Abductive science inquiry using mobile devices in the classroom

Sohaib Ahmed*, David Parsons

Institute of Information & Mathematical Sciences, Massey University, Auckland 0632, New Zealand

Abstract

Recent advancements in digital technology have attracted the interest of educators and researchers to develop technology-assisted inquiry-based learning environments in the domain of school science education. Traditionally, school science education has followed deductive and inductive forms of inquiry investigation, while the abductive form of inquiry has previously been sparsely explored in the literature related to computers and education. We have therefore designed a mobile learning application ‘ThinknLearn’, which assists high school students in generating hypotheses during abductive inquiry investigations. The M3 evaluation framework was used to investigate the effectiveness of using ‘ThinknLearn’ to facilitate student learning. The results indicated in this paper showed improvements in the experimental group’s learning performance as compared to a control group in pre-post tests. In addition, the experimental group also maintained this advantage during retention tests as well as developing positive attitudes toward mobile learning.

Keywords: Abductive science inquiry, Inquiry-based learning, Mobile learning, Mobile learning application, Science education, Technology-assisted learning

1. Introduction

The rapid development toward mobile technology deployment at school level offers students new opportunities for increasing engagement, motivation and learning (Lin, Fulford, Ho, Iyoda, & Ackerman, 2012). In fact, mobile technologies are revolutionizing school education and transforming the conventional classroom with interactive classroom applications that have the potential to enhance students’ learning experiences (Scornavacca, Huff, & Marshall, 2009). This is due to the affordances of these technologies that offer many different levels of engagement (Churchill & Churchill, 2008; Looi, Wong, So, & Seow, 2009). Thus, these technologies may increasingly become a convincing choice of technology for providing learning experiences inside the school (e.g. lab, classroom, library) but also allow students to perform learning activities in other environments (e.g. park, woodland, museum) (Price & Rogers, 2004).

Inquiry-based Learning (IBL) is an educational approach in which learners can get knowledge through exploration and investigation within authentic settings, and may enhance their critical thinking skills (Hwang & Chang, 2011; Li & Lim, 2008). In an authentic IBL science activity, learners are involved in science learning by hands-on activities and they are also provided with resources to assist them to understand domain specific knowledge by engaging in scientific reasoning processes; hypothesis generation, experimentation and evidence evaluation (de Jong, 2006). This seems to have twofold advantages; learners are able to develop understanding and knowledge about the scientific phenomena that they observe in the physical world, and they also identify how to perform the steps of scientific inquiry like scientists (Bell, Urhahne, Schanze, & Ploetzner, 2010). These two aspects are inter-related in that learners will not learn from inquiry without knowing how to do inquiry, and on the other hand, a particular domain is always required for practical inquiry skills (van Joolingen & Zacharia, 2009).

From the school science perspective, a number of mobile learning projects have been discussed in the literature, reflecting the diversity of inquiry investigations. Among these, the Ambient Wood Project (Price & Rogers, 2004), MPLS (Huang, Lin, & Cheng, 2010) and Savannah (Facer et al., 2004) are well-known projects in which learners are engaged in science learning activities by interacting with natural and virtual environments in outdoor settings. For indoor school settings, BioKIDS Sequence (Parr, Jones, & Songer, 2004) and WHIRL (Yarnall, Shechtman, & Pennel, 2006) use mobile technologies in science classrooms in order to support more frequent assessment practices.

* Corresponding author.

E-mail addresses: s.ahmed@massey.ac.nz (S. Ahmed), d.p.parsons@massey.ac.nz (D. Parsons).
HandLearn (Avraamidou, 2008), LET’s Go! (Vogel, Spikol, Kurti, & Milrad, 2010) and nQuire (Mulholland et al., 2012) are projects for mobile science learning activities in both indoor and outdoor settings.

Most of these mobile science learning applications generally follow a deductive or inductive means of inquiry where learners process their ideas (or hypotheses) to explore the observed phenomena (Grandy & Duschl, 2007). In contrast, abductive science inquiry mainly focuses on the formation of hypotheses about the observed phenomena (Oh, 2011). In the technology-assisted learning literature, this kind of inquiry has been sparsely investigated (Grandy & Duschl, 2007; Oh, 2011). No previous studies have demonstrated the benefits of mobile learning in hypothesis formation activities in the context of abductive science inquiry investigations. This provides us with an opportunity to look at some new ways to deal with science inquiry learning activities using mobile learning technologies.

With this in mind, we have developed a mobile learning application, ‘ThinknLearn’, in order to explore abductive science inquiry activities in classroom settings. Through learning with ‘ThinknLearn’, students can get a deeper understanding of a particular domain (heat energy transfer, in this particular study) and the assistance provided can help them to interpret data into meaningful hypotheses. To evaluate its effectiveness on students’ learning performance, the M3 evaluation framework (Vavoula, Sharples, Rudman, Meek, & Lonsdale, 2009) has been used. The corresponding results and analyses are discussed later in this paper.

