Journal of Education and Learning (EduLearn)
Vol. 17, No. 1, February 2023, pp. 113~126
ISSN: 2089-9823 DOI: 10.11591/edulearn.v17i1.20630

Effects of technology-integrated formative assessment on students' conceptual and procedural knowledge in chemical equilibrium

Tadesse Hagos[1], Dereje Andargie[2]
[1]Department of Chemistry, Addis Ababa University, Addis Ababa, Ethiopia
[2]Institute of Education, Debre Berhan University, Debre Birhan, Ethiopia

Article Info

Article history:

Received Jul 13, 2022
Revised Sep 23, 2022
Accepted Nov 25, 2022

Keywords:

Conceptual knowledge
Formative assessment
Procedural knowledge
Technology

ABSTRACT

The impacts of a technology-integrated formative assessment technique on students' conceptual and procedural knowledge in studying chemical equilibrium are studied in this study. To attain the purpose, a quasi-experimental pretest-posttest strategy nested with a qualitative study method was adopted. The study has three groups: two experimental and one comparative. A random sample strategy was utilized to select two intact classes for treatment and one intact class for comparison groups. Data was collected using the chemical equilibrium conceptual test, the procedural test, and classroom observation. The data was examined using descriptive (mean and standard deviation) and inferential statistics (one-way ANOVA, one-way MANOVA, and Pearson product moment correlation). According to the findings, technology-integrated formative assessment processes outperformed traditional techniques and formative assessment strategies alone in enhancing students' conceptual and practical understanding of chemical equilibrium. Similarly, when technology-integrated formative assessment processes are used, classroom observations show that students have a strong motivation to learn and that the instructor is more skilled than the other two teachers. Technology-integrated formative assessment processes were shown to be more effective than the other two groups in promoting students' conceptual and procedural understanding when learning chemical equilibrium.

Corresponding Author:

Tadesse Hagos
Department of Chemistry, Addis Ababa University
Addis Ababa, Ethiopia
Email: tadessehagos36@gmail.com

1. INTRODUCTION

Science education reformers have long urged students to engage in the study of science, learn about evidence-based reasoning and higher-order cognitive skills, and be taught how to solve problems creatively [1], [2]. Because of this, the instructional design has gone through a number of adjustments throughout its lifetime. There have been shifts in focus from an external to an interior perspective of learning, from behaviorism to cognitivists to constructivism [2], [3]. At this stage of the learning process, students must be seen as knowledge makers rather than knowledge receivers: students who generate knowledge by relating their prior experiences and knowledge to present situations, and students who have learning techniques to help them do so [4]. As a result, a good and successful education places a strong emphasis on teaching abilities that facilitate students' understanding of the material they are studying. The effectiveness of various learning-supporting instructional strategies in improving students' understanding of science has thus been the subject of several research investigations.

Even though promoting students' learning is a top goal everywhere, including in Ethiopia, many teachers neglect to implement effective assessment methods and instructional strategies that could facilitate scientific concepts and processes [5]. As a result, even after instruction, many students still have misconceptions after taking science classes [6]. One viewpoint holds that students leave science classes with misconceptions because teachers use instruction that primarily emphasizes students' information acquisition as a strategy to help students retain information for the test. This could be challenging in areas like chemistry, where students are frequently required to use scientific principles to resolve challenging algorithmic and conceptual problems [7]. In light of this, one of the main objectives of chemistry education as a component of science education is to understand how students learn chemistry, how to teach chemistry effectively, and how to improve learning outcomes by changing teaching methods and assessment techniques in order to shift students away from memorization of facts and toward understanding and applying core chemistry principles [8].

According to Slout *et al*. [9], in order to develop a higher level of cognitive understanding and processes in chemistry, students must study the macroscopic, sub-microscopic, and symbolic levels of chemical knowledge. Students of chemistry at all educational levels must have a thorough understanding of scientific principles at all three levels as well as the capacity to synthesize their learning [10], [11]. These three levels must be connected in order to understand how chemical knowledge is applied in daily life. However, it affects the others if students have trouble at one of these levels [12].

To address micro-macro thinking capacities, many techniques have been employed in face-to-face settings as well as in technology-enhanced learning (TEL) contexts. These technologically-based solutions have a great deal of potential for displaying dynamic phenomena that change over time and for making the invisible visible. A variety of representations of physical events, including graphs and diagrams, can be created and tested using simulations [13]. Researchers and educators have created TEL environments to focus on one or more micro- and macro-thinking skill areas.

For chemistry to be mastered, understanding chemical concepts and problem solving are essential. Students can use this knowledge to discuss chemical topics, such as what they have discovered in regard to particular chemicals or chemical compounds they've utilized, and they can do experiments to better understand the concepts involved in chemistry. Conceptual knowledge is the understanding of theoretical chemistry and conceptual ideas, whereas procedural knowledge is the ability to apply newly learned concepts in a variety of problem-solving contexts [14]. The results of studies on this topic indicate that while many students were able to solve algorithmic problems, they did not comprehend the chemistry principles tested [15], [16]. Even high-achieving students may not have a fundamental understanding of fundamental principles, as students struggle to connect quantitative representations to underlying chemical concepts [17]. Due to this, many students at all levels, from elementary to tertiary, find chemistry difficult to understand and fail to master it. As a result, many students find it difficult to understand the ever-more-complex concepts that are built upon these fundamental ideas [18]. As a result, students, teachers, and instructors consider chemistry to be a challenging and abstract subject [19].

Many methods have been suggested recently to help students learn chemistry in general and chemical equilibrium in particular. These include the necessity for a methodological shift in the way it is taught, a thorough investigation of misconceptions, and the recognition of the persistent nature of errors [20][20]. In this sense, the most effective kind of strategy for boosting student learning is formative assessment. It frequently aids in the process of individualized instruction, promotes student participation, gathers detailed diagnostic information, and offers prompt feedback. Teachers can utilize formative assessment more successfully with the aid of technology. The majority of educators, however, still struggle to use technology in their formative assessment processes [21]. In order to facilitate the development of 21st century abilities, a movement in pedagogy toward dynamic problem-based and inquiry-based learning is growing in popularity, demanding changes in formative assessments. With the development of new learning technologies, it is now possible to use technology to support formative assessment for learning [22]. With a focus on secondary schools in Addis Ababa, the purpose of this study is to determine how a carefully thought-out formative assessment strategy, which incorporates technology and discipline-based activities, affects students' conceptual and procedural knowledge in learning chemical equilibrium. To address the above objective, the researchers came up with two specific research questions: i) Does the use of technology-integrated formative assessment have an impact on students' conceptual and procedural knowledge? and ii) How may technology-integrated planned formative assessment aid in the teaching and learning of chemical equilibrium during the intervention lessons?

