2108–2122.

Abiodullah, H. M. (2009). Using Assessment for Improving Students Learning: an analysis of University Teachers' Practices. *Bulletin of Education and Research* , 47-59.

Adrián Domínguez, J. S.-d.-N.-M.-S. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education* , 380-392.

Ahmed, S. (2015). *Formative assessment and productive pedagogy in Finnish Classroom assessment:In the lens of curriculum materials.* University of Oslo.

Allen, M. (2004). *Assessing Academic Programs.* Pleasant Hills, CA: Anker.

Altman, D. G. (n.d.). *Parametric v non-parametric methods for data analysis.* From the BMJ: doi: http://dx.doi.org/10.1136/bmj.a3167

Arieli-Attali, Y. A. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education* , 57-63.

Association, I. R. (2012). *Organizational learning and knowledge : concepts, methodologies, tools and applications.* USA: Hershey, PA : Business Science Reference.

Aziz, S. M. (2003). Online Quizzes for Enhancing Student Learning in a First Year Engineering Course. *International Conference on Engineering Education.* Valencia,Spain.

Beckett, L. V. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice . *Canadian Journal of Education* , 239-255.

Ben Shneiderman, C. P. (2010). *Designing the User Interface: Strategies for Effective Human-computer Interaction.* Person.

Beverley Bell, B. C. (2001). The Characteristics of Formative Assessment in Science Education. 536 – 553.

Burr, S. (2009). Integrating assessment innovations in medical education. *British Journal of Hospital Medicine* , 162-163.

Christopher Cheong, F. C. (2013). Quick Quiz: A Gamified Approach for Enhancing. *Pacific Asia Conference on Information Systems*, (p. 15).

Consuelo G.Sevilla, J. A. (1992). *Research methods.* Phillipine: Manila : Rex Book Store.

Cowie, B. B. (2001). *The Characteristics of Formative Assessment in Science Education.* New Zealand: John Wiley & Sons.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* US: Thousand Oaks : SAGE Publications.

Development, O. f.-o. (2008). *Assessment for Learning Formative Assessment.* Retrieved 2016 йил 5-November from Organisation for Economic Co-operation and Development: http://www.oecd.org/site/educeri21st/40600533.pdf

Dictionary and Thesaurus. (2016). Retrieved 2016 йил December from Merriam-Webster: https://www.merriam-webster.com/

Doyle, P. B. (2014). Gamification and student motivation. *Interactive Learning Environments* , 1162-1175 .

Dylan, P. B. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan* , 139-148.

Gee, J. (2007). What Video Games Have to Teach Us About Learning and Literacy.

Hafiz Muhammad Iqbal, S. A. (2009). Using Assessment for Improving Students Learning: an analysis of University Teachers' Practices. *Bulletin of Education and Research* , 47-59.

Investopedia, L. (2016). *Investopedia - Sharper insight. Smarter Investing.* Retrieved 2016 йил December from Investopedia: http://www.investopedia.com/

Jennifer C. Jacoby, S. H. (2014). Enhancing learning through formative assessment. *Innovations in Education and Teaching International* , 72-83.

Katie Larsen McClarty, A. O. (2012). *A Literature Review of Gaming in Education* . PEARSON.

Kerri-Lee Harris, R. J. (2006). Facilitating reflection on assessment policies and practices: A planning framework for educative review of assessment.

Khattak, S. G. (2012). Assessment in schools in Pakistan. *SA-eDUC JOURNAL* .

Kiviniemi, B. C. (2009). The Effect of Online Chapter Quizzes on Exam Performance in an Undergraduate Social Psychology Course. *Teach Psychol.* , 33–37.

Laerd. (2013). *SPSS Statistics Tutorial and Statistical Guide*. From Laerd Statistics: https://statistics.laerd.com/

Louis Cohen, L. M. (2007). *Research Methods in Education.* New York: Routledge.

Lowry, R. (2005). Computer aided self assessment – an effective tool. *Chemistry Education Research and Practice* , 198-203.

Mann-Whitney U test using SPSS Statistics. (2015). Retrieved from Laerd Statistics: https://statistics.laerd.com/

McGraw-Hill. (2001). *Digital game-based learning.* New York.

Nielsen, J. (1993). *Usability Engineering.* USA: Morgan Kaufmann Publishers Inc. San Francisco.

Norbert Pachler, C. D. (2009). Formative e-assessment: practitioner cases. *Computers and Education.*

Norbert Pachler, C. D. (2010). Formative e-assessment: Practitioner cases. *Computers & Education* , 715-721.

Pakistan, G. o. (2014). *National Assessment Report.* Islamabad: National Education Assessment System (NEAS).

Parametric v non-parametric methods for data analysis. (n.d.). From the BMJ: doi: http://dx.doi.org/10.1136/bmj.a3167

Prensky, M. (2001). *Digital game-based learning.* New York: McGraw Hill.