2. Abductive science inquiry

An abductive science inquiry is based on the theory of abduction that was considered as a form of inference by C. S. Peirce (1839–1914) (cited in Raholm, 2010). In this trait of inquiry, learners tend to generate plausible hypotheses on the basis of background theories and observations (Raholm, 2010) and use their critical thinking to explain the observed phenomena (Oh, 2011). Thus, this type of inquiry exhibits different behavior from the other two commonly used methods in science inquiry investigations; deductive and inductive, where learners are required to validate data or generate a rule using known hypotheses. In both cases, hypotheses are known entities, while for abductive science inquiry investigations it is the other way round (hypotheses are initially unknown). The following example, taken from our domain of study, will show the relationships more clearly. Here, the ideas relate to black surfaced tins containing hot water losing heat more quickly than white or shiny surfaced tins. In these examples, the Case (Hypothesis), Result (Observation) and Rule (Condition or Suggestion) are defined to show the differences in order.

Deduction:

- **Rule** – The water particles in a black surfaced tin vibrate faster than the other tins.
- **Case** – A black surfaced tin absorbs more heat energy than the other tins.
- **Result** – A black surfaced tin cools more quickly than the other tins.

Induction:

- **Case** – A black surfaced tin absorbs more heat energy than the other tins.
- **Result** – A black surfaced tin cools more quickly than the other tins.
- **Rule** – The water particles in a black surfaced tin vibrate faster than the other tins.

Abduction:

- **Rule** – The water particles in a black surfaced tin vibrate faster than in the other tins.
- **Result** – A black surfaced tin cools more quickly than the other tins.
- **Case** – A black surfaced tin absorbs more heat energy than the other tins.

From this example, it can be observed that in both deduction and induction, a Case (Hypothesis) is processed with either a Rule or a Result to generate the other component, while in abduction, the Rule and Result are used together to find a Case. This trait of abduction is well-suited to inquiry problems in which learners are challenged to formulate scientific hypotheses and leads them toward new explanations on the basis of observed phenomena (Oh, 2011).

Scientific explanations have two important aspects in science (Peker & Wallace, 2011): first, they provide a unified picture of how various scientific phenomena fit together; second, they help us to comprehend how things work in the world. Further, it is suggested in cognitive science literature that scientific explanation is highly recommended as a tool for constructing learners’ cognitive thinking skills (Krupa, Selman, & Jacquette, 1985). A number of studies (Oh, 2011; Peker & Wallace, 2011) have demonstrated that if appropriate resources are provided to learners then they are able to make meaningful hypotheses and may articulate their comprehension about a given problem in better ways. However, there is a lack of exploration of mobile applications in terms of their guidance in facilitating science inquiry learning, particularly in abductive inquiry investigations. Thus, a mobile web application ‘ThinknLearn’ has been designed that can allow us to evaluate technology-assisted hypothesis generation with high school students performing abductive inquiry investigations, as described in the subsequent sections.

3. ThinknLearn: a mobile web application

In consultation with the science teachers from a local high school, we agreed on one of the science inquiry topics (i.e. heat energy transfer) from the New Zealand standard science curriculum as the experimental context to test a mobile learning application that supports abductive science inquiry.

3.1. Conceptual design

The conceptual design of the application ‘ThinknLearn’ defines an abstract viewpoint of the research objectives that deals with the development of a mobile web application for abductive science inquiry investigations, as depicted in Fig. 1. One of the main reasons for developing a mobile web application in classroom settings is to provide an environment which assists students where they are performing hands-on-activities. In comparison with desktop machines, it is suggested that the use of mobile devices within classroom activities offers...
different learning experiences in which students can access digital information and means of communication with other students or instructors with convenience, efficiency and immediacy (Curtis, Luchini, Bobrowsky, Quintana, & Soloway, 2002).

In this conceptual design, the four main components are defined as Knowledge Testing, Learning, Prediction & Selection, and Observation & Measurement. These main components are connected with a defined domain ontology via the Learning Content Generation component, which is responsible for generating learning content in the form of Multiple Choice Questions (MCQs), suggestions and hypotheses in the application.

An ontology is a hierarchy that defines concepts of a domain and their relationships along with the set of inference rules that powers its reasoning functions (Lee, Tsai, & Wang, 2008). In the given example, an ontology is defined which is based on one of the standard school science topics (i.e. heat energy transfer). It contains all the relevant aspects of heat energy that have impact on the movement of water particles, and the loss and absorption of heat energy due to different colored surfaces.