2. RESEARCH METHOD

2.1. Research design

The embedded/nested mixed research design was employed for this investigation. According to Creswell [23], the embedded approach uses a main quantitative method to guide the investigation and a secondary qualitative method to support the processes. In other words, the secondary method is nested within the primary method or integrated within it. In this mixed research design, the researcher can collect both types of data simultaneously through intervention, offering the study the advantages of both quantitative and qualitative data [23].

The current study randomly assigned experimental and comparison groups to examine the impact of interventions on students' conceptual and procedural expertise in studying chemical equilibrium. As a result, the pretest-posttest study design contains one comparison group and two experimental groups. According to the research design, students in experimental group one were exposed to technology-integrated formative formative assessment (TIFA), students in experimental group two were exposed to formative assessment (FA) alone, and students in the comparison group were exposed to existing instruction. Similarly, qualitative data was collected to supplement the quantitative data and provide a more in-depth examination of the treatment's implementation during the teaching and learning process [23]. The diagrammatic representations of the nonequivalent comparison group study design are shown in Table 1. The experimental group is exposed to a variety of treatments, including TIFA and FA alone, which were employed in this research.

Table 1. The diagrammatic representations of nonequivalent comparison group research design

Intervention groups	Pre-test	Treatments	Post-test
Experimental group one	O_1	TIFA	O_2
Experimental group two	O_1	FA alone	O_2
Comparison group	O_1	X	O_2

O_1 = pre-test for the experimental and comparison groups
O_2 = post-test for experimental and comparison groups
TIFA = treatment for experimental group 1 (received technology integrated formative assessment)
FA = treatment for experimental group 2 (received formative assessment alone)
X = treatment for comparison group (received the actual existing instruction

2.2. The participants of the study

The population of this study was 11-grade students in government secondary schools in Addis Ababa, Ethiopia. Out of ten sub-cities of Addis Ababa, three sub-cities were selected using simple random sampling techniques as the target population. Next, from each of the three sub-cities, one secondary school was selected using lottery methods as a sample. Next to this, simple random sampling techniques were employed to select three intact classes within the schools, and the three sections were just randomly assigned, two for treatment and one for comparison. And then, one relatively well-qualified and experienced chemistry teacher was purposefully selected for each school. The study included 132 eleventh-grade students from the selected governmental secondary schools.

2.3. Variables of the study

The intervention groups served as the study's independent variables. Within the intervention group, there are three levels. Comparison method, TIFA, and FA alon. The study's dependent variables were students' conceptual and procedural knowledge scores from tests on chemical equilibrium.

2.4. Data gathering instruments

In order to answer the study's research questions, data were collected using a variety of data collection tools. Data for this study were gathered by classroom observation, conceptual and procedural tests of chemical equilibrium.

2.4.1. Chemical equilibrium conceptual test (CECT)

This test consisted of 25 multiple-choice questions. For each question, there is only one correct response and four distractions. All questions were taken from literature related to chemical equilibrium and modified to suit the demands of the study in order to evaluate the students' learning outcomes in conceptual knowledge. The exam's questions were made to assess the students' broad conceptual knowledge both before and after the intervention. The conceptual test had a minimum and maximum score range of 0 and 45, respectively. All conceptual test items have an internal consistency rating of 0.74 or above [24].

2.4.2. Chemical equilibrium procedural test (CEPT)

The 15 multiple-choice items that make up the CEPT were modified and adjusted for the study's assessment of students' procedural knowledge learning outcomes [25]. Using Le Chatelier's principle, calculating out all the equilibrium constant, Kc, altering Kc with temperature, and comparing Kc and Kp equilibrium constants were among the procedural concepts put to the test in these exams. Similar to the conceptual questions, this exam also included questions to assess the students' general procedural knowledge before and after the treatment. The lowest and highest marks on the procedural test were 0 and 45, respectively. All procedural test items had an internal consistency reliability rating of 0.75 or above [25].

2.4.3. Classroom observation

The researchers spent 45 minutes each observation in each classroom during scheduled class room visits. Three weeks were spent with one classroom observation every week for this reason. For this study, observation data were gathered using the formative assessment classroom observation protocol (FACOP) [26]. The FACOP domains are as follows: domain A, describing learning objectives and success criteria; domain B, facilitating effective classroom discussions (questioning); domain C, facilitating effective classroom discussions (collaboration); domain D, carrying out learning tasks; and domain E, giving feedback on instruction. The activities that teachers and students engaged in at the start, during and end of a particular lesson in terms of the practice of both scheduled lessons and spontaneous learning activities were the primary topics of the classroom observations. All observed courses were videotaped, ensuring that all participants were aware of the researchers' presence.

2.5. Validity and reliability of the instruments

The instrument's items were a collection of questions published by other researchers. The questions reflected all areas of the misconceived topics covering the chemical equilibrium syllabus. The chemical equilibrium conceptual and procedural exams were examined for both content and face validity. Instruments for data collection: experts in chemistry were given pre-conceptual tests, post-conceptual tests, pre-procedural tests, and post-procedural exams. PhD candidates in chemistry education and secondary school chemistry teachers assessed the test items for compatibility with the textbook objectives and the items, as well as for clarity and errors in the answer key. Finally, the expert's opinions and recommendations were taken into account while making adjustments. The content validity of the classroom observation form was further evaluated by two chemistry education specialists. Furthermore, the researchers believed that estimating the internal consistency (the Cronbach alpha coefficient) and dependability of quantitative research instruments during the pilot test was adequate to verify the instrument's reliability, and its results were reported in the pilot study.

2.5.1. Pilot study

To enhance the research methods and quasi-experiment methodologies, a pilot study was conducted. The study was conducted at one school that was not included in the study sample of 40 students in 12th grade. Students who volunteered to assist with the instrument and study design piloted the pre- and post-tests for conceptual and procedural chemical equilibrium. The Kuder-Richardson formula 20 (K-R20) was used to obtain an estimate of the reliability coefficient for chemical equilibrium conceptual tests that was approximately 0.72 and 0.75 for chemical equilibrium procedural tests, respectively.

To evaluate the research design, formative assessment with technology and formative assessment without technology were used for two weeks in a real classroom while teaching the chemical kinetics topic, which is the pre-request of chemical equilibrium. Relevant data was obtained during the implementation of the instructional techniques through classroom observation and conversation with the students and teachers who took part in the pilot research. The time it took to implement the five formative assessment techniques within 45 minutes received a lot of attention. The time allowed for individual and peer activities was insufficient, according to the researcher, especially for those formative assessment groups that just got treatment. When the teacher introduced the lesson objectives and outlined the success criteria for achieving the lesson objectives, the teacher used up the majority of the given time. Based on participant input and classroom observations, the researcher updated how to execute the five formative assessment procedures.

2.6. Preparation of instructional material and intervention procedure

This intervention's teaching materials were created using formative assessment ideas and methodologies. Under unit five of the grade 11th text book, the instructional content addressed five subjects from chemical equilibrium ideas. The notion of chemical equilibrium, equilibrium constants, magnitude of chemical equilibrium constants, chemical equilibrium computations, and Le Châtelier's Principle were among the subjects covered. The study's goal is to increase students' conceptual and procedural understanding of studying chemistry in general and chemical equilibrium in particular. As a result, the researchers devised an instructive activity.

