

Integrating Assessment for Learning into the Teaching and Learning of Secondary School Biology in Tanzania

ALBERT TARMO¹

∞ The paper is about a study that investigated how the integration of assessment for learning enhances learning achievement among secondary school biology students in Tanzania. A quasi-experimental design involving pre-test and post-test of non-equivalent control and experimental groups was used to ascertain how the integration of assessment for learning into teaching and learning processes enhances students' learning achievement. Two boarding secondary schools located in the suburbs of Dar Es Salaam were selected. Students in the two schools had maintained equivalent performances in national examinations in previous years. The results showed that the students taught using teaching and learning processes integrating assessment for learning outperformed those taught using conventional approaches. The integration of assessment for learning is likely to have contributed to the higher learning achievement in the experimental group. The study contributes to our understanding of how teachers in resource-constrained classrooms can integrate assessment for learning techniques into their day-to-day lessons, thereby harnessing the power of assessment to enhance learning and raise standards.

Keywords: assessment, educational assessment, learning achievement, Tanzania

1 School of Education, University of Dar es Salaam, Tanzania; tarmo50@outlook.com.

Vključevanje ocenjevanja za učenje v poučevanje in učenje pri pouku biologije v Tanzaniji

ALBERT TARMO

≈ Prispevek predstavlja raziskavo, ki je preučevala, kako vključevanje ocenjevanja za učenje izboljša učne dosežke učencev biologije v Tanzaniji. Uporabljen je bil kvaziekperimentalni model, ki je vključeval pred- in potest neekvivalentne kontrolne in eksperimentalne skupine z namenom ugotavljanja, kako vključevanje ocenjevanja za učenje v proces poučevanja in učenja izboljša učne dosežke. Izbrani sta bili dve šoli iz predmestja Dar Es Salaam. Učenci obeh šol so imeli enake dosežke pri nacionalnem preverjanju znanja v letu pred izvedbo raziskave. Rezultati so pokazali, da so bili uspešnejši učenci iz skupine poučevanja z uporabo ocenjevanja za učenje kot tisti, ki so bili deležni poučevanja s konvencionalnimi pristopi. Vključevanje ocenjevanja za učenje je verjetno prispevalo k boljšim učnim dosežkom eksperimentalne skupine. Raziskava prispeva k razumevanju, kako lahko učitelji v z viri omejenih okoliščinah vključujejo ocenjevanje za učenje v vsakodnevno poučevanje ter tako izkoristijo moč ocenjevanja za izboljšanje učenja in dvig standardov.

Ključne besede: ocenjevanje, izobraževalno ocenjevanje, učni dosežki, Tanzanija

Introduction

Assessment is widely considered a powerful tool for enhancing students' learning achievement when embedded in the teaching and learning process (Black & Wiliam, 2018; Ellegaard et al., 2018; Wiliam et al., 2004; Wiliam, 2011). When integrated into the teaching and learning process, assessment serves to elicit evidence about students' learning progress. For example, assessment during instruction provides opportunities for students to display their understanding and uncover the strengths and weaknesses of their thinking (Greenstein, 2010). Teachers and learners can use such evidence to make decisions about subsequent learning steps (Wiliam, 2011). For instance, teachers can tailor subsequent lessons in response to students' learning needs and support their students' cognitive growth (Greenstein, 2010).

Moreover, when teachers ask questions to elicit students' prior experiences at the start of a new lesson, they generate evidence that becomes readily available to inform instructional decisions. Teachers can use information about students' prior knowledge to determine students' learning needs for a new lesson. For both teachers and students, learning needs are the bases for planning lessons and setting learning objectives and expectations. When such objectives are made explicit to students, they can take charge of their learning and work towards meeting these expectations (Wiliam, 2011). Most importantly, information about students' prior knowledge helps both teachers and students make connections between previous lessons and new topics to enhance meaningful learning (Greenstein, 2010).

Assessment during teaching allows teachers to continuously check students' learning and adjust instructional processes to meet learners' just-in-time needs (Wylie & Lyon, 2015). For example, teachers may intersperse their verbal descriptions with questions-and-answers to test students' comprehension of a topic. As students respond to questions, teachers can spot individual learners who are struggling to learn certain concepts or skills. In such cases, teachers may provide feedback that identifies gaps in students' thinking and redirect learning by showing the next steps students need to follow (Greenstein, 2010).

Using feedback mechanisms, teachers can focus students' attention on the areas in which they have demonstrated learning success and those that require more practice. Moreover, teachers can support learners to devise learning plans for achieving desired learning outcomes. Teachers facilitate students in directing their learning and make them active participants when they supply the required information about learning progress and provide support for subsequent learning steps. Most importantly, evidence of learning success

motivates students and enhances their resilience when faced with learning difficulties that are within their capabilities (Berry, 2008). Generally, assessment shapes subsequent instruction and learning when teachers and students have continuous access to evidence showing learners' current levels of learning.

Formative assessment (FA) and assessment for learning (AfL) are two closely related and widely used concepts to describe the use of assessment to enhance future teaching and learning (Black & Wiliam, 2009; Wiliam, 2011). The meanings of these concepts remain widely debated (Jonsson et al., 2015; Hopfenbeck, 2018; Wiliam, 2011). Black and Wiliam (2005) defined FA as activities by teachers and students aimed at generating information about students' learning progress and the use of such information as feedback to modify teaching and learning processes to meet learners' needs. In the contemporary literature, however, the term "assessment for learning" rather than "formative assessment" is favoured for describing assessment that promotes learning (Black & Wiliam, 2018; Broadfoot et al., 2002; Hopfenbeck, 2018). This is because FA is used in diverse ways. For example, in some contexts, FA is conceived as "early warning summative" assessment that provides information about the "likely performance of students on the state mandated tests" (Wiliam, n.d., p. 4). Feedback is given to students telling them the items they got right and wrong regardless of the use they make of such feedback (Wiliam, n.d.). In the Tanzanian context, FA often means regular monthly, terminal and annual testing to reduce overdependence on the single final examination that students sit at the end of each education cycle (Kyaruzi et al., 2018). On the other hand, the Assessment Reform Group defined AfL as "the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there" (Broadfoot et al., 2002, p. 2). The present study uses the term "assessment for learning" and draws on the key attributes of assessment that enhance learning as summarised by the Assessment Reform Group (Broadfoot et al., 2002).