In these four components, the Knowledge Testing component is the one which is used for assessing students’ knowledge by asking MCQs about a given topic. The Learning component is designed for giving suggestions according to the nature of the problem (i.e. science inquiry). The Prediction & Selection component guides learners in hypothesis generation and selection after learning about the topic. In the Observation & Measurement component, students can interact with an environment to observe and measure natural phenomena. In addition, they need to use a mobile web user interface to access the information through the application as depicted in Fig. 1. Further information regarding the technical details related to the design of the application can be found in Ahmed, Parsons, and Mentis (2012).

3.2. Abductive inquiry model

The ‘ThinknLearn’ mobile web application follows the Abductive Inquiry Model (AIM) (Oh, 2011) that defines four phases: exploration, examination, selection and explanation. This model may be utilized for pedagogical purposes of enriching students’ knowledge as well as engaging them in scientific inquiry practices.

In this particular example, three tins with different surface colors are provided in order to compare the way they radiate heat energy. Tin A is painted White, tin B Black and tin C is Silver or Shiny. In the exploration phase, students have to record the temperature of each tin at a particular time interval after pouring boiling water into these tins. In this example, the temperature ranges lie between 20 and 100 °C while the time can be recorded between 0 and 20 min. The application asks for the temperature of the given tins with different colored surfaces to be measured at a particular time. The main task for students in this phase is to investigate the given scientific phenomena by observing data. Fig. 2a shows one of the measurement screens from the application, which is asking for two inputs from the student; the temperature of a Shiny tin and the time at which the temperature of that tin is recorded.

The examination phase of this model may help learners to generate scientific hypothesis for this investigated phenomena. In this phase, the application asks a series of MCQs regarding the gathered values of the measures taken during the exploration phase. This feature makes students use their observational and critical thinking abilities to answer the given questions, as depicted in Fig. 2b. This phase of the model guides students toward a point where they are able to construct hypotheses about the observed measures and understand the knowledge presented in this application. The main reason for not giving correct answers straightaway is to exploit learners’ critical thinking skills for comprehending the given domain concepts.

In the selection phase, students are asked to select one of the suggested hypotheses about the given problem (see Fig. 2c). This application defines two types of hypotheses about the underlying domain: first, the vibration of the water particles and the loss of heat energy from the different colored tins; second, the heat absorption and the loss of heat energy from the different colored tins. The application extracts one of these hypotheses using a random function. In this way, students may get different hypothesis questions while performing the same task in a classroom. In addition, the application extracts all the possible hypotheses including one correct and three other distracters by using the defined ontology and its inter-related concepts.
In the final phase of the AIM – explanation – students propose complete explanations for the given problem after getting assistance through the application. These explanations help students to gain a comprehension of the topic, which may be subsequently tested by the teacher to assess learning performance during the inquiry investigation. The evaluation of the application and the learning assessment of the students are both discussed in the subsequent sections.

4. Methodology

In mobile learning applications, evaluation can serve as a means to examine the effectiveness of the application to enable learning and offer new learning opportunities with the support of the underlying technology (Sharples, 2009). For evaluation purposes, part of the M3 evaluation framework (Vavoula et al., 2009) was applied. This framework consists of three levels of granularity (Vavoula & Sharples, 2009): Micro, which examines the individual activities of technology users and assesses the utility of the application; Meso, which investigates the learning experience as a whole, and learning breakthroughs and breakdowns; and Macro, which examines the impact of the new technology on established learning practices. We do consider the Macro level in this study because this level would involve the establishment of abductive science inquiry practices, supported by mobile learning, in school science education. At this stage of the research, this level of evaluation is not yet possible.

In this study, two experiments were conducted with high school students for evaluating the application using the M3 evaluation framework (Vavoula et al., 2009) as shown in Fig. 3. In the first experiment, regarded as the pilot study, only Micro level evaluation was applied, with usability (ISO, 2003) and mobile quality (Parsons & Ryu, 2006) aspects being evaluated. Both quantitative and qualitative data were gathered about these aspects using mixed methods (Johnson & Onwuegbuzie, 2004): questionnaires and semi-structured group

Fig. 3. Evaluation of the application ‘ThinknLearn’.
discussions. The same Micro level evaluation was also conducted with a larger sample size in the final experiment, along with the Meso level evaluation (Vavoula et al., 2009). This Meso level was used to assess the impact of the application on learners’ assessments between experimental and control groups in their pre-post and retention tests. In this paper, only the corresponding results and analyzes of the final experiment are discussed.

4.1. Experimental design

The rationale for this experimental design was to evaluate ‘ThinknLearn’ in terms of its utility and educational issues as defined in Table 1. For that purpose, two levels (i.e. Micro and Meso) of the M3 evaluation framework (Vavoula & Sharples, 2009) were applied. The utility covers the use of the application for guiding students in order to construct scientific hypotheses about the underlying domain, as measured through usability and mobile quality aspects of the application. However, for the educational issues, this experiment was designed and evaluated between control and experimental groups.