According to a review of the literature, the most challenging concept taught in general chemistry is chemical equilibrium. According to their findings, students' lack of conceptual and procedural knowledge will obstruct meaningful learning of a subject, and conventional teaching methods will not foster these understandings. As a result, in order to develop instructional materials that would allow students to learn both conceptual and procedural knowledge, the researchers utilized a constructivist approach. Both experimental groups employed a student-centered teaching approach to introduce chemical equilibrium throughout the study, in keeping with social-constructivist pedagogy. As a result, the teacher served as a facilitator and a mentor during class discussions, including the students in a process of inductive learning that involved producing meaning by questioning, supervising, validating, and elaborating on ideas.

On the one hand, the formative assessment alone group was exposed to interaction-based formative assessment activities that aim to develop conceptual and procedural knowledge by using a variety of examples of conceptual and procedural problems. On the other hand, every activity in and out of the classroom was delivered without supporting technology. The technology-integrated planned formative assessment group, on the other hand, was explicitly introduced to technology-supported discourse that includes the three elements of macro-micro-symbolic teaching as well as every formative activity supported by technological tools and software over the course of the study. They also received the same content. When teaching chemistry to the comparison group, the teacher followed his usual teaching strategy.

Technological tools used in this study included a computer desktop, a plasma screen, a laptop, a white board, a microphone, and a smart phone. The programs used included Telegram, PowerPoint, and internet access. Making use of such technological software and hardware was intended to facilitate the application of formative assessment methods both within and outside of the classroom. The teacher created the course objectives and success criteria using Power Point, as well as individual and peer formative tasks. The formative exercises were divided into two categories: peer and individual. With the use of a plasma screen and computer desktop activities, the teacher introduced the lesson objectives to the students. The teacher gives enough time for both individual and group discussions on the formative assignments during this time. They discuss their ideas and information with their peers.

In this classroom, the teacher's role was to assist and direct the students. After presenting the formative activities, the teacher also showed the scientific solutions on the plasma screen. To help people understand the idea of chemical equilibrium at different levels, he also downloaded a number of related lecture films and animations (at the micro, symbolic, and macro levels). Additionally, a Telegram group was established by the teacher and the students. The use of telegrams is evident in this group's work, and the teacher usually included conceptual and procedural homework assignments with them so that students may do them at home. Every time a student committed a mistake, the teacher would also send a telegram to let them know. Furthermore, the teacher used this telegram group to link the necessary instructional resources, helping the students develop their conceptual and procedural knowledge.

As a consequence, seven-week courses (totaling 21 periods) were developed based on the chemical equilibrium scope of topics specified in the 11th grade chemistry textbook. To encourage discussion amongst the students, the lessons were usually performed through individual and cooperative group work. Eight groups were established, each consisting of four to five individuals. Teachers used the formative assessment techniques concept mapping, conceptual diagnosis, observation, self-assessment, quiz, portfolio check, oral questions, think-pair-share, think-write-pair-share, one question and one comment, a three-minute pause, and a one-minute essay in the classroom. This means that when teachers use formative assessment strategies to teach for the advancement of students' higher-order cognitive knowledge, they must offer meaningful feedback during each task. On the other hand, in the comparison group, the teacher used the traditional lecture-style course delivery over seven weeks on these five topics. Teachers and students who will participate in the intervention received training after the intervention's instructional materials were developed. The main emphasis of the training was on how to implement formative assessment practices in the classroom. The training was conducted by the researchers and lasted seven days (one hour per day) for the students and fourteen days (two hours per day) for the teachers assigned to this experiment. The program included an in-depth overview of the five different formative assessment techniques as well as a hands-on demonstration of how to develop a formative daily lesson plan using actual examples from the classroom. In addition, crucial instruction on using technology in the classroom was given. After the research period, the conceptual and procedural knowledge exams were administered as a post-test.

2.7. Methods of data analysis

The results obtained from all the instruments administered were coded and analyzed by the researcher. The quantitative data was analyzed using descriptive and inferential statistics. To see if there were any statistically significant differences between the means of two treatments and one comparison group, a one-way ANOVA was used. To examine the influence of the independent factors on both dependent variables at the same time, one-way MANOVA statistics were utilized. Finally, to assess the link between conceptual and procedural knowledge, the Pearson product-moment correlation coefficient was used. The required assumptions were researched and tested prior to the study. The assumptions of univariate and multivariate normality, homogeneity of variances, and variance-covariance homogeneity were examined in this way. Mahalanobis distance values for each dependent variable were computed to determine extreme values in terms of multivariate normality. The statistical package for social sciences (SPSS) computer package version 26 was used for this investigation. Finally, the themes were used for interpretation based on how they related to the research questions during the qualitative data analysis of the study.

2.8. Consideration of ethical issues

This study was carried out after receiving official approval from the school administrator. The research was conducted in accordance with standard ethical guidelines. The participants in the study were asked to give their informed permission, which they did. Participants were informed that their participation was entirely voluntary and that they might withdraw at any moment or refuse to participate in any research-related learning activities. They were also told that their privacy and identities would be respected. There were no names or personal information divulged, and the material was kept private and solely utilized for research purposes. As a result, in all transcripts and written material, the researchers used special codes to conceal the names of the participants and schools (e.g., school A, school B, school C, and teacher A, teacher B, teacher C).

3. RESULTS

3.1. Analysis of quantitative pre-test results among groups

Because there were three groups, pretest mean scores for the two experimental and one comparison group were compared using one-way ANOVA based on data acquired from the pre-administration of the conceptual and procedural knowledge tests. In Tables 2 and 3, the statistical data of each group were evaluated and presented. Table 2 indicates the differences in mean and standard deviation for each group at pretest, based on the two dependent variables under investigation. According to descriptive statistics, the mean value for all dependent variables, such as chemical equilibrium and chemical procedural knowledge, was practically the same for each research groupan value for all dependent variables, such as chemical equilibrium and chemical procedural knowledge, was practically the same for each research group. After the descriptive statistics were analyzed, a one-way ANOVA was used to see if there was a significant difference between groups on their two dependent pre-tests. The assumptions of ANOVA, such as normality and homogeneity of variance, were validated before doing the analysis of pre-test scores. In the three dependent variables, the skewness and kurtosis of the pretest data were within acceptable limits, shown in Table 4. This indicates that the data was fairly regularly distributed. The Levene test, which was not significant for all dependent variables, pre-conceptual and pre-procedural knowledge tests, and other assumptions of ANOVA, such as homogeneity of variance, were also examined, shown in Table 5. This indicates that for the population of the groups, the variance of scores on each measure is similar. As a result, the ANOVA assumptions were not violated.