The idea of integrating assessment into instruction to enhance learning has been widely embraced at the national and regional levels (Hopfenbeck & Stobart, 2015), to the extent that its adoption has been described as a "research epidemic" (Steiner-Khamsi, 2004, p. 2). AfL is a tool for enhancing learning by making learning expectations explicit to students and providing them with continuous feedback in order to inform them about their learning progress and the next steps they need to take to improve their learning achievement (Hopfenbeck, 2017). Cases of large-scale implementation include Sweden (Jonsson et al., 2015) and four high-needs US districts (Wylie & Lyon, 2015), where AfL successfully transformed assessment practices and improved the collection of evidence about

students' learning through questions-and-answers. In the Tanzanian context, AfL has received policy attention despite the scarcity of exemplary implementation practices at the classroom level, as further discussed below.

Learning Assessment in Tanzania

In the latest revision of the secondary education curriculum, the government of Tanzania stressed the need to integrate assessment activities with everyday instruction using authentic approaches such as practical tasks, project work, portfolios and verbal questioning (Ministry of Education and Vocational Training (MoEVT), 2007). The aim was to widen the range of learning achievement that could be assessed and use the information to guide and improve teaching and learning processes. Such assessment is aimed at promoting learning through building confidence and developing students' belief in their capacity to attain learning success. This assessment is envisioned to be formative in nature, as it monitors learning progress throughout a given education cycle (MoEVT, 2007). Generally, the curriculum calls for a change in assessment approach by adopting AfL and minimising overdependence on paper-and-pencil tests. However, local research suggests that efforts to improve learning achievement rarely make use of assessment as a means of raising standards (Kira et al., 2013; Kitta & Tilya, 2010; World Bank, 2008). High-stake, large-scale and centrally administered examinations, which are used for certification and placement purposes, remain dominant in Tanzania (Kyaruzi et al., 2019). Such examinations have lasting effects on students' life chances because the results are used to select students for highly valued places in further education and workplaces.

The government introduced Continuous Assessment (CA) to reduce overdependence on high-stake examination, assess students on a continual basis, and combine results with those obtained in final examinations to determine students' final grades (Kyaruzi et al., 2019). However, studies suggest that teachers often do not implement CA in such a way that the information collected could be used to improve instruction (Lema & Maro, 2018). Instead, teachers' assessment practices largely mimic the system-wide high-stake examinations. At the classroom level, paper-and-pencil assessment through quizzes, tests and examinations, which assesses memorisation and test-taking skills, dominates. Classroom observation studies suggest that during actual teaching, teachers largely ask closed questions and favour single answers, often known beforehand (UNICEF Tanzania, 2018). Classroom questioning often involves inviting students in turns to give answers until the correct answer that the teacher favours is provided. Teachers either do not provide feedback or provide only

general feedback indicating the gaps in students' knowledge that made them give incorrect answers (Lema & Maro, 2018). Furthermore, paper-and-pencil assessment provides limited useful information for teachers and students to adjust instructional processes in ways that can improve achievement (Kippers et al., 2018). Paper-and-pencil assessment provides scores and grades, which are not particularly useful in guiding instructional improvements.

Since school success is typically judged based on students' performance in high-stake examinations, teachers are often compelled to resort to teaching to the test instead of promoting meaningful learning (O-saki & Njabili, 2003). They train students' techniques for answering examination questions instead of facilitating the development of higher-order skills as stipulated in the curriculum. Often teachers do not teach topics that are not tested in the national examination, or give them only marginal attention (World Bank, 2008). Moreover, the emphasis on grades as a determinant of access to higher education and employment often drives students to strive for higher grades instead of a deeper understanding of school subjects. When classroom cultures reward "gold stars" through grades or ranks, "students often play dirty to score higher grades" (Black & Wiliam, 2005).

Generally, the envisioned transformation in assessment practice through the adoption of assessment techniques that enhance learning achievement remains largely unrealised. Most importantly, the curriculum lacks practical examples showing how assessment reforms can be implemented in classrooms. Moreover, teacher education courses often focus on standardised assessment methods and how to enhance their psychometric properties (Kyaruzi et al., 2019). In this context, where teachers often lack assessment skills, the most logical option for teachers is to rely on traditional assessment approaches, mainly the tools provided by textbooks and instructional material publishers (Lema & Maro, 2018), which often replicate high-stake national examinations. Furthermore, there are relatively few studies on how teachers can integrate AfL into classroom lessons in the Tanzanian context (Kyaruzi et al., 2018, 2019; Lema & Maro, 2018). Thus, there is scant evidence regarding how teachers in resource-constrained classrooms can integrate AfL into their lessons and how AfL contributes to students' learning achievement in such contexts (Kyaruzi et al., 2019). It is therefore imperative that more research focusing on this be conducted.

Two studies conducted by Kyaruzi et al. (2018, 2019) explored teachers' and students' perceptions of FA and how these perceptions predicted self-professed feedback use and student performance. The results suggested that the perceived quality of teacher feedback predicted feedback use and student performance. Moreover, teachers claimed to formatively use assessment information

for self-reflection, improving their approaches, correcting errors and conducting remedial classes to support weaker students. They further reported summative use of assessment information such as ability grouping, accountability reporting and reprimanding low achievers. These findings are limited, however, as no attempts were made to observe whether teachers' favourable perceptions of FA and their avowed use of feedback manifested in actual practice.