Usability can play an important part in the success of any learning application. If a learning application is not usable enough then it impedes students’ learning; the students would spend more time learning how to use the application than learning the contents (Wong, Nguyen, Chang, & Jayaratna, 2003). Generally, usability is measured by assessing learners’ performance, by identifying their difficulties with the interface, and by asking them about their opinions of the application (Sim, Farlane, & Read, 2006). For that purpose, three standard ISO metrics namely learnability, operability and understandability were used (ISO, 2003) for measuring usability aspects of the application. Besides these usability aspects, three quality aspects of mobile learning applications were considered: metaphor, interactivity and learning content (Parsons & Ryu, 2006). These aspects with their explanations are depicted in Table 2.

In this experimental design, a control group was used to perform the heat energy experiment in the science laboratory using a “pre-test — > heat energy experiment — > post-test — > retention test” method, where the students carried out the learning activities without using ‘ThinknLearn’. On the other hand, the experimental group used this application while performing the same experiment in the science laboratory using a “pre-test — > heat energy experiment using ThinknLearn — > post-test — > retention test” method. Both experimental and control groups received the identical experimental context, performed identical learning activities, and participated during the same period. The learning activities involved pre and post tests during hypothesis construction activities in the context of abductive science inquiry. These tests consist of MCQs and an open-ended question regarding a hypothesis and its explanation which assessed learners’ knowledge about the topic covered in their science classes earlier. An additional test (i.e. retention test) was also performed after a span of two months, to measure the learning retention of both groups.

This experiment was a between-subject design. The two ways of generating hypotheses and improving learning (i.e. with and without using the application ‘ThinknLearn’) are considered as the independent variables, while the dependent variables are learning outcomes and knowledge retention in this experiment. The measurement of the learning outcome assesses how well each participant has learnt the given science content (i.e. heat energy transfer) in order to construct scientific hypotheses. For the knowledge retention, the investigation measures how each participant has remembered the given knowledge after some time has elapsed.

4.2. Participants

In this experiment, 161 NCEA level 1 science students from six different classes voluntarily participated, divided into two groups. One of the groups was treated as an experimental group, which comprised 86 students from three science classes. The other 75 students were a control group. In the experimental group, all students filled in a questionnaire and participated in group discussions regarding the utility of the application. Out of these 86 students, 81 students participated in pre-post activities. In the control group, all 75 students were involved in pre-post activities without using ‘ThinknLearn’. After two months, the retention test was conducted in their classrooms. In this retention test, only 125 students participated. Amongst them, 67 were those students who had previously been in the experimental group, while 58 had been control group students.

For the invitation to participate, a procedure had been followed in which students needed to fill in the individual consent forms for the evaluation if they were at least 16 years old. For those students who were less than 16 years of age, they were asked to get these consent forms signed by their parents. This was one of the requirements of the low risk ethics notification agreed by the school and the researchers’ university. These forms briefly explained the nature of the experiment but did not mention the learning topic. At the time of the experiment, most of the students were equipped with their own mobile devices which indicated that the students were previously mobile users. The students’ mobile technological background makes them good evaluators of a mobile application that deals with their science learning. The experiment was undertaken as a learning activity as part of their regular science classes using mobile devices equipped with WiFi. This learning activity involved using a mobile application that was conducted in their science laboratory, so the environment was not a novel factor for the students.

For the distribution of the groups (i.e. experimental and control), science teachers were consistent on keeping the class structure intact. Therefore, students could not be randomly assigned to any of the groups. However, three classes apiece were selected as experimental and control groups respectively. In each class, there were previously 8–9 sub-groups for performing their classroom activities. No modification

<table>
<thead>
<tr>
<th>M3 evaluation framework level</th>
<th>Evaluation aspects</th>
<th>Form of evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro Level</td>
<td>Utility Issues (Usability and quality aspects)</td>
<td>Questionnaire Semi-structured group discussions (Experimental &amp; Control groups)</td>
</tr>
<tr>
<td>Meso Level</td>
<td>Educational Issues (Learning outcomes)</td>
<td>Pre-post tests Retention tests</td>
</tr>
</tbody>
</table>
was made to this class structure. The experiment was conducted in February 2012, when the 1st term of the school year was underway. It involved six science classes, so the experiment was run for the whole second week of February. The retention test was undertaken almost two months later (i.e. the first week of April, 2012).

4.3. Apparatus

Three tins with different colored surfaces including Black, Shiny (Silver), & White were provided for comparing the way these tins radiate heat energy. The experimental group of students were equipped with WiFi enabled mobile devices to access the application which may assist them to understand the concepts about the given topic (i.e. heat energy transfer). However, the control group students were required to perform the experiment in the traditional way (i.e. without any assistance from the application).

