International Journal of Educational Excellence
(2021) Vol. 7, No. 1, 79-97
ISSN 2373-5929
DOI: 10.18562/IJEE.064

Collaborative Testing on Students' Achievement and Retention of Chemistry Concepts: An Exploratory Study in the Field of Science Education for Sustainable Development

Bimbola Dupe Oludipe [a], Adewale Owodunni Saka [a],
& Babatunde Opeyemi Awolo [a]

Received: 24 March 2021 • Accepted: 7 May 2021

Abstract: Science education is a sine-qua-non for sustainable development, but students' performance in the science subjects is worrisome. Meanwhile, the collaborative testing strategy has been reported as effective in learning certain subjects but not chemistry. Hence, this study investigated the effects of collaborative testing strategy on students' achievement in senior school chemistry with gender as a moderating variable. The study adopted a pretest-posttest control group quasi-experimental research design. The study targeted all the students offering chemistry in the twenty (20) public senior secondary schools in Alimosho Local Government Area of Lagos State (Nigeria). Sixty (60) students (control=30, experimental=30) selected from two purposively selected schools participated in the study. The research instrument was the Chemistry Achievement Test (r= 0.75). Data collected were analyzed using inferential statistics. The results showed that the collaborative testing strategy significantly improved students' achievement and reduced the gender achievement gap in senior school chemistry but did not significantly improve the retention of learned concepts. Finally, the study recommended that the teachers adopt the collaborative testing strategy for learning Chemistry.

Key Words: achievement; collaborative testing; retention; secondary chemistry; sustainable development.

---

[a] Olabisi Onabanjo University (Nigeria). Correspondence: Adewale Owodunni Saka, Science and Technology Education Department (STED). Faculty of Education, Olabisi Onabanjo University, Ago-Iwoye, Ogun State, Nigeria, wale.saka@oouagoiwoye.edu.ng ORCID 0000-0003-2417-4956. Coauthor: Bimbola Dupe Oludipe ORCID 0000-0002-2692-4055, bimbola.oludipe@oouagoiwoye.edu.ng. Coauthor: Babatunde Opeyemi Awolo ORCID 0000-0002-3055-7075, opeyemit642@gmail.com

1. Introduction

Every teacher desires to improve the quality of learning of the students. Quality learning requires the teacher to adopt effective teaching methods. Meanwhile, the traditional teaching methods have been criticized for lacking capability to provide the kind of support needed by the learners to function in today's competitive world.

Science and Technology Education (ST&E) policy through the Federal Ministry of Education (FME, 2018) stated that teaching should enable the students to think and clarify their values, understand and appreciate the value of thinking to make them resourceful and well-positioned to compete favorably among their peers in the global world of work. This means that the method of teaching that will improve learning quality should put the students at the center of teaching-learning activities instead of the teacher-do-all approach inherent in the traditional methods of teaching.

Quite often, high-stake tests measure learning quality, which usually cause test anxiety. Test anxiety can manifest in the learners as worry, thinking disorder, tension, bodily symptoms and inferiority complex (Thomas, 2021). Test anxiety has been found to negatively impact students' learning (Galal, Vyas, Hackett, Rogan, & Nguyen, 2021; Pate et al., 2021; Thomas, 2021).

Studies have shown that testing, which is traditionally a means of assessing students' cognitive development, can also be used as a teaching strategy to engage the learners and reduce their anxiety actively. Efu (2018) observed that examinations are fast becoming learning tools contrary to the general belief that they gauge individual students' learning. An example of a testing method reported in the literature to serve as a learning tool and reduce test anxiety is collaborative testing.

Collaborative testing is an umbrella term covering any test method where students work together to share knowledge and skills to complete an examination (Phillips, Munn, & George, 2019). It is team-based testing where learners discuss a test question extensively and reach an agreement before answering (Vogler & Robinson, 2016). The strategy is based on the constructivism learning belief that social interaction among learners leads to an exchange of ideas, knowledge and skills capable of moving the learners from their zones of proximal development to an independent level of learning. Collaborative testing is a form of active learning strategy that can improve learners' academic prowess (Helms, Keith, & Walker, 2019) and reduce their test anxieties (Breane-Bovee, 2016; Eastridge & Benson, 2020). Kapitanoff and Pandey (2019) noted that the quality of engagement among the learners contributes to the effectiveness of collaborative learning.