Ethnographic studies of teachers' practice in Tanzania suggest that while teachers may verbally commit to innovative pedagogies, their actual classroom practices often contrast with their perceptions (Vavrus, 2009; Vavrus & Bartlett, 2012). Indeed, findings from classroom observations by Lema and Maro (2018) and UNICEF Tanzania (2018) contradict teachers' and students' avowed use of assessment information, as reported by Kyaruzi et al. (2018, 2019). Lema and Maro (2018), for example, observed that teacher feedback constituted exclamatory verbal comments such as "excellent", "very good", "good try" and "that's fair" for students who answered questions correctly, whereas for those who got questions wrong teachers commented "work hard", "lazy" and "poor". Similarly, UNICEF Tanzania (2018) reported that teachers often gave very general feedback to explain why students made mistakes or answered questions incorrectly. Together these studies suggest that teachers lack skills for providing constructive feedback to help students improve their learning.

It was against this background that the present study redesigned biology teachers' lessons, integrating AfL techniques into the teaching and learning process to exemplify how teachers in resource-constrained schools in Tanzania can use AfL in actual lessons (see section 2.2). The aim was to assess the contribution of integrating AfL into the instructional process to students' learning achievement in biology. The question addressed was: What is the contribution of integrating AfL techniques into the teaching and learning process to students' learning achievement?

Method

A quasi-experimental design involving pre-test and post-test of non-equivalent control and experimental groups was used to establish how the integration of AfL into the teaching and learning of secondary school biology enhances students' learning achievement. Non-equivalent control and experimental group design is a form of quasi-experimental design in which the participants cannot be assigned randomly into experimental and control groups simply because the researcher has no control over the randomisation of treatment, unlike in true experimentation (Mitchell & Jolley, 2010). This was the most feasible design for

the school context in which students were organised in intact streams. In such a setting, the random placement of students into control and experimental groups was restricted, as it could have caused learning disruption. Therefore, two intact streams of students, each from a different school, were randomly designated as experimental group ($N = 44$) and control group ($N = 45$) by tossing a coin. The use of existing streams also maximised the ecological validity of the findings.

Research setting

The setting was two boarding secondary schools located in the suburbs of the metropolitan city of Dar Es Salaam, Tanzania. Over the previous five years, both schools had maintained an overall Grade Point Average of 4.6 in the national Certificate of Secondary Education Examination. Thus, the students in the two schools had equivalent academic performance. Furthermore, the schools had similar learning environments because both were located in different parts of the same ward, had relatively similar student populations, and had class sizes of 40–45 students. Both were government schools and thus had similar timetabling, teacher recruitment, remuneration and supply of resources. The matching of the groups based on various characteristics, as well as their random assignment into control and experimental groups, sought to further strengthen the equivalence (Mitchell & Jolley, 2010).

The study involved form one students aged 13–14 years. These students were about to begin learning the topic *Cell Structure and Organisation* (MOEC, 2005). This topic comprises abstract content, which makes it among the most difficult school biology topics for form one students to comprehend (Ozcan et al., 2014). In their study of students' perceptions of difficult biology topics, the researchers found that topics related to the cell, cell division, heredity, DNA and genetic code were among the most difficult to comprehend. The intervention procedures of the current study are described next.

Procedures

Designing lessons

The literature covering the key principles of AfL (Black & Wiliam, 2009; Broadfoot et al., 2002) and exemplary practices in various contexts (Hopfenbeck, 2018; Jonsson et al., 2015; Wylie & Lyon, 2015) was surveyed to identify guidelines for lesson design. Copies of lesson plans from previous years for the topic of Cell Structure and Organisation were then requested from biology teachers at ten schools in the same district. These were analysed to establish

whether they reflected any of the principles and practices of AfL. Moreover, the teachers' lessons other than those covering Cell Structure and Organisation were observed and detailed notes were written to establish whether AfL practices were incorporated in their actual lessons.

Overall, the lesson plans had similar patterns and did not reflect any AfL practices (see Appendix I). Typically, the lessons began with an introduction in which the students reviewed the previous topic. The teacher-directed presentation of new content was interspersed with illustrative visuals and observations, followed by questions-and-answers. The lessons concluded with a summary of key points and instructions for the next lesson. Teachers predominantly asked closed questions requiring single-word or simple affirmative factual answers. Moreover, they mainly gave affirmative feedback using words such as "okay", "correct" or "exactly" to approve students' responses. These observations were consistent with recent research on teachers' assessment practices in Tanzania (Lema & Maro, 2018; UNICEF Tanzania, 2018). After the lesson analysis and observation, the lesson plans were redesigned to incorporate AfL techniques. Verbal questions were added with increased wait-time, rubrics, small project reports, observational checklists, presentations and worksheets in order to broaden the range of assessment formats. Opportunities for the collaborative setting of learning objectives, self- and peer review of work before submission, sharing of assessment criteria in the form of rubrics, and provision of written and verbal feedback were also included (see Appendix II). Assessment tools were constructed, such as worksheets, rubrics and observational checklists, which were used at different stages of the lesson during the intervention (see Appendix III). Finally, two lesson plan formats were established: plans with AfL techniques integrated and the original lesson plans the teachers provided.

The AfL techniques embedded in the redesigned lessons reflected the research-based principles of AfL in various ways. For example, the teachers in the experimental group assisted the students using questions to identify the learning objectives and activities they needed to perform. In addition, they provided assessment rubrics showing different levels of performance when they assigned class work. Such practices reflect the principle of AfL that states that lesson planning should include "strategies to ensure that learners understand the goals they are pursuing and the criteria that will be used to assess their work" (Broadfoot et al., 2002, p. 2). In this case, the collaborative setting of learning objectives was a strategy to help learners understand the learning goals and rubrics were intended to communicate the assessment criteria. The redesigned lesson plans were used with the experimental group and the original lesson plans that the teachers had provided were used with the control group.