An MCQ quiz was given along with instructions appropriate to each group for performing pre and post learning activities. In addition, a questionnaire was also provided to each participant of the experimental group to investigate his/her individual learning experience about the given application. For the retention test, students from both groups were required to answer the same MCQs asked earlier in their pre-post activities.

4.4. Procedure

In the beginning, the information regarding the experiment, and the data collection process, were described by science teachers to their students. For the experimental group, students were asked to answer pre-post quizzes, filled in questionnaires and were involved in semi-structured group discussions while answering the questions posed by the researcher. However, for the control group, they were only required to answer pre-post quizzes.

In this evaluation process, both groups of students were initially asked to answer the pre-test which comprised four MCQs. Following this, they were required to perform the heat energy transfer experiment as depicted in Fig. 4. In this experiment, they found some data values related to each investigated tin. These data values helped them to understand some key concepts discussed in the given topic. At the end of the experiment, they were again asked to answer the same MCQs with the addition of one open-ended question related to the hypothesis, with its explanation. This question was used to understand how well the students engaged in the learning and critical thinking process during this abductive form of inquiry investigation. However, both groups were provided with different instructions in their post-tests as the experimental group was additionally given mobile devices with which they could access the web application ‘ThinknLearn’.

For evaluating the usability and the utility of the application, experimental group students were also required to individually rate a 9-question questionnaire on a five-point Likert scale. Subsequently, they were also engaged in semi-structured group discussions in which students were asked to consider three questions related to their overall learning experiences about the application. For the retention test, the same four MCQs were again asked to the students of both groups after a span of two months. For that purpose, they were only required to answer these questions in their classrooms instead of performing science experiments again. The corresponding results are discussed in the following section.

Table 2
Usability and quality aspects for evaluating ‘ThinknLearn’ (adapted from Chua & Dyson, 2004; Parsons & Ryu, 2006).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sub-characteristic</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usability aspects</td>
<td>Learnability</td>
<td>Can the student learn the system easily?</td>
</tr>
<tr>
<td></td>
<td>Operability</td>
<td>Can the student use the system without much effort?</td>
</tr>
<tr>
<td></td>
<td>Understandability</td>
<td>Does the student comprehend how to use the system easily?</td>
</tr>
<tr>
<td>Quality aspects</td>
<td>Interactivity</td>
<td>Is there an adequate level of interactivity between the student and the system?</td>
</tr>
<tr>
<td></td>
<td>Learning Content</td>
<td>Does the student feel that the learning component is of high quality?</td>
</tr>
<tr>
<td></td>
<td>Metaphor</td>
<td>Does the student have an overall vision of the learning process?</td>
</tr>
</tbody>
</table>

Fig. 4. Experimental group students involved in experimental activities.
5. Results

The M3 evaluation framework supports the technology development process from the very early stages of the design to the final assessment of the technology in a learning context (Vavoula et al., 2009). In this experiment, two levels (Micro and Meso) were applied to explore the use of the mobile learning application. The corresponding results and analyses are described in the following sub-sections.

5.1. Micro level evaluation

In this micro level evaluation, the utility of the application ‘ThinknLearn’ was investigated. For this purpose, the responses from a questionnaire and group discussion questions were gathered from the experimental group students about their learning experiences with the application. The control group students were not involved in this aspect of the evaluation.

5.1.1. Questionnaire responses

The questionnaire was given to the students after they had completed the science experiment. The 9 statements in the questionnaire attempted to address different aspects of usability (learnability, understandability, operability) and mobile quality (metaphor, interactivity, learning content) as described in Table 3. A five-point Likert scale was used in which 1 was ‘strongly disagree’ while 5 was denoted as ‘strongly agree’. When this application was tested with science students, their overall responses were positive and encouraging.

As depicted in Table 3, three questionnaire statements (S2, S3, and S4) were intended to address the fundamental aspects of usability found in ‘ThinknLearn’: learnability, operability, and understandability (ISO, 2003). The responses to statements ‘S2’ and ‘S3’ revealed that the students found this application was not difficult to use and navigation was straightforward. The ratings on the statement ‘S4’ also revealed that our respondents believed that the guidance toward hypothesis generation and the whole learning process was very easy to understand. A one sample t-test against the neutral value 3.00 confirmed these interpretations (tS2 = 6.14, p < .01 for S2; tS3 = 7.88, p < .01 for S3; tS4 = 4.20, p < .01 for S4).