There have been groundswell studies on the effect of collaborative testing strategy on students' learning. Eastwood, Kleinberg, and Rodenbaugh (2020) examined the effects of collaborative testing strategy on preclinical medical students' long-term retention of basic knowledge. The mixed-methods study also collected information on students' perceptions of the strategy. The results disclosed that students engaged in collaborative testing performed better than their counterparts exposed to the individual testing. The students exposed to collaborative testing also retained the knowledge of the test topics better than those involved in individual testing. The qualitative aspect of the research confirmed that collaborative testing improved knowledge retention due to repeated opportunities for knowledge retrieval during the test. Additional results showed that some students were displeased with the interpersonal conflicts that usually occurred during collaboration. This means that mechanism for conflict resolution is needed for collaborative testing to yield desired benefits to the learners and teachers.

Mahoney and Harris-Reeves (2019) also explored the benefits of collaborative testing on students' overall performance and performance in higher-order thinking questions. The results revealed that low ability and average ability students performed better in the collaborative test than in the traditional (individual test) where no improvement was observed in the results of the upper performing students. This means that the high-ability students groups help the other categories of learners to learn the contents which is the essence of collaborative learning. This effort bridged the knowledge gaps which existed among the learners before the commencement of the collaboration. The improvement of high ability learners manifests in the form of self-reflection, no matter the group composition (Conejo, Barros, Guzman, & Garcia-Vinas, 2013).

In a related effort, Eastridge and Benson (2020) determined the effect of collaborative testing on students learning and retention using a mixed-method research approach. The study revealed that students that engaged in the group testing outperformed their counterparts in the control group, which engaged in individual testing. The outcomes further disclosed that the collaborative testing approach reduced students' test anxiety. Similarly, Breane-Bovee (2016) examined the impact of collaborative testing on students' test anxiety at professional health institutions. The results showed that students that took tests collaboratively demonstrated a reduction in test anxiety compared to students that took the test individually. Another interesting finding was the students' improvement in test scores, confidence, critical thinking and satisfaction. Vazquez-Garcia (2018) also examined the relationship between second-year medical students' group performance and individual performance in a collaborative learning environment. The result showed that the mean performance scores of students in individual testing and collaborative testing differed significantly in support of the collaborative testing group. The finding

also indicated that collaborative testing improved knowledge retention. The finding is in accord with that of Kleinberg, Eastwood, and Rodenbough (2018) that collaborative testing promoted students' long-term knowledge retention.

However, the research of Caboral-Stevens and Fox (2020), which compared the assessment scores of Baccalaureate nursing students exposed to collaborative testing to those exposed to the individual testing method, reported the mean retention score of students in the collaborative testing group was poor. This may be due to the period of eight months that elapsed before the retention test was conducted. The students tend to have forgotten most of the contents learned especially if they do not put them to use frequently. Despite short-term retention, students expressed satisfaction with the use of collaborative testing as they gained social skills, self-confidence and learning accountability. The study of Leight, Saunders, Calkins and Withers (2012) similarly found that collaborative testing significantly improved performance but not content retention and concluded that there was no improvement in retention when students were tested using the collaborative testing method.

Studies on the effects of collaborative learning/testing on student's performance by gender are also abound in the literature. Takeda and Homberg (2014) examined the effect of gender on group work process and performance using the results of self and peer assessment. The results showed that students in gender-balanced groups (a group with an equal number of males and females) had improved learning performance. The strong determination of group members to learn propelled them to collaborate productively, which manifested in improved performance. The researchers maintained that collaborative efforts would not result in learning improvement if the group composition is defective. For instance, the study reported underperformance by all-male groups. It also indicated that a male student in a male gender exception groups (a group with one male with other members as females) exhibited negative behaviors during collaboration.

The findings from the study of Kapitanoff and Pandey (2019) also revealed that female students who interacted well during mid-term collaborative testing obtained better grades than in the individual tests conducted earlier in the term. The improved performance in the collaborative test also extended to the final examinations. This was attributed to the quality and quantity of the interaction enjoyed during the collaborative testing. The extension of knowledge acquired during midterm to the final examination suggests the effectiveness of collaborative testing in enhancing knowledge retention, contrary to the findings of researchers such as Caboral-Stevens and Fox (2020). Kapitanoff and Pandey noted that unproductive discussion during collaborative testing may hinder students' knowledge retention.