Teacher training on the use of AfL

Four biology teachers from the school designated as the experimental group were invited to a week-long workshop on the principles and practice of AfL. In addition to in-depth discussion about AfL, its core principles and exemplary practices, the workshop involved orienting the teachers on how to implement the redesigned lesson plans and the challenges they were likely to face when implementing AfL techniques in their classroom contexts. Finally, the teachers were given copies of the redesigned lesson plans to implement according to their school subject timetables.

Designing the achievement test

Although the purpose of the AfL approach is to enhance authentic learning achievement (William et al., 2004), the students in both the control and experimental groups would eventually sit the National Form Two Examination, which largely tests their knowledge and understanding of biology concepts (Hakielimu, 2012). While AfL may have contributed both to the students' authentic learning and academic performance, the present study aimed to establish its contribution to their academic performance only. An achievement test was therefore constructed and used to measure the students' knowledge and understanding of Cell Structure and Organisation.

The test questions measured all of the learning objectives, covering definitions, characteristics, types and parts of cells, as listed in the syllabus under the topic Cell Structure and Organisation. The test was reviewed for content validity and error reduction by two experienced biology teachers. The necessary amendments were made following the review and the test was piloted in a secondary school comparable to the sampled schools. Immediately after the test, a reflective discussion focusing on the test's item clarity, difficulty and timing was held with ten randomly selected students from the pilot class. The test was then revised to create the final version, which was used as a pre-test and post-test. A typical test item is provided in Figure 1.

Figure 1*Typical Test Item*

6. Carefully study the figure below and answer questions a-c @ 2 marks.

a. This cell is typical of _____ cells

b. What evidence is there to support your answer in 'a' above?

i. _____

ii. _____

iii. _____

c. Name the cell parts labeled A, B, C and F

The intervention

The experimental and control groups were pre-tested using the designed test to assess the prior learning achievement of the students before the topic Cell Structure and Organisation was taught. One teacher who had not participated in the training on AfL then taught the control group using the conventional lesson plan. Meanwhile, one of the four teachers who had participated in the training on AfL taught the experimental group using the redesigned lesson plans. The teachers who taught the control and experimental groups respectively were selected after carefully matching their demographics. They each held a Bachelor of Science with Education degree, had eight years of teaching experience, and were at the same salary level. These teachers had no other commitments apart from teaching and serving as class teachers.

In order to enhance the external validity of the results, the teaching in both groups followed the official syllabus and the school timetables. As per the syllabus, the topic Cell Structure and Organisation is supposed to be taught

over four 80-minute periods (MoEC, 2005). These four periods cover three weeks of instructional time according to the school timetables. With an additional week for pre-testing and post-testing, the intervention lasted one month.

The researcher monitored teaching in the experimental group to verify that the redesigned lessons were implemented as intended. After teaching, the post-test was administered to both the experimental and control groups using the test described above. Pre-test and post-test scores were used to assess the difference in learning achievement between the control and experimental groups.

T-test analysis

The variation in the students' performance from pre-test to post-test in both the control and experimental groups was assessed using a paired sample t-test. Moreover, an independent sample t-test was used to ascertain whether the difference in mean scores between the experimental group and the control group was significant ($p < .05$). The aim was to establish whether the experimental group had higher learning achievement as a result of the treatment. Furthermore, the qualitative data that had been collected during the teachers' professional development and the monitoring of lesson implementation was analysed thematically following the example of Miles, Huberman and Saldana (2014). However, the present paper is based on the quantitative data.

Results

The study redesigned biology teachers' lesson plans to integrate AfL techniques into the teaching and learning process. Furthermore, it assessed the contribution to students' learning achievement of embedding AfL techniques in the instructional process. The results are presented next.

Difference in learning achievement for the control group before and after teaching

A paired sample t-test was used to compare the mean pre-test and post-test scores in order to determine whether the control group had a statistically significant difference in learning achievement before and after teaching using the conventional lesson plans. The results are presented in Table 1.

Table 1*Mean Pre-Test and Post-Test Scores for the Control Group*

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre-Test	14.12	45	4.82	.72
	Post-Test	18.62	45	4.46	.66

The results in Table 1 show that the mean post-test score was higher than the mean pre-test score, with a difference of 4.5. A paired sample t-test was performed to determine the statistical significance of the difference in mean scores between the pre-test and post-test. The results show a statistically significant increase in test scores from pre-test ($M = 14.12$, $SD = 4.82$) to post-test ($M = 18.62$, $SD = 4.46$), $t(44) = -8.18$, $p < .001$. The eta squared statistic (.6) indicated a large effect size. This suggests that the control group achieved some learning when the conventional lesson plans were used.

Difference in learning achievement for the experimental group before and after teaching

A paired sample t-test was used to compare the mean pre-test and post-test scores to determine whether the experimental group had a statistically significant difference in learning achievement before and after teaching. The results are presented in Table 2.

Table 2*Mean Pre-Test and Post-Test Scores for the Experimental Group*

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pre-Test Scores	14.7	44	5.12	.77
	Post-Test Scores	33.18	44	9.21	1.38

The results (see Table 2) show that the mean post-test score was higher than the mean pre-test score with a mean difference of 18.48. A paired sample t-test was performed to determine the statistical significance of the difference in mean scores between the pre-test and post-test. The results show a statistically significant increase in test scores from pre-test ($M = 14.7$, $SD = 5.12$) to post-test ($M = 33.18$, $SD = 9.21$), $t(43) = -14.995$, $p < .001$. The eta squared statistic (.83) indicated a large effect size. This suggests that the experimental group achieved learning with the use of AfL-integrated lessons.

In both the control and experimental groups, there were gains in learning achievement. However, AfL-integrated lessons appear to have had a higher impact ($\eta^2 = .83$) compared to conventional lessons ($\eta^2 = .6$). In order to ascertain whether the difference in learning achievement between the experimental and control groups was statistically significant, an independent t-test was run to compare the mean post-test scores of the two groups. The results are presented next.