Table 2. Summary on students' pre-test scores in conceptual test, and procedural test among the three groups

Dependent variable	Group	N	Mean	Std.deviation
Pre-test conceptual knowledge	TIFA group	45	7.87	2.64
	FA group	43	6.95	3.08
	CM group	44	8.27	2.490
	Total	132	7.70	2.78
Pre-test procedural knowledge	TIFA group	45	4.09	1.62
	FA group	43	3.40	2.52
	CM group	44	3.84	1.96
	Total	132	3.78	2.07

Table 3. ANOVA summary table comparing the three groups on scores of pre-test of conceptual and procedural knowledge test scores

Dependent variables	Source	SS	Df	MS	F	Sig.
Pre-test conceptual knowledge	Between groups	39.64	2	19.82	2.63	.076
	Within groups	971.83	129	7.53		
	Total	1011.48	131			
Pre-test procedural knowledge	Between groups	10.82	2	5.41	1.27	.283
	Within groups	547.81	129	4.25		
	Total	558.63	131			

Table 4. Normal distribution analysis study for students' pre-post conceptual and procedural tests among the three groups

DV	Group	Normality test							
		N	Skewness	SE	z-value	Kurtosis	SE	z-value	Sig.
Pre-conceptual knowledge test	TIFA	45	.59	.35	1.69	.97	.70	1.39	.230
	FA only	43	-.05	.36	-.14	.73	.71	1.03	.301
	CM	44	.20	.36	.56	-.26	.70	-.37	.373
Pre-procedural knowledge test	TIFA	45	.08	.35	.23	-.51	.70	-.73	.007
	FA only	43	.49	.36	1.36	.35	.71	.49	.000
	CM	44	.33	.36	.92	.61	.70	.87	.000
Post-conceptual knowledge test	TIFA	45	.06	.35	.17	-1.22	.70	-1.74	.130
	FA only	43	-.59	.36	-1.64	-.65	.71	-.92	.170
	CM	44	-.45	.36	-1.25	-.52	.70	-.74	.120
Post-procedural knowledge test	TIFA	45	-.27	.35	-.77	-1.25	.70	-1.79	.080
	FA only	43	-.07	.36	-.19	.20	.71	.28	.051
	CM	44	.13	.36	.36	-.96	.70	-1.37	.060

Table 5. Levene's test of homogeneity of variances for students' conceptual and procedural knowledge test scores among the three groups

Dependent variables	Levene statistic	df1	df2	Sig.
Pre-test conceptual knowledge	.03	2	129	.969
Pre-test procedural knowledge	1.94	2	129	.148
Post-test conceptual knowledge	1.07	2	129	.345
Post-test procedural knowledge	.44	2	129	.646

The ANOVA analysis revealed that there was no statistically significant mean difference between the comparison and treatment groups for the conceptual and procedural tests: F (2,129)=2.63, p=.076 for the conceptual test, and F (2,129)=1.27, p=.283 for the procedural test, implying that the groups were similar in terms of their conceptual and procedural test scores, shown in Table 3. It's worth noting that the pretest findings for both the treatment and comparison groups are nearly similar. This means that there was no significant difference in the three groups' acquisition of conceptual and procedural knowledge prior to using the technology-integrated planned formative assessment knowledge. As a result, the researchers found that the three groups' mean conceptual and procedural knowledge test scores were similar at the start of the investigations.

3.2. Analysis of quantitative post-test date

The primary goal of this study was to see if there were any significant mean differences between the groups in the two dependent variables of conceptual and procedural exam scores. To assess the effects of groups on the combined dependent variables, the researchers used a one-way MANOVA. To avoid inflating the type 1 error rate in the follow-up ANOVA and post-hoc comparisons, a MANOVA was first run on the means. However, before performing the MANOVA, a Pearson correlation between the dependent variables was performed to test the MANOVA assumption that the dependent variables would be moderately correlated.The dependent variables of conceptual and procedural test scores showed a significant pattern of correlations (r=.25, p=.003), indicating that a MANOVA was suitable. In terms of multivariate normality, the Mahalanobis distance was computed to look for outliers. For the two dependent variables (conceptual and procedural test scores), the estimated maximum Mahalanobis distance value was 8.24, which is less than the critical point (13.82). As a result, no uncommon combinations of scores on the conceptual and procedural test scores of dependent variables in the distribution were discovered.

Second, the multivariate normality distribution was examined; the means of each dependent variable in each cell, as well as all linear combinations of dependent variables, were found to be approximately normal, shown in Table 4. According to Huberty and Petoskey's standards [27], the box's M value of 11.54

was linked with a p value of.080, which was deemed non-significant (i.e., p>.005). For the purposes of the MANOVA, the covariance matrices between the groups were assumed to be similar. According to Box's test, equal variance can be assumed. As a consequence, Wilk's lambda will be utilized as the test statistic, and the MANOVA findings will be presented in Tables 6–9.

Table 6. Means and standard deviations for conceptual and procedural test scores by groups

Group	Conceptual test scores		Procedural test score	
	M	SD	M	SD
TIFA group	18.93	2.30	11.56	1.49
FA only group	16.63	3.86	10.13	1.45
CM group	15.05	4.06	9.45	1.37

Table 7. Multivariate tests for conceptual and procedural test scores by group differences

Effect	Wilk's Lambda value	F	Sig	η^2
Wilks' Lambda	.64	15.85^b	.000	.194

a. Design: Intercept + Group; b. Exact statistic; c. The statistic is an upper bound on F that yields a lower bound on the significance level; d. Computed using alpha = .025

Table 8. ANOVA summary table for separate conceptual and procedural test scores among groups

Source	Dependent variable	Type III SS	Df	MS	F	Sig	η^2
Group	Post- conceptual test scores	340.54	2	170.27	12.71	.000	.17
	Post-procedural test scores	102.85	2	51.43	24.90	.000	.28

a. R Squared=.165 (adjusted R Squared=.152); b. R Squared=.279 (adjusted R Squared=.267); c. Computed using alpha=.025

Table 9. Post hoc results for combined conceptual and procedural test scores by group differences

Dependent variable	Group of students (I)	Group of students (J)	Mean difference (I-J)	Std. Error	Sig.	99.13% confidence interval Lower bound	Upper bound
Post-conceptual test scores	TIFA group	FA group	2.31	.78	.011	-.06	4.68
		CM group	3.89*	.78	.000	1.53	6.24
	FA group	TIFA group	-2.31	.78	.011	-4.68	.06
		CM group	1.58	.79	.138	-.80	3.97
	CM group	TIFA group	-3.89*	.78	.000	-6.24	-1.53
		FA group	-1.58	.79	.138	-3.97	.80
Post-procedural test scores	TIFA group	FA group	1.44*	.31	.000	.51	2.37
		CM group	2.10*	.30	.000	1.18	3.03
	FA group	TIFA group	-1.44*	.31	.000	-2.37	-.51
		CM group	.66	.31	.101	-.27	1.60
	CM group	TIFA group	-2.10*	.30	.000	-3.03	-1.18
		FA group	-.66	.31	.101	-1.60	.27