Another study by Cheng (2019) reported that female students enjoy learning in the same balanced gender-mixed groups while male students can benefit from any kind of gender-mix composition. It suggested using the same-gender composition or balanced mixed-group activities to help students learn effectively. Cheng further argued that individuals do not benefit from collaboration in the same proportion since the level of achievement in collaborative efforts depends on diverse factors such as gender difference and the nature of the tasks. This implies that any teacher interested in collaborative testing strategy should be conversant with group dynamics. However, another study by Acharya, Acharya, and Shrestha (2020) indicated no gender difference in the way students collaborate to learn. This also suggests that group composition will not affect interaction among the learners during group learning.

Scientific development is a panacea to sustainable development and can be facilitated with quality science education. Science and technology education is to inculcate the skills of creativity in the citizens and aid them in participating productively in the world shaped by knowledge and innovation (FME, 2018). It is also to equip citizens with skills needed to understand the interaction between science and society and its contribution to cultural heritage. Further, science and technology education is to develop in the learners the skills required for a sustainable society and economic development of Nigeria.

Meanwhile, Nigeria lacks technical know-how aside from the persistent low performance of students in examinations conducted by bodies such as West African Examinations Council (WAEC), National Examination Council (NECO), and National Business and Technical Examinations Board (NABTEB) and others. The pass rate of students in examinations conducted by these bodies in science, technology and mathematics (STM) subjects has been consistently less than 50%. Another worrisome indicator is the ranking of 124 in the Global Competitive Index (GCI) among 140 countries and subsequent ranking as 132nd based on the quality of science education (FME, 2018).

Although the three core science subjects at the secondary education level are Physics, Chemistry, and Biology, this study focuses on chemistry due to its importance in attaining the 2030 sustainable development agenda that Nigeria is a signatory. Juntunen (2016) maintained that chemistry could contribute immensely to the actualization of the agenda for sustainable development. Chemistry is important for the growth of human capacity and to increase the skilled workforce for modernization (Lerman, 2014). Shiqiqi, Permanasari, and Hermani (2020) also argued that learners need knowledge of chemistry to solve existing and future problems. Solving present and future problems is very important to develop society sustainably. Chemistry is also crucial for the socio-economic development of a nation (Elschami &

Kumemer, 2020), and the application of chemistry guarantees a healthy environment, one of the pivotal agendas for sustainable development.

The research reports on students' performance in chemistry in Nigeria align with the ST & E policy statement that students' performance in science, technology, and mathematics subjects is below expectation. For instance, findings from Nbinna (2012) and Ojukwu (2016) studies revealed that students' performance in chemistry at the secondary school level was poor. Likewise, the West Africa regional secondary education examinations body, West African Examinations Council (WAEC) Chief Examiners' Reports (2014-2018) on students' performance in chemistry in Nigeria also revealed fluctuation in the results. The students' mean performance was 25 with a standard deviation of 9.06 in 2014; 27 with a standard deviation of 8.83 in 2015; 25 with a standard deviation of 7.81 in 2016; 26 with a standard deviation of 8.37 in 2017; and 24 with a standard deviation of 9.95 in 2018. The fluctuations in the mean performances of students suggest that attainment of sustainable development and other national development using chemistry may be like chasing a shadow and the country may be left behind which is against the declaration in the Agenda that no country would be left behind (Rosa, 2017).

Similarly, studies on the gender gap in chemistry learning have reported conflicting findings. For example, researchers such as Popoola and Solomon (2017); Chabari et al. (2018) and Ekpen (2020) found no gender difference in chemistry learning among senior secondary school students, whereas Ejedegbe (2016) and Okeke (2018), who sought to determine the effect of gender on mean achievement scores of chemistry students found that gender was a significant factor in students' achievement in chemistry. This means that gender still counts in students' learning of chemistry. The suggests that further research on reducing the learning gap between male and female students offering chemistry in secondary school is necessary. The individuals, regardless of their gender, contribute to the country's scientific development.

The poor state of science and education has been linked to factors such as quality of teaching and learning. The ST&E policy maintains that traditional methods are responsible for students' declining interest and consequent low enrolment and poor performance in science and technology education subjects and programs. It also recommends that the teaching of science and technology of the 21st-century should be student-centred. Interestingly, previous studies and reports have shown that collaborative testing focuses on the learners rather than the teachers and can be used as a learning strategy. It has also been found to improve learning and retention, mostly in higher institutions. This study, therefore, examined the effectiveness of collaborative testing strategy on students' learning of chemistry in secondary schools. It also determined the moderating effect of gender on students' achievement and retention of knowledge chemistry concepts.