According to the results, the students experienced ‘ThinknLearn’ as an effective learning application. The results of the interactivity aspects confirm these interpretations (tS5 = 4.62, p < .01 for S5; tS6 = 2.38, p < .01 for S8). The ratings on the statements S1 and S9 revealed that the students considered that this application was an effective learning application because it provides an enjoyable learning experience to them (tS9 = 5.74, p < .01 for S1; tS5 = 5.11, p < .01 for S9). These statements cover Learning content which describes learners’ feelings about the quality of the application. Similarly, the responses to the other quality aspect, Metaphor showed that the students experienced an overall vision of the learning processes (tS7 = 7.10, p < .01 for S5; tS6 = 5.19, p < .01 for S7).

In summary, the questionnaire responses of the students were promising which suggests that ‘ThinknLearn’ has embodied considerable software quality measures including usability and the quality aspects required by any mobile learning application.

5.1.2. Group discussions

Qualitative data were also gathered in semi-structured group discussions. Three of the questions (see Table 4) in the discussions were asked of the students for evaluating the application in terms of their mobile learning experiences, hypothesis generation process and comprehension about a given topic. These questions were used to define the usability and mobile quality aspects of the software quality measures.

In these group discussions, 25 groups from three science classes participated. As far as the question 1 responses were concerned, most of the students considered that ‘ThinknLearn’ was easy to use and they did not find any difficulty while using it. However, there were a few who found this application difficult in terms of its guidance toward hypothesis generation. One of the groups highlighted that “... questions were difficult and the given suggestions were not easy to understand”. Those students who regarded the application as difficult may not understand the deliberate intent of this application to exploit their critical thinking skills by posing challenges, but this does not negate the possibility that their understanding was enhanced nonetheless.

In answering the second question in the group discussions, almost all the students were positive about their learning experiences and they enjoyed using the application. One of the students stated that “we really enjoyed using it. This application was pretty good and engaging, it helped you to learn about your course (science)”. Another student commented that “this type of application keeps you on focus and requires better attention”. In contrast, one student disliked this application. According to them, “it was boring and confusing and therefore, I did not like it”. Despite this, overall, students enjoyed the innovative way of learning, and found the application interesting and engaging.

The responses for question 3 were promising and most of the students felt that the given suggestions were relevant and made them think. One of the students indicated that “these suggestions are relevant to the answers but they make us think”. On the other hand, there were a few who remarked that “… more detail should be provided” and “… relevant but they (suggestions) did not explain much”. These comments

<table>
<thead>
<tr>
<th>No.</th>
<th>Statements</th>
<th>Evaluation aspects</th>
<th>Mean ± standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>This mobile learning experience was enjoyable.</td>
<td>Learning Content</td>
<td>3.66 ± 0.11</td>
</tr>
<tr>
<td>S2</td>
<td>This mobile application was easy to use.</td>
<td>Learnability</td>
<td>3.66 ± 0.11</td>
</tr>
<tr>
<td>S3</td>
<td>Navigation through this application was easy.</td>
<td>Operability</td>
<td>3.87 ± 0.11</td>
</tr>
<tr>
<td>S4</td>
<td>This application guides me to formulate a hypothesis.</td>
<td>Understandability</td>
<td>3.45 ± 0.11</td>
</tr>
<tr>
<td>S5</td>
<td>The given suggestions in the application were relevant.</td>
<td>Metaphor</td>
<td>3.74 ± 0.10</td>
</tr>
<tr>
<td>S6</td>
<td>This application helps me understand the relationships between different variables</td>
<td>Interactivity</td>
<td>3.50 ± 0.11</td>
</tr>
<tr>
<td>S7</td>
<td>The given suggestions help me to understand the topic.</td>
<td>Metaphor</td>
<td>3.54 ± 0.10</td>
</tr>
<tr>
<td>S8</td>
<td>This application helps me to improve my reasoning skills.</td>
<td>Interactivity</td>
<td>3.28 ± 0.12</td>
</tr>
<tr>
<td>S9</td>
<td>It is an effective learning application.</td>
<td>Learning Content</td>
<td>3.55 ± 0.11</td>
</tr>
</tbody>
</table>
showed that this application presents some challenges to the students and made them think about the given topic. It may be argued that a certain level of challenge was maintained in this application to make it more engaging and interesting. However, some ways may be needed to convince those students about the value of this approach.

On the whole, the group discussion responses suggest that the application was engaging and the given suggestions make learners think about the knowledge space under investigation, and may exploit their critical thinking skills.

5.2. Meso level evaluation

In this level, learning assessments between experimental and control group students were examined. It involved pre and post activities, including answering MCQs and writing hypotheses with explanations while performing science experiments. Further, the same test was conducted with these students to evaluate how well the students retained their knowledge after two months. The results and analyzes of this evaluation level are covered in the following sections.

5.2.1. Learning assessments

In the pre-post tests, students were asked to answer MCQs related to the learning domain. These test results showed that both groups of students gained knowledge about the learning domain after conducting science inquiry experiments, as shown in Table 5. Two paired-sample t-Tests were used to compare pre and post test means for both groups; experimental and control. A paired-sample t-Test found a significant difference ($p < .01$) between scores obtained by the experimental group students during their pre and post tests. However, another paired-sample t-Test, which compared mean scores of control group students during pre and post tests, did not show statistically significant differences ($p = .088$).