Difference in learning achievement between the control and experimental groups

Pre-test results for the experimental and control groups

In order to establish whether the students in both the control and experimental groups had the same level of prior knowledge and understanding of Cell Structure and Organisation, the mean pre-test scores of the two groups were compared. Table 3 shows the mean pre-test scores of the control and experimental groups.

Table 3

Pre-Test Mean Scores for the Experimental and Control Groups

	Group	N	Mean	Std. Deviation	Std. Error Mean
Pre-Test Scores	Experimental Group	44	14.7	5.12	.77
	Control Group	45	14.12	4.81	.72

The results displayed in Table 3 show that the experimental group had a mean score of 14.7 while the control group had a mean score of 14.12, with a mean difference of .57. In order to ascertain whether the mean difference in the pre-test scores between the experimental and control groups was statistically significant, an independent sample t-test was performed. The independent sample t-test for equality of means found no statistically significant difference in the mean scores between the experimental group ($M = 14.7$, $SD = 5.12$) and the control group ($M = 14.12$, $SD = 4.81$), $t(87) = .553$, $p = .582$. The magnitude of the difference in the means (mean difference = .58, 95% CI [-1.51, 2.67]) was very small ($\eta^2 = .003$). The results suggest that prior to the treatment, both the experimental and control groups had the same level of knowledge and understanding of Cell Structure and Organisation. The post-test was administered after teaching the topic using AfL-integrated lessons in the experimental class and conventional approaches in the control group. The post-test results are presented next.

Post-test results of the experimental and control groups

The mean post-test scores of the experimental and control groups were compared to assess the contribution to students' learning achievement of integrating AfL techniques into the teaching and learning process. The post-test results of the experimental and control groups are summarised in Table 4.

Table 4

Post-Test Mean Scores for the Experimental and Control Groups

	Class	N	Mean	Std. Deviation	Std. Error Mean
Post-Test Scores	Experimental Group	44	33.25	9.13	1.37
	Control Group	45	18.62	4.46	.66

The results in Table 4 show that the experimental group, which was taught using AfL-integrated lessons, had a mean score of 33.25, while the control group, which was taught using conventional lessons, had a mean score of 18.62. The mean difference between the two groups was 14.63.

In order to assess whether the mean difference in the post-test scores between the two groups was statistically significant, an independent sample t-test was carried out. The results showed that there was a statistically significant difference in the post-test scores between the experimental group ($M = 33.25$, $SD = 9.13$) and the control group ($M = 18.62$, $SD = 4.46$), $t(62.12) = 9.569$, $p < .001$. The magnitude of the difference in means (mean difference = 14.63, 95% CI [11.57, 17.68]) was very large (eta squared = .51). The experimental group had a higher mean score compared to the control group. This suggests that the experimental group achieved higher learning compared to the control group. Higher learning achievement by the experimental group is likely to have been the result of the intervention, which involved integrating AfL techniques into the teaching and learning process, as discussed next.

Discussion

The most significant finding from this study is the higher learning achievement observed in the experimental group. Previous studies (Wiliam et al., 2004) show that teachers' use of AfL techniques in secondary school science and mathematics leads to increased quality of learning, and subsequently to higher learning achievement. The findings from the present study, which involved biology teachers in resource-constrained schools in the suburbs of Dar es Salaam, confirm those of previous studies. It is likely that embedding AfL

techniques in biology lessons enhanced the learning achievement of the students in the experimental group in various ways.

First, asking open-ended, thought-provoking questions such as those indicated in the lesson plan (see Appendix II) is likely to have enhanced the students' mental engagement through classroom interactions and dialogues. Classroom interactions provide context for students to comment on each other's work, which makes them feel positive about their learning (Webb & Jones, 2009). In Tanzania, teachers often ask closed, factual questions with very brief wait-times (Kira et al., 2013). When no students volunteer to answer or when none answer as expected, teachers either seek answers from bright students or provide the correct answers themselves. This often limits classroom interactions to routinised, factual questions-and-answers with limited learning value (Hardman et al., 2012). The teachers in the experimental group allowed relatively more time for the students to think and generate well-thought-out ideas. In this way, these teachers demonstrated that they valued elaborate, well-thought-out contributions, as opposed to the short affirmative responses that characterise classroom questioning in Tanzania (Kira et al., 2013).

Second, although the lessons in both the control and experimental groups began with activities aimed at eliciting the students' prior knowledge of the topic, the teachers in the experimental group explicitly used the evidence of prior learning to plan the next learning steps (see Appendix II). In the control group, the teachers mostly adhered to the rigid lesson plans, regardless of the students' learning needs. Unlike those in the control group, the teachers in the experimental group not only shared the lesson objectives as indicated in the syllabus, but also collaborated with the students to adapt the lesson objectives in light of the students' prior knowledge of and experience with the topic. This collaborative setting of learning objectives enabled the students to understand what they were supposed to learn and to self-assess their progress accordingly (William et al., 2004). In this way, the teachers in the experimental group best served the students' learning needs. When students are involved in setting learning objectives, they adopt relevant strategies to learn and improve their achievement in spelling and punctuation (Black & William, 2018).

Third, while the teachers in the control group were mainly concerned with the correctness of the students' responses and taught to help the students produce correct answers known beforehand, the teachers in the experimental group asked questions to encourage thinking, and thus their students produced more thoughtful answers. The teachers in the experimental group were concerned with what they could learn from the students' answers and how they could provide feedback to help the students adjust their learning pathways. To

this end, the teachers in the experimental group provided constructive feedback to enhance the students' confidence, optimism and determination. Such feedback specified the learning outcomes that the students had or had not achieved and the learning pathways they needed to follow. The quality of interactive feedback is a critical feature in determining the quality of the learning activity (Black & Wiliam, 2006).