2. Objetives and Hypotheses

2.1. Objectives of the Study

The study determined the effect of collaborative testing on students' achievement and retention of learned Chemistry concepts. Specifically, the study:

i. examine students' achievement and retention in chemistry after exposure to collaborative testing.

ii. Investigated gender difference in achievement and retention in chemistry after exposure to collaborative testing.

2.2 Hypotheses

The following null hypotheses tested at the 0.05 level of significance guided this study:

$H_0$1. There is no significant main effect of testing strategy (collaborative and individual) on students' achievement in senior secondary school chemistry.

$H_0$2. There is no significant main effect of testing strategy on students' retention of learned chemistry concepts.

$H_0$3. There is no significant main effect of gender on students' achievement in senior secondary school chemistry.

$H_0$4. There is no significant main effect of gender on students' retention of learned senior chemistry concepts.

$H_0$5. There is no significant interaction effect of testing strategy and gender on students' achievement in senior secondary school chemistry.

$H_0$6. There is no significant interaction effect of testing strategy and gender on students' retention of learned chemistry concepts.

3. Methodology

The study adopted a pretest-posttest control group quasi-experimental research design involving a 2x2 factorial. This design allows the researchers to use the participants without randomly assigning them to groups (Maciejewski, 2020). It also enables the researchers to cross the testing strategy with gender as a moderating variable. So, the participants were not randomized to the groups but used as found in their entire classes. The testing strategy and gender were the independent and moderating variables, respectively, whereas students' achievement in chemistry and retention of learned chemistry concepts were the dependent variables.

3.1. Target Population

The target population for this study was all the senior secondary 2 (SS 2) students offering chemistry in the twenty (20) public senior secondary schools in Alimosho Local Government Area of Lagos State. The SS 2 students were selected because they were not preparing for any external examinations. Such examinations come up in senior secondary school 3 (SS 3).

3.2. Sample and Sampling Technique

The sample comprised sixty (60) students offering chemistry in the two public senior secondary schools purposely selected for the study on the criteria that they had qualified chemistry teachers who were willing to participate in the study. The two schools were also not close to each other to avoid interaction among the students from both schools. However, the schools were randomly assigned to experimental and control groups. Coincidentally, both the experimental and control groups had 30 students each.

3.3. Instrumentation

3.3.1. Chemistry Achievement Test (CAT)

A researcher-developed Chemistry Achievement Test (CAT) was the only instrument used for data collection. The instrument had sections A and B. Section A elicited demographic information of the respondents, whereas section B consisted of twenty multiple-choice items covering topics such as oxidation, reduction and organic chemistry. All the topics were part of Nigeria's senior secondary II Chemistry curriculum.

Initially, the CAT had eighty items extracted from past WAEC and NECO questions. It was administered to eighty (80) students offering chemistry from schools that did not participate in the study but had comparable academic standards with the participating schools. The item analysis indicated that only forty (40) items could be used. The test copies with forty items were given to two experts in test construction and two secondary school chemistry teachers for content and face validities. The critiques and suggestions of the experts and teachers were used to refine the items in the test. Twenty items made the final CAT from the pool of the forty items.

The CAT was administered on twenty (20) SS 2 students offering chemistry in a school different from the targeted sample but had a comparative standard. The split-half reliability method yielded a coefficient of 0.75.

3.4. Method of Data Collection

The researchers approached the Heads of participating schools for permission to conduct the study ethically. The chemistry teachers were

briefed about the purpose of the research and agreed unconditionally to participate. They served as research assistants throughout the eight weeks of the experiment. The chemistry teachers in both schools were allowed to teach the topics penciled down for the study for four weeks using their usual conventional methods. The researchers did not interface with the students in both groups until after the teachers had taught all the topics for the study.

In the fifth week, the researcher sought the permission of the students to participate in the study. They agreed to participate after assuring them that their information would be kept confidential and that results from the tests would only be used for research purposes alone. The students in the experimental group were then trained on the guidelines of collaborative testing such as the need to discuss in the group all the questions before answering them individually; every student would submit a worksheet to attract individual mark; all students have the rights to contribute to the discussions in the group; nobody has control over the other and; any conflict should be reported to the researchers or the assistants. After that, the pretests were conducted in both groups through individual testing.