In comparing these two groups, an independent sample t-Test was used to find out the learning performance differences. The results showed a significant difference ($p = .025$) between the group using ‘ThinknLearn’ and the other students. As a matter of fact, the control group students got marginally better scores in their pre-tests as compared to the experimental group students. However, in the post-tests, both groups improved but the experimental group gained more in learning performance than the control group as depicted in Fig. 5.

In the post-tests, both group students were asked to write a hypothesis about the color of any of these three tins with the explanation in the open-ended dialog box. As far as the marking of the open-ended question was concerned, it was mutually decided with the science teachers to mark thus: ‘0’ for wrong (or no) hypothesis; ‘0.5’ for a correct hypothesis but a wrong explanation; ‘1’ for a correct hypothesis with its explanation. As an example, one of the answers from the students who got ‘1’ mark for a correct hypothesis with its explanation was “Black tin absorbs more heat energy than the other tins and loses more heat energy than the others therefore it keeps the water cool from the inside”. Given that such answers are open to interpretation, and the marking scheme is coarse grained, there is the potential for bias which should be taken into account when analyzing our results.

Table 4

<table>
<thead>
<tr>
<th>No.</th>
<th>Software quality measures</th>
<th>Group discussion questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Usability</td>
<td>What type of difficulty do you find in using this application?</td>
</tr>
<tr>
<td>2</td>
<td>Mobile Quality</td>
<td>How do you feel after using this application?</td>
</tr>
<tr>
<td>3</td>
<td>Mobile Quality</td>
<td>What do you think about the suggestions given in the application?</td>
</tr>
</tbody>
</table>

Table 5

<table>
<thead>
<tr>
<th>Group</th>
<th>Test</th>
<th>Students (N)</th>
<th>Mean (M)</th>
<th>Standard deviation (SD)</th>
<th>Standard error (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Pre</td>
<td>81</td>
<td>2.05</td>
<td>1.06</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>81</td>
<td>2.67</td>
<td>1.02</td>
<td>0.11</td>
</tr>
<tr>
<td>Control</td>
<td>Pre</td>
<td>75</td>
<td>2.09</td>
<td>1.00</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Post</td>
<td>75</td>
<td>2.30</td>
<td>0.96</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Fig. 5. Pre-post tests comparison between experimental and control groups.
According to the applied independent sample t-Test, the results find a significant difference \((p = .017)\) between both groups. The experimental group students got improvements in their critical thinking and learning while formulating hypotheses about the learning domain. However, the control group students did not appear to understand the given topic so well and therefore were not able to formulate hypotheses and their explanations at the same level as the experimental group. The students’ scores in percentages confirming these interpretations are illustrated in Fig. 6.

5.2.2. Learning retention

The learning retention test was carried out with both groups after two months. In this retention test, only MCQs were asked about the same topic that was explored in the earlier tests. Both groups’ students performed well in their retention tests, but the number of students participating was reduced from the previous pre-post tests.

Two paired-samples t-Tests were used to evaluate the mean differences between the scores obtained only from the students in both groups who were involved in both the post and retention tests as listed in Table 6. The difference in mean values from those in Table 5 is thus a consequence of only assessing a subset of the original student groups in the retention tests. With respect to the comparison between these two groups, an independent sample t-Test was used to find out the learning retention differences. The results demonstrated a difference between groups after two months that was statistically significant \((p = .032)\) in favor of the experimental group.

The results were interesting in that both the experimental and control groups’ students improved their scores from the tests conducted two months previously (see Fig. 7). One of the reasons for these results might be that the students’ final examination would be starting in a few weeks time. Hence, it may be speculated that both groups’ students were involved in revising their theoretical knowledge of a given topic which assists them to perform better in these retention tests. However, it is important to note that experimental group maintained their advantage over the control group.

6. Discussion

The results and analyses in this study indicate that the mobile learning environment being evaluated can help learners to enhance their learning performance and can also retain this advantage in domain knowledge over a period of time. However, there are diverse results found in the literature regarding the benefits of mobile learning in the classroom in terms of learning performance and learning retention. In this regard, some significant work has been done in the past in which researchers found that the use of mobile learning applications fostered positive attitudes toward school education, particularly in learning assessments (Cavus & Ibrahim, 2009; Chen & Hsu, 2008; Lai, Yang, Chen, Ho, & Chant, 2007; Zurita & Nussbaum, 2004). In the literature related to school sciences, there are many studies (Huang et al., 2010; Hwang & Chang, 2011; Uzunboylu, Cavus, & Ercag, 2009) stating that students equipped with mobile learning applications while performing science learning activities enhanced their knowledge compared with those who were involved in traditional ways of science learning. The results of these studies showed that students performed better in their learning activities using mobile learning environments. However, in some other studies the results showed that there was no such difference in students’ learning performance with and without using a mobile learning application (Park, Parsons, & Ryu, 2010).