Note: Based on observed means; The error term is Mean Square (Error)=2.065; *The mean difference is significant at the .0087 level

A one-way multivariate analysis of variance (MANOVA) was used to see if the combination of conceptual and procedural test scores differed across the three groups, Wilk's Lambda=.64, F (2, 129)=15.85, p<.001. This large F shows that on a linear combination of the conceptual and procedural test results, there are significant differences between the intervention groups. The multivariate $\eta 2$=.19 shows that the group factor was responsible for around 19% of the multivariate variation of the dependent variables. As a follow-up to the MANOVA, two one-way ANOVAs on each of the conceptual and procedural test scores were investigated. To protect against Type I error, the researcher can use a traditional Bonferroni procedure and test each ANOVA at the .025 level (.05/2). Follow-up univariate ANOVAs revealed that the three groups' conceptual and procedural test results were significantly different: F(2,129)=12.71, p<.001, $\eta 2$=.17 and F(2,129)=24.90, p<.001, $\eta 2$=.28, respectively, see in Table 9. The TIFA group had somewhat greater levels of conceptual knowledge (M=18.93, SD=2.30) and procedural knowledge (M=11.56, SD=1.49) than the FA group, which had slightly lower levels of conceptual knowledge (M=16.63, SD=3.86) and procedural knowledge (M=10.13, SD=1.45). The FA group alone and the CM group exhibit nearly similar levels of conceptual and procedural knowledge in their mean scores, shown in Table 6.

Despite the fact that the between-subjects effect tests revealed a statistically significant mean score difference between groups in all linear combinations of dependent variables, the results did not reveal which group was different from the other. As a result, post hoc multiple comparisons were used to determine which group distinguished itself from the others. Both univariate ANOVAs had previously been tested at the.025

alpha level to account for Type I error. The Bonferroni approach was also used to control for Type I error across the pairwise comparisons of conceptual and procedural knowledge test scores, and each comparison was tested at the alpha level for the ANOVOs divided by the number of comparisons (.025/3=.0083).

With the exception of the TIFA and CM groups, the post-hoc multiple comparison result indicated that there was no statistically significant mean difference in post-conceptual test mean scores between any pair of groups (p>.0087). Post hoc multiple comparison findings, on the other hand, indicated a statistically significant mean difference in post-procedural test mean scores between each pair of groups (p<.0087), with the exception of the FA alone group and the CM group (p>.0087). In every case, the impact had a linear trend. That is, the TIFA group had higher mean scores in learning outcomes on average than the FA alone group, while the FA alone group had higher mean scores in learning outcomes on average than the CM group. Table 8 shows the effect size calculated using Cohen's r. It can be demonstrated that the procedural had the greatest impact, with an average Cohen's value r of.28, indicating a greater effect in accordance with Cohen's [28] recommendations.

3.3. Analysis of classroom observation data

The researchers looked at five dimensions of classroom practice selected from research on successful formative assessment practices to see how often teachers and students use formative assessment and adjust their instruction depending on formative data. Learning intents and success criteria, engineering successful classroom discussions (questioning), engineering effective classroom talks (collaboration), learning tasks (executed), and feedback on teaching are the five formative assessment strategies. A classroom observation rubric was used to grade the practice of the five formative assessment procedures during the observation classes. Each domain had four to five components that the observer evaluated and scored on a one-to-four-point scale (1=beginning, 2=developing, 3=effective, and 4=exemplary). In order to explore the data by subscale and compare it qualitatively with the field note data acquired, descriptive statistics were used to assess components within each subscale and uncover variances across the three teachers. We compared these average scores to pre-established score ranges, it can bee seen in Tables 10 and 11.

Table 10. Descriptive statistics result from classroom observation checklist

Domain of formative assessment	Teacher C			Teacher A			Teacher B		
	Obs.1	Obs.2	Obs.3	Obs.1	Obs.2	Obs.3	Obs.1	Obs.2	Obs.3
Domain A	Descriptive statistics for sharing learning intentions and success criteria								
Connection to future learning	2	2	1	2	2	2	2	2	3
Learning goal quality	1	1	1	2	2	2	3	3	3
Learning goal implementation	1	1	1	2	2	2	3	3	3
Presentation of criteria	1	2	1	2	2	2	3	3	3
Average	1.25	1.5	1	2	2	2	2.75	2.75	3
Domain B	Descriptive statistics for effective classroom questioning								
Use of questioning	1	2	1	3	3	3	3	3	3
Wait time	1	1	1	2	3	3	3	3	3
Eliciting evidence of learning	1	1	1	3	3	3	3	3	3
Determining progress	1	1	2	2	3	3	2	3	3
Average	1	1.25	1.25	2.5	3	3	2.75	3	3
Domain C	Descriptive statistics for effective classroom collaboration								
Climate	2	2	2	3	3	3	3	3	3
Student collaboration	2	2	2	2	3	3	3	3	3
Student viewpoints	2	2	2	2	2	3	3	3	3
High expectations	1	2	2	2	2	2	2	3	3
Average	1.75	2	2	2.25	2.5	2.75	2.75	3	3
Domain D	Descriptive statistics for learning tasks implementation								
Connection to learning goals	1	1	1	3	3	3	3	4	4
Clarity of task	1	2	1	3	3	3	3	3	4
Adjust instruction within the lesson	1	1	1	3	3	3	3	3	3
Use evidence to inform future instruction	1	1	2	2	3	3	3	3	3
Average	1	1.25	1.25	2.75	3	3	3	3.25	3.5
Domain E	Descriptive statistics for instructional feedback								
Assessing progress during lesson	1	2	1	2	2	2	2	3	3
Individualized feedback	1	1	1	1	1	1	1	1	1
Self-assessment	1	1	1	2	1	1	2	2	2
Peer assessment	1	1	2	2	1	1	2	2	2
Feedback loops	1	1	1	1	1	1	2	2	2
Average	1	1	1.2	1.2	1.2	1.2	1.8	2	2

Table 10. Descriptive statistics result from classroom observation checklist (*continued*)

Domain of formative assessment	Teacher C			Teacher A			Teacher B		
	Obs.1	Obs.2	Obs.3	Obs.1	Obs.2	Obs.3	Obs.1	Obs.2	Obs.3
	Descriptive statistics by subscale in the FACOP								
	Teacher C			Teacher A			Teacher B		
Learning intentions and criteria	1.25	1.5	1	2	2	2	2.75	2.75	3
Effective classroom questioning	1	1.25	1.25	2.5	3	3	2.75	3	3
Effective classroom collaboration	1.75	2	2	2.25	2.5	2.75	2.75	3	3
Learning tasks implemented	1	1.25	1.25	2.75	3	3	3	3.25	3.5
Feedback on instruction	1	1	1.2	1.2	1.2	1.2	1.8	2	2
Average	1.2	1.4	1.675	2.14	2.34	2.39	2.61	2.8	2.9
Total mean		1.43			2.29			2.77	

Table 11. FACOP score value with corresponding average score range

Score value	Score range
Beginning	1.00-1.75
Developing	1.76-2.75
Effective	2.76-3.75
Exemplary	3.76-4.00

3.4. Learning intentions and criteria for success

Teachers showed greater strengths in this domain in connecting current lessons to future learning and addressing the learning goal throughout the lesson (learning goal implementation). They introduce the lesson content, instructional objectives, and the type of formative assessment to be used. In many of the classes, the majority of the students were observed while taking notes. Some students were also listening attentively to the teacher without taking notes. Whenever the students attempt to fulfill the assessment criteria, the teachers motivate them to a large extent. Among the three teachers, A and B were ranked at developing and effective levels, respectively, in all areas of the sharing learning intentions and criteria for success subcategory. On the other hand, teacher C was ranked at the beginning level for the sharing learning intentions and criteria for success subcategory.