In the sixth week, the experimental group conducted the collaborative test. The test involved four students in balanced-mixed-gender groups (equal number of males and females). Each learner was allowed to pick partners from a different group to ensure knowledge gaps among the learners about the contents to learn. The grouping was done using students' results in their last terminal examination without their knowledge. This was to ensure that the groups were made up of mixed-ability students. Collaboration is only effective when there are knowledge gaps to fill by group members (Retnowati, Ayres, & Sweller, 2018). The test was conducted in a more significant class to reduce noise level and facilitate face-to-face interaction among group members as suggested by experts (Millis and Cottell cited in Slusser & Erickson, 2006).

During the test, the students were encouraged and allowed to discuss the test questions with members of their groups before they answered them. They shared ideas, argued about conflicting answers. However, they were not allowed to search for the answers from their notebooks, and textbooks. The discussions were also limited to within the groups. The researchers and the assistants moved around the class to monitor students' interactions. Most times, they moderated the students' discussion to avoid conflicts. Every student was mandated to submit separate answer sheet despite the collaboration. However, the control group still maintained their testing approach. Meanwhile, the questions were reshuffled to prevent the threat of examination wisdom on students in both groups.

In the eighth week, individual tests were conducted in both experimental and control groups. The scores obtained by the students in the test formed the retention scores.

3.5. Method of Data Analysis

The data collected were analysed using inferential statistics that involved Analysis of covariance (ANCOVA). Multiple classification analyses were used to determine the magnitude of the testing effects. The data were analysed with the aid of IBM SPSS Statistics 23. All the hypotheses formulated tested at 0.05 level of significance.

4. Results

This section presents the results obtained from data analysis using descriptive statistics for research questions and inferential statistics for hypotheses testing.

Hypotheses Testing

H_01. The testing strategy has no significant main effect on the students' achievement in senior secondary school chemistry.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	954.10	4	238.53	41.80	0.00	0.75
Intercept	1454.62	1	1454.62	254.92	0.00	0.82
Pretest	10.83	1	10.83	1.90	0.17	0.03
Test Strategy	751.52	1	751.52	131.71	0.00	0.71
Gender	12.05	1	12.05	2.11	0.15	0.04
Group * Gender	2.12	1	2.12	0.37	0.54	0.01
Error	313.83	55	5.71			
Total	26812.00	60				
Corrected Total	1267.93	59				

Table 1. Summary of analysis of variance of students' achievement by testing strategy and gender

The results in table 1 show a significant main effect of testing strategy on students' achievement in senior secondary school chemistry ($F_{(1,55)}$= 131.71, p=0.00

The magnitude of the effect of the testing approach on students' achievement as determined by multiple classification analysis is shown in Table 2.

				Predicted mean		Deviation	
			N	**Unadjusted**	**Adjusted for factors**	**Unadjusted**	**Adjusted for factors**
Achievement	**Testing Group**	Individual	30	16.70	16.79	-3.93	-3.84
		Collaborative	30	24.57	24.47	3.93	3.84
	Gender	Male	32	19.34	20.30	-1.29	-0.33
		Female	28	22.11	21.01	1.47	0.37

Table 2. Multiple classification analysis of students' achievement in chemistry by strategy and gender

Table 2 indicates that students exposed to the collaborative testing recorded a higher adjusted post-test mean achievement score of 24.57 than those who engaged in individual testing with an adjusted post-test mean achievement score of 16.70. The difference of 7.87 in their mean achievement score is significant, as shown in Table 1.

$H_0 2$. There is no significant main effect of testing strategy on students' retention of learned chemistry concepts.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	282.00	4	70.50	10.63	0.00
Intercept	147.98	1	147.98	22.30	0.00
Posttest	44.69	1	44.69	6.74	0.01
Test Strategy	0.35	1	0.35	0.05	0.82
Gender	3.88	1	3.88	0.58	0.45
Test strategy * Gender	34.94	1	34.94	5.27	0.03
Error	364.93	55	6.64		
Total	28468.00	60			
Corrected Total	646.93	59			

Table 3. Summary of analysis of variance of students' retention by testing strategy and gender

The results in table 3 show a non-significant main effect of testing strategy on students' retention of learned chemistry concepts ($F_{(1,55)}$= 0.05, p=0.82>0.05). This means that the post-test mean retention scores of chemistry concepts by students exposed to individual testing and collaborative testing do not differ significantly. Hence, the hypothesis states that there is no

significant main effect of testing strategy on students' retention of learned chemistry concepts.