Regarding learning retention, in one study Sandberg, Maris, and de Geus (2011) stated that the students were not able to retain knowledge over a period of time when using a mobile learning application in formal and informal learning environments. In another study, the learning performance of the experimental and the control groups was not significantly improved after five weeks duration (Zhang, Song, & Burston, 2011). In contrast, there are a few studies which resulted in favor of the use of mobile devices for long term learning retention such as Fozdar and Kumar (2007), and Shih, Chu, Hwang, and Kinshuk (2011). These studies demonstrate different aspects of learning retention using mobile devices in school education. The varying results of these studies do not give a clear indication whether mobile

Table 6

<table>
<thead>
<tr>
<th>Group</th>
<th>Test</th>
<th>Students (N)</th>
<th>Mean (M)</th>
<th>Standard deviation (SD)</th>
<th>Standard error (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental</td>
<td>Post</td>
<td>67</td>
<td>2.61</td>
<td>1.04</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>67</td>
<td>2.97</td>
<td>0.85</td>
<td>0.10</td>
</tr>
<tr>
<td>Control</td>
<td>Post</td>
<td>58</td>
<td>2.34</td>
<td>0.89</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Retention</td>
<td>58</td>
<td>2.60</td>
<td>1.04</td>
<td>0.14</td>
</tr>
</tbody>
</table>
learning can enhance students' performance in their learning activities or if comprehension of domain knowledge can be retained for a longer period of time. These diverse results demand further exploration of the potential benefits of mobile learning.

Specifically to hypothesis formation activities, there are some studies found in the literature that show that students have capabilities to make scientific hypotheses and their explanations when appropriate resources are provided to them during science inquiry investigations (Mulder, Lazonder, & de Jong, 2010; Oh, 2011; Peker & Wallace, 2011). However, none of these studies demonstrated the benefits of mobile learning in hypothesis formation activities in the context of abductive science inquiry investigations.

7. Conclusion and future work

This study provides some insights to science educators and researchers as to why the use of this innovative application ‘ThinknLearn’ might be an effective way of assisting high school students not only in generating hypotheses but also in exploiting their critical thinking skills. The main contribution of this study is the practical demonstration of the abduction theory in school sciences using mobile learning technologies which was not seriously considered in the previous literature. Further, it can be suggested that this kind of application may be useful to enhance both learning performance and critical thinking skills where students are engaged in exploring and experimenting in real environments. This can promote deeper understanding of a particular science domain and can guide students in interpreting data to create meaningful hypotheses.

Although the results discussed above are promising, there are some limitations to this work. In this study, science teachers had designed pre-post and retention tests. However, for marking these tests, the science teachers were busy in their school work at that time. In consultation with the science teachers about this issue, a marking scheme was agreed with the researchers, following the answers suggested by the science teachers. For the MCQs, it might not make any difference but there may be the potential for bias while marking open-ended questions, related to the hypotheses with their explanations. In addition, this study represents a sample from a single science inquiry context which would need to be repeated in similar contexts to validate our results. We cannot state to what extent these results may be generalizable to other technology-assisted science inquiry based learning activities. Further, we had no control over the grouping of the students, and since they performed the experiments in groups, there may be a chance that they worked together in answering questions and writing hypotheses with their explanations.

As far as future research is concerned, an extended version of ‘ThinknLearn’ may be designed for teachers or instructors so that they may generate learning content according to their needs. For instance, learning content (suggestions) may be provided according to students’ knowledge levels (i.e. novice or expert) or their learning styles. This extended version may be further evaluated in order to comprehend how such adaptivity features in these kinds of applications may affect students’ learning performance and their learning experiences. Moreover, this study only compares the use of ‘ThinknLearn’ with the traditional pedagogy approach (i.e. paper-based). However, this work may be extended to compare these two approaches with the context-aware learning approach which considers content adaptation and device-independence to serve different learners and their situated mobile learning environments.

The Macro level of the evaluation framework is not considered in this study because it is used to examine the longer term impact of the new technology on established learning practices. If this new inquiry practice (i.e. abductive) emerges in supporting science learning in schools, as ‘ThinknLearn’ is designed to do, then we may be able to perform Macro level evaluation in the future.

Acknowledgments

We acknowledge the Higher Education Commission (HEC), Pakistan for financially supporting the principal author’s doctoral research. We are also grateful to Diana Hartley for assisting us in conducting experiments with her students at Albany Senior High School, Auckland, New Zealand.

References