Press, O. U. (2017). *Self-assessment - defination od self-assessment in English*. From English Oxford Living Dictionairies: https://en.oxforddictionaries.com/definition/us/self-assessment

Redish, J. S. (1999). *A Practical Guide to Usability Testing.* USA: Intellect Books.

Reeves, D. (2007). *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning.* Indiana: Solution Tree.

Rehmani, A. (2012). Changing assessment practices in Pakistani schools : A case of AKU-EB middle school assessment framework. *An International Conference at Aga Khan University Institute for Educational*, (pp. 285-295).

Rehmani, A. (2003). Impact of public examination system on teaching and learning in Pakistan. *ANTRIEP* .

Ron Fisher, J. C. (2011). Assisting transition to university: using assessment as a formative learning tool, Assessment & Evaluation in Higher Education. *Assessment and Evaluation in Higher Education* , 225-237.

Ruland, J. W. (2011). *The impact of Using Formative Assessment Attributes in Daily Instruction on Student Affect.* Chicago.

Shafqat Hussain, M. S. (2013). Students' Perceptions of Improving Secondary School. *Scientific & Academic Publishing* , 15-19.

StatsDirect. (n.d.). *StatsDirect Statistical Analysis Software*. From Stats Direct: http://statsdirect.com/

Thiru Aravinthan, W. L. (2011). Effectiveness of formative online quizzes in learning and teaching a structural engineering course. *AAEE Conference.* Fremantle, Australia.

UNESCO. (1998). *International Bureau of Education*. From IBE-UNESCO: http://www.ibe.unesco.org/curriculum/Asia%20Networkpdf/ndreppk.pdf

Wiliam, P. B. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. 139 - 148.

Zahediasl, A. G. (2012). Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. *International Journal of Endocrinology and Metabolism* , 4.

Appendices

Appendix A: Usability Checklist

User Friendly	
Easy to Use	
The icons are clear	
The content of the app is clear	
The language used is easy to understand	
Drag & Drop clear (Expected Result) The label stops on the tagline	
Audio: Is the sound overall clear and understandable?	
Back button (Expected Result) Back button should navigate the user to previous screen.	
Replay (Expected Result) Replay button should reload that scene	

Application Start (Expected Result) Application must not take more than 1-2s to start.	
Exit Button (Expected Result) Exit button should exit the scene/game	

1 Strongly Disagree	2 Disagree	3 Neither Agree nor Disagree	4 Agree	5 Strongly Agree

Appendix B: Usability Form

(To be filled by the students)

Usability Test

Name: _____

1. **I'm satisfied with this science quiz application.**

Strongly Disagree Strongly Agree

1	2	3	4	5

]

2. **The content of application is clear to me.**

Strongly Disagree Strongly Agree

1	2	3	4	5

3. **The application is easy to use.**

Strongly Disagree Strongly Agree

1	2	3	4	5

4. **It was easy to drag and drop the labels.**

Strongly Disagree Strongly Agree

1	2	3	4	5

5. **The content written is easy to understand.**

Strongly Disagree Strongly Agree

1	2	3	4	5

6. **The icons are easy to understand.**

Strongly Disagree Strongly Agree

1	2	3	4	5

Appendix C: Pre-Test

Name: _____ Class: _____

PRE-TEST

1. What protects the eyes from light and danger?
   (a) Eyelashes
   (b) Eyelids
   (c) Pupil

2. Which is not the function of skin?
   (a) Helps you feel things
   (b) Stop water loss
   (c) Make the air to get in and out of body

3. Which organ controls everything you taste, smell, see and hear?
   (a) Brain
   (b) Tongue
   (c) Wind Pipe

4. The parts of human eye are:
   (a) Outer eye, Inner eye
   (b) Wind pipe, Air sacs, Joints
   (c) Pupil, Upper lid, Lower Lid, Eyelashes
   (d) Outer Jaw, Lower Jaw, Incisors

5. What is taken away by the lungs when you breathe out?
   (a) Oxygen
   (b) Carbon monoxide
   (c) Carbon dioxide

6. Which organ helps you in crushing the food and gives you a nice smile?
   (a) Mouth
   (b) Teeth
   (c) Stomach

7. Your friend is sick and you are feeling sad. Which body organ is responsible for causing such feeling?
   (a) Heart
   (b) Brain
   (c) Joints