Fourth, by engaging the students in self- and peer assessment of their work, the teachers motivated them to improve the standard of their work. Peer assessment provides opportunities for students to serve as instructional resources for one another (Black & Wiliam, 2006). The students in the experimental group were receptive to the comments made by their peers, probably because the comments were in a language they could relate to. By building on their peers' comments, the students in the experimental group were able to adjust their learning beyond what they would have done if they had not engaged in self- and peer assessment. Consequently, the students in the experimental group seemed to believe more strongly in their own learning success (Black & Wiliam, 2009).

Lastly, by sharing assessment criteria, the teachers made the students aware of the achievement benchmarks from the start of the topic. Therefore, both teachers and students could monitor learning progress based on the shared assessment benchmarks and lesson objectives. They planned the next learning pathways and managed their learning advancement. When learners participate in setting success criteria, they are able to monitor their thinking, performance and understanding (Davies, 2003). In other words, they use the assessment criteria to monitor their learning.

Conclusion

The present study set out to redesign biology teachers' lessons to integrate AfL techniques into the teaching and learning process. It further assessed the effect of integrating AfL techniques on students' learning achievement in form one biology. Independent sample t-tests revealed that the form one students in the experimental group exhibited higher performance than those in the control group on a test measuring understanding of the topic Cell Structure and Organisation. This suggests that the students in the experimental group, which was taught using AfL-integrated lessons, achieved higher learning compared to those in the control group, which was taught using conventional approaches. Overall, these results strengthen the idea that the integration of AfL into teaching and learning enhances students' learning achievement (Ellegaard

et al., 2018; Wiliam et al., 2004). The present study contributes to our understanding of how teachers in resource-constrained classrooms such as those in sub-Saharan Africa can integrate AfL techniques into their day-to-day lessons, thereby harnessing the power of assessment to enhance learning and raise standards.

The findings of the study are, however, limited in some important ways. The most important limitation lies in the fact that the sample was small, which limits the generalisability of the findings. Furthermore, the training itself may have motivated the teachers in the experimental group to provide better and novel instruction. Consequently, the students in the experimental group may have benefited from such novelty (Mertens, 2010). Lastly, although efforts were made to match the control group and the experimental group along several key variables, including age, learning environment, teacher demographic, learning achievement, etc., the two groups were not equivalent. This is because the groups were not randomly assigned (Mitchell & Jolley, 2010). If interpreted cautiously, however, the findings may still prove useful in supporting the conclusions.

Future research could assess how students benefit from different AfL techniques and whether each of the techniques contributes equally to students' learning achievements. For example, research comparing the contribution of constructive feedback with the contribution of peer-assessment is needed. Furthermore, in resource-constrained classroom contexts, a follow-up study assessing the sustained use of AfL techniques by teachers in the intervention group is imperative. Such a study is needed in order to establish whether or not teachers continue to use AfL techniques after participating in continuous professional development aimed at improving their assessment practices.

References

- Berry, R. (2008). *Assessment for learning*. Hong Kong University Press.
- Black, P., & Wiliam, D. (2005). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.
- Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81–100). Sage.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575. <https://doi.org/10.1080/0969594X.2018.1441807>

- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002). *Assessment for learning: 10 principles*. University of Cambridge, School of Education.
- Davies, A. (2003). Learning through assessment: Assessment for learning in science classroom. In J. M. Atkin & J. E. Coffey (Eds.), *Science educators essay collection: Everyday assessment in science classroom* (pp. 13–26). NSTA Press.
- Ellegaard, M., Damsgaard, L., Bruun, J., & Johannsen, B. F. (2018). Patterns in the form of formative feedback and student response. *Assessment & Evaluation in Higher Education*, 43(5), 727–744. <https://doi.org/10.1080/02602938.2017.1403564>
- Greenstein, L. (2010). *What teachers really need to know about formative assessment*. ASCD.
- Hakielimu. (2012). *School children and national examinations: Who fails who?: A research report on the relationship between examination practice and curriculum objectives in Tanzania*. Hakielimu. <https://learningportal.iiep.unesco.org/en/library/school-children-and-national-examinations-who-fails-who-a-research-report-on-the>
- Hardman, F., Abd-Kadir, J., & Tibuhinda, A. (2012). Reforming teacher education in Tanzania. *International Journal of Educational Development*, 32(6), 826–834. <https://doi.org/10.1016/j.ijedudev.2012.01.002>
- Hopfenbeck, T. N. (2018). Classroom assessment, pedagogy and learning – twenty years after Black and Wiliam 1998. *Assessment in Education: Principles, Policy & Practice*, 25(6), 545–550. <https://doi.org/10.1080/0969594X.2018.1553695>
- Hopfenbeck, T. N. (2017). Balancing the challenges of high-stakes testing, accountability and students' well-being. *Assessment in Education: Principles, Policy & Practice*, 24(1), 1–3.
- Hopfenbeck, T. N., & Stobart, G. (2015). Large-scale implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 22(1), 1–2. <https://doi.org/10.1080/0969594X.2014.1001566>
- Jonsson, A., Lundahl, C., & Holmgren, A. (2015). Evaluating a large-scale implementation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy & Practice*, 22(1), 104–121. <https://doi.org/10.1080/0969594X.2014.970612>
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75, 199–213. <https://doi.org/10.1016/j.tate.2018.06.015>
- Kira, E., Komba, S., Kafanabo, E., & Tilya, F. (2013). Teachers' questioning techniques in advanced level chemistry lessons: A Tanzanian perspective. *Australian Journal of Teacher Education*, 38(12), 66–79.
- Kitta, S., & Tilya, F. (2010). The status of learner-centred learning and assessment in Tanzania in the context of the competence-based curriculum. *Papers in Education and Development*, 29, 77–91.
- Kyaruzi, F., Strijbos, J.-W., Ufer, S., & Brown, G. T. L. (2018). Teacher AfL perceptions and feedback practices in mathematics education among secondary schools in Tanzania. *Studies in Educational Evaluation*, 59, 1–9. <https://doi.org/10.1016/j.stueduc.2018.01.004>
- Kyaruzi, F., Strijbos, J.-W., Ufer, S., & Brown, G. T. L. (2019). Students' formative assessment