The magnitude of the effect of the testing strategy on students' retention as determined by multiple classification analysis is shown in Table 4.

				Predicted mean		**Deviation**	
			N	**Unadjusted**	**Adjusted for factors**	**Unadjusted**	**Adjusted for factors**
Retention	**Testing Group**	Individual	30	19.80	19.90	-1.73	1.64
		Collaborative	30	23.27	23.17	1.73	1.64
	Gender	Male	32	20.78	21.19	-0.75	-0.34
		Female	28	22.39	21.93	0.86	0.39

Table 4. Multiple classification analysis of students' retention of learned chemistry concepts by strategy and gender

Table 4 discloses that students in the collaborative group recorded a higher mean retention score of 23.17 regarding the learned chemistry concepts than their counterparts in the control group with a mean retention score of 19.80.

*H_03***.** There is no significant main effect of gender on students' achievement in senior secondary school chemistry.

The results in Table 1 reveal no significant effect of gender on students' achievement in senior secondary school chemistry ($F_{(1,55)} = 2.11$, $p=0.15>0.05$). This indicates that male and female students' post-test mean achievement scores in senior secondary school chemistry do not vary significantly. Thus, the hypothesis which states that there is no significant main effect of gender on students' achievement in senior secondary school chemistry is retained.

The magnitude of the effect of gender on students' achievement in senior secondary chemistry as determined by multiple classification analyses is shown in Table 2. The results show that female students recorded higher adjusted post-test mean achievement score of 21.01 compared to the adjusted post-test mean achievement score of 20.30 recorded by the male students. However, this means difference (0.71) is not significant as shown in Table 1.

*H_04***.** There is no significant main effect of gender on students' retention of learned senior chemistry concepts.

The results in Table 3 reveal no significant effect of gender on students' retention of learned senior secondary school chemistry concepts ($F_{(1,55)}$ = 0.58, p=0.45>0.05). This indicates that the post-test mean scores of male and female students regarding the retention of learned senior secondary school chemistry concepts do not vary significantly. Thus, the hypothesis which states that there is no significant main effect of gender on students' retention of learned senior chemistry concepts is retained.

The magnitude of the effect of gender on students' retention of chemistry concepts as determined by multiple classification analyses is shown in Table 4. The results show that female students recorded a marginally higher adjusted post-test mean retention score of 21.93 compared to the adjusted post-test mean retention score of 21.19 recorded by the male students. However, this means difference (0.74) is not significant, as shown in Table 1.

*$H_0$5***.** There is no significant interaction effect of testing strategy and gender on students' achievement in senior secondary school chemistry.

The results in table 1 show a non-significant main interaction effect of testing strategy and gender on students' achievement in senior secondary school chemistry ($F_{(1,55)}$= 037, p=0.54>0.05). This discloses that the adjusted post-test mean achievement of male and female students do not vary significantly across the two levels of the testing strategy.

Students achievement by gender across the testing strategy is as shown in figure 1:

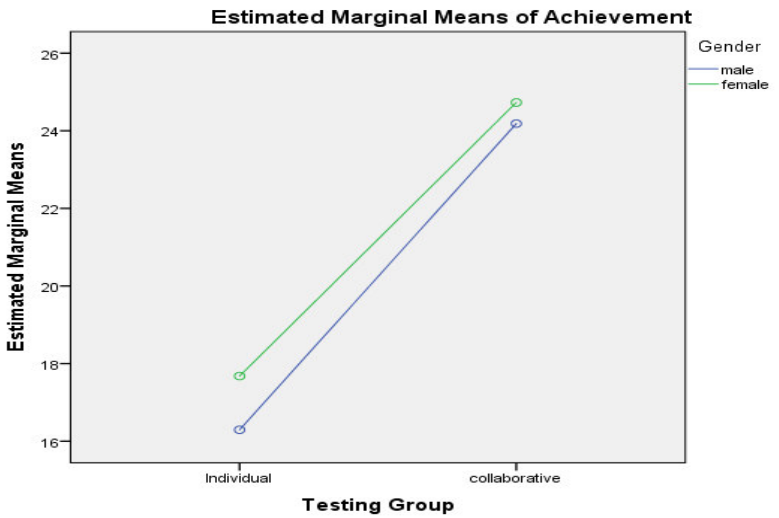

Figure 1. Graph of the interaction effect of testing strategy and gender on achievement

The graph in figure 1 indicates that the students in the collaborative testing group regardless of their gender outperformed their counterparts in the individual testing group. It also reveals that the difference in the adjusted post-test mean achievement scores of males and females in the group engaged in the collaborative testing is smaller than the males and females who used the individual testing approach.