8. What can damage your lungs?
   (a) Air
   (b) Heavy pollution in cities
   (c) Oxygen

9. The organ that is sort of like a pump and pushes blood through the body:
   (a) Liver
   (b) Heart
   (c) Lungs

10. When you eat where does all that food go?
   (a) Mouth
   (b) Stomach
   a) Liver

11. Your body is completely covered by:
   (a) Hair
   (b) Skin
   (c) Muscles

12. Inside your skull, which organ controls everything you do?
   (a) Stomach
   (b) Lungs
   (c) Brain

13. Canines are used for:
   (a) Grinding
   (b) Tearing
   (c) Seeing

14. Which node of heart causes the heartbeat?
   (a) Relay station
   (b) Pulse
   (c) Pace maker

15. Which organ is like a mixer, churning and breaking down the small pieces of food?
   (a) Teeth
   (b) Wind pipe

Appendix D: Post-Test

Name: _____ Class: _____

POST-TEST

1. What makes our lungs soft and spongy?
   (a) Wind pipe
   (b) Blood
   (c) Air Sacs

2. Which organ helps you in remembering things that happened long ago?
   (a) Brain
   (b) Ear
   (c) Skin
   (d) Heart

3. The node of heart that causes the heart beat is:
   (a) Pace maker
   (b) Pulse
   (c) Relay station

4. The type of tooth that helps you in tearing and piercing the food:
   (a) Molars
   (b) Canine
   (c) Incisors

5. The largest part of your body that makes you feel pain, sense the temperature and pressure is:
   (a) Eyes
   (b) Skin
   (c) Ears

6. The “J” shaped muscular bag that lies below the heart:
   (a) Stomach
   (b) Lungs
   (c) Liver

7. Your brain is protected inside your:
   (a) Ribcage
   (b) Skull
   (c) Eyes

8. Each eye has two flaps that protect them from danger:
   (a) Nerves
   (b) Joints
   (c) Eyelids

9. What helps you in chewing the apple?
   (a) Teeth
   (b) Ear
   (c) Lips

10. An elastic and soft layer that covers your body is:
   (a) Skin
   (b) Joints
   (c) Muscles

11. When you breath in air, you bring oxygen into your lungs and blow out:
   (a) Carbon dioxide
   (b) Hydrogen
   (c) Oxygen
12. Stomach breaks down food into smaller pieces and mix it with:
   (a) Mucus
   (b) Digestive Juice
   (c) Nerves
13. The organ that sends and receives messages from different parts of the body, telling them what to do:
   (a) Lungs
   (b) Bones
   (c) Brain
14. The parts of human eye are:
   (a) Outer Jaw, Lower Jaw, Incisors
   (b) Pupil, Upper lid, Lower Lid, Eyelashes
   (c) Wind pipe, Air sacs, Joints
   (d) Inner eye, outer eye
15. The organ that is protected by ribcage and keeps the blood moving in our body:
   (a) Heart
   (b) Stomach
   (c) Liver

Appendix E: Consent Letter

Research Title: Effectiveness of Formative Assessment in Improving Learning Quality: Using Interactive Science Quiz to Enhance Student Learning Outcomes at Primary Level

Researcher: Rida Imran

Supervisor: Farzana Ahmad

To the Head of the institution,

Your school is invited to participate in a research study. This consent letter provides you with the information on the research project. It is important that you fully understand the research in order to make an informed decision.

Purpose:

The purpose of this study is to determine the effectiveness of formative assessment using interactive science quiz application.

Procedures:

This research will not alter the standard practices used by your teachers. The course teaching methods and assessments used in your classrooms will remain in place.

Privacy:

Information collected for this research will remain confidential and only accessed by the researcher.

This project is approved by National University of Sciences and Technology (NUST). Kindly sign the form if you have read the consent form and agree to allow your school to participate in this research.

Head's signature _____ Date _____

Appendix F: Student Questionnaire

Name: _____

For Engagement

1. I wanted to complete the quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

2. I wanted to explore all the options available to me.

Strongly Disagree Strongly Agree

1	2	3	4	5

3. I found the quiz satisfying.

Strongly Disagree Strongly Agree

1	2	3	4	5

4. I felt absorbed in the quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

5. I felt that time passed quickly.

Strongly Disagree Strongly Agree

1	2	3	4	5

6. I felt excited during the quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

For Enjoyment

1. I felt exhausted when playing Science Quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

2. I felt unhappy when playing Science Quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

3. I felt worried when playing Science Quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

4. I felt happy when playing Science Quiz.

Strongly Disagree Strongly Agree

1	2	3	4	5

For Learning

1. Playing Science Quiz improves my learning quality.

Strongly Disagree Strongly Agree

1	2	3	4	5

2. Playing Science Quiz enhances my learning effectiveness.

Strongly Disagree Strongly Agree

1	2	3	4	5

3. Playing Science Quiz improves my learning productivity.

Strongly Disagree Strongly Agree

1	2	3	4	5

4. Playing Science Quiz helps to get better grades.

Strongly Disagree Strongly Agree

1	2	3	4	5

5. Playing Science Quiz improves my learning performance.

Strongly Disagree Strongly Agree

1	2	3	4	5

Appendix G: Validity form for pre/post –test

Name: _____ School: _____

1. Do you think the content written is appropriate for the students of Grade 4?

   _____

   _____

2. The pre/post tests are made keeping in mind the following Learning Outcomes:
   - **Identify major parts of human body**
   - **State functions of the major parts of the body**

Do they fulfill these learning outcomes?

_____

_____

3. Is the vocabulary used appropriate?

   _____

   _____

4. Any comments:

   _____