- perceptions, feedback use and mathematics performance in secondary schools in Tanzania. *Assessment in Education: Principles, Policy & Practice*, 26(3), 278–302.
<https://doi.org/10.1080/0969594X.2019.1593103>
- Lema, G., & Maro, W. (2018). Secondary school teachers' utilization of feedback in the teaching and learning of mathematics in Tanzania. *Papers in Education and Development*, 36, 162–184.
- Mertens, D. M. (2010). *Research and evaluation in education and psychology*. SAGE Publications.
- Ministry of Education and Culture. (2005). *Biology syllabus for secondary schools*. Ministry of Education and Culture.
- Ministry of Education and Vocational Training. (2007). *Syllabus for biology subject I-IV*. Ministry of Education and Vocational Training.
- Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained*. WADSWORTH.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE.
- O-saki, K., & Njabili, A. (2003). *Secondary education sector analysis*. Ministry of Education and Culture and the World Bank.
- Ozcan, T., Ozgur, S., Kat, A., & Elgun, S. (2014). Identifying and comparing the degree of difficulties biology subjects by adjusting its reasons in elementary and secondary education. *Procedia - Social and Behavioral Sciences*, 116, 113–122. <https://doi.org/10.1016/j.sbspro.2014.01.177>
- Steiner-Khamsi, G. (Ed.). (2004). *The global politics of educational borrowing and lending*. Teachers College, Columbia University.
- UNICEF Tanzania (Ed.). (2012). *Cities and children: The challenge of urbanisation in Tanzania*. UNICEF Tanzania.
- Vavrus, F. (2009). The cultural politics of constructivist pedagogies: Teacher education reform in the United Republic of Tanzania. *International Journal of Educational Development*, 29(3), 303–311.
<https://doi.org/10.1016/j.ijedudev.2008.05.002>
- Vavrus, F., & Bartlett, L. (2012). Comparative pedagogies and epistemological diversity: Social and materials contexts of teaching in Tanzania. *Comparative Education Review*, 56(4), 634–658.
<https://doi.org/10.1086/667395>
- Webb, M., & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 16(2), 165–184. <https://doi.org/10.1080/09695940903075925>
- Wiliam, D. (2004, June). Keeping learning on track: Integrating assessment with instruction [Invited address]. 30th Annual Conference of the International Association for Educational Assessment, Philadelphia, PA.
https://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/IAEA%2004%20paper.pdf
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49–65.
<https://doi.org/10.1080/0969594042000208994>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14.
<https://doi.org/10.1016/j.stueduc.2011.03.001>

World Bank. (2008). *Curricula, examinations and assessment in secondary education in sub-Saharan Africa* (World Bank Paper no. 128). World Bank.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, 22(1), 140–160.

<https://doi.org/10.1080/0969594X.2014.990416>

Biographical note

ALBERT TARMO is a lecturer in the field of science education at the School of Education, University of Dar es Salaam, Tanzania. His main areas of research include educational assessment, biology education, science teacher education, inquiry-based science teaching, and learner-centred pedagogy.

Appendix I: Lesson Plan for Control Class

Preliminary Information

Subject	Date	Stream	Period	Time	Number of students	
					Registered	Present

Main Topic: Cell Structure and Organisation

Sub Topic: The concept of a cell

General Objective: Students should understand the concept of a cell

Specific objectives: By the end of 80 minutes, each student should be able to:

- explain the meaning of a cell correctly;
- mention at least four characteristics of a cell correctly;
- differentiate various types of cells.

Resources: Charts showing various types of cells, biology text book.

Lesson Development

Stage	Teaching Activities	Learning Activities	Assessment Procedures
Introduction (5 min.)	Introducing a new lesson. Asking questions about the meaning of cell. <ul style="list-style-type: none"> • What is a cell? 	Listening. Answering questions.	Verbal questions.
Presentation (50 min.)	Describing the concept of a cell. Guiding students to observe charts showing different types of cells. Asking students to identify and write down the characteristics of various types of cells. Writing notes on the chalkboard.	Listening. Observing charts showing various types of cells. Identifying and writing down the characteristics of various types of cells. Taking notes in exercise books.	Verbal questions.
Reinforcement (10 min.)	Provide a reading activity for students to differentiate various types of cells.	Reading biology textbooks in groups to differentiate various types of cells.	
Reflection (10 min.)	Guiding students to discuss the use of the knowledge learned in their daily life.	Discussing the use of the new knowledge in their daily life.	Verbal questions.
Conclusion (5 min.)	Guiding students to summarise the lesson learned.	Summarising the lesson learned.	

Appendix II: Lesson Plan for Intervention Class

Preliminary Information

Subject	Date	Stream	Period	Time	Number of students	
					Registered	Present

Main Topic: Cell Structure and Organisation

Sub Topic: The concept of a cell

General Objective: Students should understand the concept of a cell

Specific objectives: By the end of 80 minutes, each student should be able to:

- explain the meaning of a cell correctly;
- mention at least four characteristics of a cell correctly;
- differentiate various types of cells.

Resources: Charts showing various types of cells, biology text book.