We did not see a lot of evidence in the three teachers' observations that they shared the success criteria with the students. The main shortcomings identified during the observations included the teacher who participated in the control group's failure to introduce instructional objectives, the students' inactivity and limited participation, and the inappropriate seating arrangements. The teacher who took part in the technology-integrated formative assessment group, in contrast, had a greater rate of carrying out learning objectives than the other two teachers.

3.5. Engineering effective classroom discussions (questioning)

This method, according to our observations, is demonstrated by the questioning style (more probing questions), the wait time for answers, the gathering of learning evidence (revealing students' thinking), the assessment of learner progress, and the use of the evidence to modify teaching. Teachers A and B were rated as having the best use of questioning techniques among the teachers. Teacher C, on the other hand, received a starting level rating for the use of questioning techniques. Although each teacher approached the students' self-assessment differently, the general procedures were essentially the same. Except for the teacher who took part in the control group, two of the other teachers claimed that questioning techniques were employed successfully during their observed lessons. The teacher who had taken part in the control group's lecturing method, on the other hand, was utilized to deliver the lesson's material by posing some convergent questions.

Teachers in the experimental study infused questions throughout the lesson to determine student understanding, provided appropriate wait time, and gauged student progress based on classroom discourse and interactions. In the meantime, some students attempted to answer questions while others listened attentively. Here are some examples of effective practices we observed in this domain: i) Questioning: In the TIFA group, the teacher asked both divergent and convergent questions throughout a lesson on the dynamic nature of chemical equilibrium. He asked students to build on one another's predictions and descriptions of scientific experiments and pushed for detailed responses to his questions. For example, he asked specific questions about macro-micro thinking; and ii) Eliciting evidence of learning: To systematically elicit evidence from all students throughout the lesson, a TIFA teacher had his students respond to daily warm-up questions and complete individual practice problems by displaying them on the plasma screen. When reviewing students' responses, the teacher was able to walk around and immediately provide feedback and ask follow-up questions. At the end of class, students submitted responses from their small work group via Telegram.

3.6. Engineering effective classroom discussions (collaboration)

The classroom climate, usage of small group discussions, utilization of student perspectives, expressing expectations to the learners, and classroom interactions were all indicators of an effective classroom discussion (collaboration) technique used in engineering, according to the researchers (student-teacher and student-student). All constructs related to promoting classroom collaboration were rated at an effective level for Teacher B. For instance, students were seen working effectively in groups while exchanging various perspectives and ideas with one another. Additionally, interactions between students and teachers had improved, and the instructor was now investing more time in coaching, engaging, and encouraging students rather than directing their learning. However, for successful classroom collaboration, teachers A and C were graded at a developing level. Here are some examples of effective practices we observed in this domain: i) Student collaboration: In FA and TIFA groups, teachers had students' desks organized into groups of five. The groupings encouraged regular discussions and collaboration on assignments without requiring students to move around. Most students seemed used to working with their groups, and most were actively engaged with their peers during discussions and collaboration; and ii) Student viewpoints: In the TIFA group, the teacher gave students conceptual and procedural problems to complete in small groups. The teacher encouraged each group to develop conceptual and procedural knowledge. Groups then shared a range of ideas and solution pathways for the same problem with the entire class.

3.7. Learning tasks (implemented)

The researchers found evidence of the connection to learning objectives (congruence) for the engineering (learning tasks) strategy as well as the clarity of the tasks (transparency), the relevance of the tasks to real-world problems (authenticity), student autonomy (student consultation on tasks), and personalized tasks (student capabilities). For the learning task domain, teachers A and B were both assessed at an effective level. Teacher C, in contrast, received a starting level rating for the implementation of the learning task domain. Teachers A and B made sure that the students understood the lesson's objective and the activities they had to do. They often checked for understanding with the class throughout the lesson and adjusted as needed. Teacher C, on the other hand, frequently missed chances to draw conclusions about development and critical times to modify instruction based on student knowledge. More specifically, in order to create a safe environment where students might feel confident, teachers A and B gave good and encouraging attention. Students are therefore inspired and stimulated to study about chemical equilibrium and actively participate in class. More significantly, the professors answer the students' inquiries favorably and logically correct their errors. Additionally, the TIFA group's teacher circles (moves about) the classroom, pays close attention to the contributions that students make during class discussions, takes notes, and poses questions. Notably, the educator demonstrates his concern for his charges by providing them with the same chances and chances to engage in the learning process as he does. Most frequently, the teacher puts out a lot of effort to encourage open communication and interpersonal interaction among the pupils while also setting high standards for them. They become interested, motivated, and totally focused on the instruction as a result. The ultimate goal of these possibilities is to foster a sense of class belonging and promote cooperative involvement in class activities. As a result, the students feel at ease.

3.8. Feedback on instruction

Feedback on instruction, as indicated by focused and action-oriented feedback on learning goals, self-assessment, individualized feedback, peer assessment, and feedback loops, was noted. All the teachers in the study scored lower in the implementation of strategies for giving feedback. All constructs in this area were scored at the beginning level, with the exception of teacher B, who scored at the developing level on the construct after assessing progress during the lesson. Students in the TIFA group were given a chance to mark their peers' assessment tasks and were given time to comment and give feedback on the performance of their peers against the instructional objectives and the assessment criteria introduced.

On the other hand, issues utilizing individual input were the typical restrictions seen in the observed classrooms. Additionally, it was discovered that some of the students in the observed courses were hesitant and unwilling to grade, provide feedback, and comment on their peers' assessment results. This was mainly because self- and peer-assessment practices were not observed across all classrooms. During the class, teachers in the experimental groups were seen going over student work and giving students meaningful comments in real time. Teachers made a very limited effort to offer personalized feedback to each student in each group, but they did not give them the chance to internalize it or use it in a useful way.

4. DISCUSSION

A new version of Bloom's taxonomy was published in 2001. The previous taxonomy was reassessed, and the revised Bloom's taxonomy was established, taking into account recent advancements in the field of education, students' learning styles, and new assessment and evaluation methods [26]. The updated taxonomy's major focus is on "what to learn" and "how to learn." As a result, the revised Bloom taxonomy represents a significant advancement in learning outcomes. It contains four layers of knowledge: factual, conceptual, procedural, and metacognitive. The scope of this study was confined to conceptual and procedural knowledge. As a result, the major goal of this research was to see how successful technology-integrated formative assessment was at improving students' conceptual and procedural understanding in studying chemical equilibrium.

A one-way MANOVA was used to assess the linear combination effect of the conceptual and procedural knowledge dependent variables across the three groups. It was discovered that the MANOVA influence was statistically significant. This large F demonstrates that there are considerable differences between the three groups based on a linear combination of the conceptual and procedural test scores. The group factor accounted for about 19% of the dependent variables' multivariate variation, according to the multivariate partial eta square.

The conceptual and procedural knowledge levels of the TIFA group were marginally higher than those of the FA alone group, which were marginally lower. In their mean scores, the FA and CM groups demonstrate roughly comparable levels of conceptual and procedural competence. The FA and CM groups exhibit nearly identical levels of conceptual and procedural expertise in their mean scores, shown in Table 6. Additionally, the results of the classroom observation demonstrate that teachers in the experimental groups saw formative assessment as a crucial component of their teaching methods and felt encouraged in doing so.

The findings of this study support other expert assertions that utilizing technology to enhance learning has a favorable effect [22], [29]. Technology platforms can be used by teachers to collect formative data [30]. To better capture what students know and do not know, [31] developed formative assessment software that makes use of the tablet PC's handwriting capabilities. When this instructional method was used, students in the experimental group showed greater learning gains than those in the control group, who received traditional lectures. Elmahdi *et al.* [21] Investigated how technology-assisted formative assessment affected students' learning. They discovered that employing Plickers for formative evaluation increases student engagement, frees up classroom time, provides fair participation opportunities, and fosters a fun and stimulating learning environment.

Kowalski *et al.* [32] conducted research on the effects of computer-assisted formative evaluation feedback on students' arithmetic learning. Two treatment groups and one control group were chosen at random for them. The control group just read pertinent materials and received no feedback following formative assessments, while the first group received in-depth instruction-based feedback, the second group received dichotomous verification feedback, and the third group received no feedback at all. Students who used verification codes scored higher on the post-test for conceptual knowledge than students who received longer remarks. The authors came to the conclusion that students must be actively involved in the learning process for extensive feedback to be effective.

Song and Sparks [1] Used game-based formative assessment to compare the usefulness of two types of feedback (answer-only versus explanatory feedback) for the argumentation abilities of 106 sixth and seventh graders. Some game elements, such as interaction, rules and limitations, challenges, objectives, and rapid task-level feedback, are included in the lesson and shown onscreen so that students may assess their current performance and development. Students who received explanatory feedback improved their reasoning abilities somewhat more than those who received simple answer-based feedback. Although most students fared similarly across feedback situations, highly skilled students did worse on explanatory input than on answer-only feedback. The current findings, on the other hand, contradict those of [33], who employed formative assessment software with laptops and tablet PCs, allowing students to write or draw their responses. According to the researchers, there was no link between the score on writing tasks and overall conceptual knowledge. Maier *et al.* [33] for example, we investigated the impact of mobile-friendly web-based formative assessment software. When compared to a control group that did not get technological assistance, the learners did not accomplish substantial learning with this new configuration.

5. CONCLUSION

Student assessment is needed to determine progress and performance, plan teaching and learning growth, and share information with relevant stakeholders. A summative evaluation examines whether education has met its goals and objectives, whereas a formative assessment evaluates students' performance during and after the teaching-learning process to determine how much they've learned. This required developing and implementing assessment processes to quantify their influence on student learning. Integration of technology into classrooms is crucial for successful teaching that enhances learning, especially in the 21st century, when students' demand for technology and digital tools inspire and encourage them to study. This study evaluated the influence of technology-integrated formative assessment on students' conceptual and procedural chemistry knowledge. The study indicated that when teachers apply technology-supported formative classroom assessment with timely feedback, pupils' academic performance improves. They were introduced to teacher-guided individual and peer evaluation, projects, and group tasks instead of exams and take-home assignments. Based on this study's findings, formative assessment for diagnostic purposes improves students' academic performance and helps them better understand topic knowledge than summative examinations. Formative tests can detect content challenges. The instructor can then deliver remedial and corrective activities to improve pupils' subject understanding and learning results.

While conducting this research, many limitations need to be taken into account. Three teachers conducted this study in three different schools. Even though the researcher tried to choose similar schools and teachers from various perspectives, provided training on instructional approaches and methods of implementation, and conducted classroom observations and discussions with the teachers during implementation to ensure the research protocols were followed, it is difficult to control all teacher- and school-related variables, so the intervention may be influenced by the teachers and the schools. The fact that this study was conducted in a natural setting since the participants designed the courses was another drawback.

ACKNOWLEDGEMENTS

The first author has been working on this part of a PhD dissertation under the supervision and with the assistance of the second author. The Education Bureau of Ethiopia's Amhara Regional State and Addis Ababa University have partnered to support the complete PhD program. We, the aforementioned authors, hereby declare that neither the university nor the Bureau have any materially conflicting financial, professional, or institutional interests that could have affected the decision to publish this paper in this journal or elsewhere. We wish to thank many others for their assistance, support, and guidance.

REFERENCES

[1] Y. Song and J. R. Sparks, "Building a game-enhanced formative assessment to gather evidence about middle school students' argumentation skills," *Educational Technology Research and Development*, vol. 67, no. 5, pp. 1175–1196, 2019, doi: 10.1007/s11423-018-9637-3.

[2] R. J. Amineh and H. D. Asl, "Review of constructivism and social constructivism," *Journal of Social Sciences*, vol. 1, no. 1, pp. 9–16, 2015.

[3] P. A. Ertmer and T. J. Newby, "Behaviorism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective," *Performance Improvement Quarterly*, vol. 26, no. 2, pp. 43–71, 2013, doi: 10.1002/piq.21143.

[4] V. Arghode, E. W. Brieger, and G. N. McLean, "Adult learning theories: implications for online instruction," *European Journal of Training and Development*, vol. 41, no. 7, pp. 593–609, 2017, doi: 10.1108/EJTD-02-2017-0014.

[5] L. Darling-Hammond, L. Flook, C. Cook-Harvey, B. Barron, and D. Osher, "Implications for educational practice of the science of learning and development," *Applied Developmental Science*, vol. 24, no. 2, pp. 97–140, 2020, doi: 10.1080/10888691.2018.1537791.

[6] S. J. Patil, R. L. Chavan, and V. S. Khandagale, "Identification of misconceptions in science: Tools, techniques & skills for teachers," *Aarhat Multidisciplinary International Education Research Journal (AMIERJ)*, vol. 8, no. 2, pp. 466–472, 2019.