Lesson Development

Stage	Teaching activities	Learning Activities	Assessment Procedures	Feedback
Introduction (15 min.)	<p>Asking open-ended related questions on the concept of a cell:</p> <ul style="list-style-type: none"> • What are living things like plants made of? • What are these parts, e.g., a leaf, made of? • If you tear up a leaf blade into the smallest units/parts, what will you end up with? • If you divide a living organism into the smallest units, what will you end up with? <p>Guide students to formulate objectives and activities to pursue based on the last two questions.</p>	<p>Brainstorming and answering questions asked in order to elicit their prior thoughts.</p> <p>Suggest objectives and activities to pursue.</p>	<p>Verbal questions (3+ minutes wait time). Verbal responses.</p>	<p>Constructive verbal feedback; teacher scaffolds, follow-up questions.</p>
Presentation (30 min.)	<p>Leading students in groups of five to observe charts showing various types of cells.</p> <p>Provide rubrics to guide peer assessment.</p> <p>Commenting on the group work and peer comments.</p> <p>Clarifying any misconceptions and queries arising from group activity.</p>	<p>Observing charts showing various types of cells. Identifying characteristics of various cells. Writing down meaning and characteristics of cells.</p> <p>Exchanging work between groups. Commenting on peer work based on the rubric provided.</p>	<p>Assessment by peers on the exchanged work focusing on weaknesses, strengths and points for further improvement.</p>	<p>Written comments by peers. Verbal comments by teacher.</p>

Stage	Teaching activities	Learning Activities	Assessment Procedures	Feedback
Reinforcement (15 min.)	Provide reading activity for students to differentiate various types of cells. Providing worksheets.	Reading biology textbooks in groups of five to differentiate various types of cells. Attempt questions on worksheets. Exchange worksheets between groups for peer comments.	Assessment of worksheets by peers identifying weaknesses, strengths and points for further improvement.	Written comments by peers. Verbal comments by teacher.
Reflection (10 min.)	Leading plenary discussion on the application of knowledge of cells in daily life, e.g., in sickle cell screening.	Discuss in plenary the application of knowledge of cells in daily life	Verbal questions (3+ minutes wait time).	Verbal comments by teacher on individual responses.
Conclusion (10 min.)	Guiding students to revisit the objectives set and summarise the major concepts learned.	Revisiting objectives set. Summarising major concepts learned.	Verbal questions. Self-assessment to determine what they have learnt.	

Appendix III

A. Rubric on observing charts with various cell types

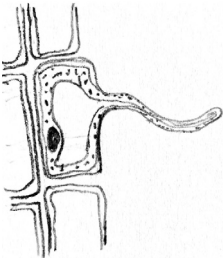
Task	Good	Better	Best
Defining cell	<i>Defined a cell fairly correctly but missed the most relevant keywords, thus meaning conveyed is vague.</i>	<i>Defined a cell correctly using some relevant keywords, thus the meaning conveyed is fairly clear.</i>	<i>Defined a cell correctly using relevant keywords, thus conveying a clear meaning.</i>
Identifying the characteristics of cells	<i>Identified at least four characteristics without explaining or giving examples.</i>	<i>Identified at least four characteristics giving a fairly clear explanation without examples.</i>	<i>Identified at least four characteristics giving a clear explanation and examples.</i>

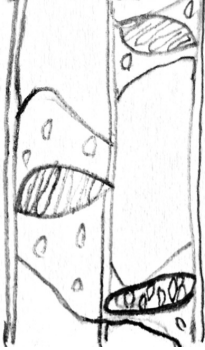
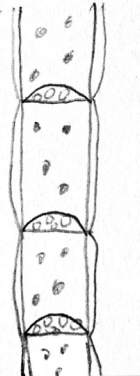
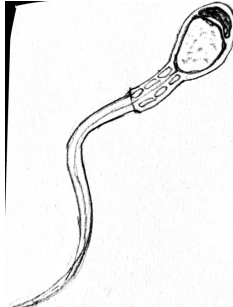
B. Worksheet

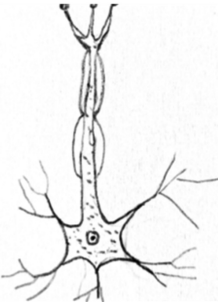
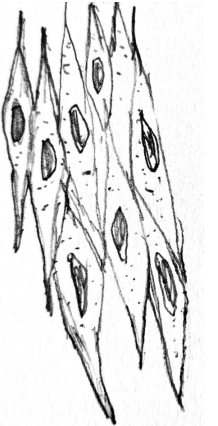
Instructions:

- Carefully read question I–III and fill in the blanks as required.
- Once you complete your work, give it to your neighbouring group, who will also give you their work.
- Once you receive the work of your peers, read it carefully and provide constructive comments highlighting both the strengths and weaknesses, and suggesting what they might do to further improve their work.

Question I

Cell type	Description
A. Root hair 	1. Name _____ 2. Location _____ 3. Functions _____ _____ _____ 4. Adaptive features _____ _____ _____

<p>B. Xylem</p> 	<ol style="list-style-type: none"> 1. Name _____ 2. Location _____ 3. Functions _____ _____ _____ 4. Adaptive features _____ _____ _____
<p>C. Phloem</p> 	<ol style="list-style-type: none"> 1. Name _____ 2. Location _____ 3. Functions _____ _____ _____ 4. Adaptive features _____ _____ _____
<p>D. Sperm cell</p> 	<ol style="list-style-type: none"> 1. Name _____ 2. Location _____ 3. Functions _____ _____ _____ 4. Adaptive features _____ _____ _____

<p>E. Nerve cell</p> 	<p>1. Name _____</p> <p>2. Location _____</p> <p>3. Functions _____ _____ _____</p> <p>4. Adaptive features _____ _____ _____</p>
<p>F. Muscle cell</p> 	<p>1. Name _____</p> <p>2. Location _____</p> <p>3. Functions _____ _____ _____</p> <p>4. Adaptive features _____ _____ _____</p>

Question II

Cells “A” are found in the specialised region of _____ called _____. The cells in this region produce a _____ that helps them burrow into the soil more easily.

Question III

Special types of proteins called _____ and _____ make up cells F. Due to the presence of such proteins, _____ can slide past each other.