[7] K. Broman and I. Parchmann, "Students' application of chemical concepts when solving chemistry problems in different contexts," *Chemistry Education Research and Practice*, vol. 15, no. 4, pp. 516–529, 2014, doi: 10.1039/C4RP00051J.

[8] J. D. Sewry and S. A. Paphitis, "Meeting important educational goals for chemistry through service-learning," *Chemistry Education Research and Practice*, vol. 19, no. 3, pp. 973–982, 2018, doi: 10.1039/C8RP00103K.

[9] M. Slouf, B. Strachota, A. Strachota, V. Gajdosova, V. Bertschova, and J. Nohava, "Macro-, Micro- and Nanomechanical Characterization of Crosslinked Polymers with Very Broad Range of Mechanical Properties," *Polymers*, vol. 12, no. 12, p. 2951, 2020, doi: 10.3390/polym12122951.

[10] J. K. Gilbert and D. Treagust, *Multiple representations in chemical education*, vol. 4. Dordrecht: Springer, 2009.

[11] A. H. Johnstone, "Chemical education research: Where from Here?," in *University Chemistry Education*, 2000, pp. 34–48.

[12] Z. D. Kirbulut and O. Geban, "Using Three-Tier Diagnostic Test to Assess Students' Misconceptions of States of Matter," *EURASIA Journal of Mathematics, Science and Technology Education*, vol. 10, no. 5, 2014, doi: 10.12973/eurasia.2014.1128a.

[13] C. Blake and E. Scanlon, "Reconsidering simulations in science education at a distance: features of effective use," *Journal of Computer Assisted Learning*, vol. 23, no. 6, pp. 491–502, 2007, doi: 10.1111/j.1365-2729.2007.00239.x.

[14] K. S. Taber, *The Nature of the Chemical Concept: Re-constructing Chemical Knowledge in Teaching and Learning*, 1st ed., vol. 3. Royal Society of Chemistry, 2019.

[15] M. H. Chiu, "Algorithmic problem solving and conceptual understanding of chemistry by students at a local high school in Taiwan," in *Proceedings-National Science Council Republic of China Part D Mathematics Science and Technology Education*, 2001, pp. 20–38.

[16] R. Hanson and S. Acquah, "Enhancing concept understanding through the use of micro chemistry equipment and collaborative activities," *Journal of Education and Practice*, vol. 5, no. 12, pp. 120–130, 2014.

[17] K. Osman and N. S. Sukor, "Conceptual understanding in secondary school chemistry: A discussion of the difficulties experienced by students," *American Journal of Applied Sciences*, vol. 10, no. 5, pp. 433–441, 2013, doi: 10.3844/ajassp.2013.433.441.

[18] N. Reid, "A scientific approach to the teaching of chemistry. What do we know about how students learn in the sciences, and how can we make our teaching match this to maximise performance?," *Chemistry Education Research and Practice*, vol. 9, no. 1, pp. 51–59, 2008, doi: 10.1039/B801297K.

[19] M. W. Melaku, A. Harrison, and E. Temechegn, "The Effect of Conceptual Difficulties of Undergraduate Chemistry students' Understanding of Energy," *Journal of Education, Society and Behavioural Science*, vol. 4, no. 9, pp. 1290–1299, 2014.
[20] J. D. Chapman and D. N. Aspin, "A problem-solving approach to addressing current global challenges in education," *British Journal of Educational Studies*, vol. 61, no. 1, pp. 49–62, 2013.
[21] I. Elmahdi, A. Al-Hattami, and H. Fawzi, "Using Technology for Formative Assessment to Improve Students' Learning," *Turkish Online Journal of Educational Technology-TOJET*, vol. 17, no. 2, pp. 182–188, 2013.
[22] C. C. Johnson, T. A. Sondergeld, and J. B. Walton, "A Study of the Implementation of Formative Assessment in Three Large Urban Districts," *American Educational Research Journal*, vol. 56, no. 6, pp. 2408–2438, 2019, doi: 10.3102/0002831219842347.
[23] J. W. Creswell, *Research design : qualitative, quantitative, and mixed methods approaches*, 4th ed. Thousand Oaks, California: SAGE Publications, 2014.
[24] G. Demircioğlu, H. Demircioğlu, and M. Yadigaroğlu, "An Investigation of Chemistry Student Teachers' Understanding of Chemical Equilibrium.," *International Journal on New Trends in Education & their Implications (IJONTE)*, vol. 4, no. 2, p. 185, 2013, [Online]. Available: http://proxy.kennesaw.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edb&AN=90519462&site=eds-live&scope=site.
[25] A. Mensah, "Promoting conceptual change in chemical equilibrium through metacognition development: Students' achievement and metacognitive skills," North-West University, South Africa, 2017.
[26] H. Özmen, "Determination of students' alternative conceptions about chemical equilibrium: a review of research and the case of Turkey," *Chemistry Education Research and Practice*, vol. 9, no. 3, pp. 225–233, 2008, doi: 10.1039/B812411F.
[27] C. J. Huberty and M. D. Petoskey, "Multivariate Analysis of Variance and Covariance," in *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, Elsevier, 2000, pp. 183–208.
[28] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Erlbaum, 1988.
[29] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman, 2001.
[30] M. L. Shirley and K. E. Irving, "Connected Classroom Technology Facilitates Multiple Components of Formative Assessment Practice," *Journal of Science Education and Technology*, vol. 24, no. 1, pp. 56–68, 2015, doi: 10.1007/s10956-014-9520-x.
[31] G. Aldon and M. Panero, "Can digital technology change the way mathematics skills are assessed?," *ZDM*, vol. 52, no. 7, pp. 1333–1348, 2020, doi: 10.1007/s11858-020-01172-8.
[32] F. V Kowalski *et al.*, "Using InkSurvey with pen-enabled mobile devices for real-time formative assessment II. Indications of effectiveness in diverse educational environments," *Workshop on the Impact of Pen and Touch Technology on Education*, pp. 307–314, 2013.
[33] U. Maier, N. Wolf, and C. Randler, "Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types," *Computers & Education*, vol. 95, pp. 85–98, 2016, doi: 10.1016/j.compedu.2015.12.002.

BIOGRAPHIES OF AUTHORS

Tadesse Hagos earned a Bachelor of Teaching (B.Ed.) in chemistry from Jimma University, a Master of Science (M.Sc.) in physical chemistry from Haramaya University, and a Doctor of Philosophy (Ph.D.) in chemistry education from Addis Ababa University. He currently works as a teacher of secondary education in private schools. He can be contacted at email: tadessehagos36@gmail.com.

Dereje Andargie earned B.E.D in Chemistry Haramaya University, MA in Curriculum and Instructiojn in Addis Ababa University, PhD in MSTE (Chemistry Education).from University of South Africa. He currently works in supervising PhD and M.Ed. students in chemistry in Addis Ababa and Bahidar Universities; as an Assistant professor in Science Education (Chemistry Education) and a vice president for academic affairs. Personal contact address: email: derejefamaye@gmail.com.