*$H_0$6***.** There is no significant interaction effect of testing strategy and gender on students' retention of learned chemistry concepts.

The results in Table 3 reveal significant effect of gender on students' retention of learned senior secondary school chemistry concepts ($F_{(1,55)} = 5.27$, $p=0.03$

This study examined the interaction effect of testing strategy and gender on students' achievement in chemistry. The findings indicated no significant interaction effect of strategy and gender on students' achievement in chemistry. This means that male and female students' post-test mean achievement scores do not vary significantly across the two testing groups (control and experimental). It further implies that learners' achievement in both testing groups does not depend significantly on their gender. However, students in the collaborative testing group, regardless of their gender, outperformed those in the individual testing group. This may be consequent upon their engagement in social interaction.

Interestingly, the findings further revealed that the gender learning gap in the experimental group is closer compared to the gap in the individual testing group. This suggests that collaborative testing can close the gender learning gap in chemistry. The reduced gender gap may be attributed to the balanced-gender group composition adopted in the study. Cheng (2019) observed that female students tend to benefit from the same gender or balanced ratio groups whereas male students can benefit from any kind of gender-mix. Collaborative efforts would not improve learning if the group composition is defective (Takeda & Homberg, 2014). The finding on the effectiveness of the collaborative testing by gender is congruent to that of Takeda and Homberg (2014) which reported that students in gender-balanced groups (a group with an equal number of males and females) had improved learning performance.

Unsurprisingly, there was a significant interaction effect of testing strategy and gender on students' retention of learned chemistry concepts. This means that the testing strategy significant effect their post-test mean retention scores. This was expected due to the lack of opportunities for the learners to interact in either of the groups during the retention test. The retention tests were taken individually in both groups. Strikingly, however, the students' post-test mean retention scores of male and female students in the groups are in reverse order. This finding implies that males in the control group retain the

chemistry concepts better than their counterparts in the experimental group. Similarly, females in the experimental group retain the chemistry concepts better than their counterparts in the control group.

However, the chemistry mean retention gap across gender was reduced in the experimental group than in the control group. This may be consequent upon the gender composition and the quality of discussion that took place during the collaborative testing, which could have resulted in a deeper understanding of the materials contrary to the situation in the control group. Kapitanoff and Pandey (2019) stated that better interaction during collaborative testing could translate to better performance in future examinations. This means that collaborative testing could have improved students' knowledge retention, although a gap existed in their mean retention scores across gender. Cheng (2019) has argued that individuals do not benefit from collaboration in the same proportion because achievement in collaborative efforts depends on diverse factors such as gender difference and the nature of the tasks. This finding agrees with Kleinberg, *et al.* (2018) that collaborative testing improved long-term retention of acquired knowledge.

5. Conclusion

The study examined the effect of collaborative testing on students' achievement and retention of learned senior secondary school chemistry concepts with gender as a moderating variable. Based on the findings, the study concludes that collaborative testing strategy significantly improved students' achievement in Chemistry. However, the effect of the strategy on students' retention of learned chemistry concepts was not potent. Still, students that engaged in collaborative testing gained more knowledge retention than those in the control group.

Also, gender is not a significant factor in students' achievement and retention of chemistry knowledge in senior secondary school in Nigeria. Meanwhile, the collaborative learning strategy reduced the gender learning gap in the subject. This means that collaborative testing can reduce the gender learning gap in chemistry. This research's findings support and extend the knowledge that a collaborative testing strategy is effective. They also serve as evidence that collaborative testing strategy can improve students' learning and retention of knowledge of chemistry in senior secondary school chemistry and reduce the gender achievement gaps in the learning of the subject for sustainable development. These findings should encourage chemistry teachers to engage students in collaborative testing before a terminal examination to have a quality discussion, exchange ideas, and reduce the anxiety that will deepen their understanding of the contents. The finding is also a call on the government to organize workshops for teachers on using collaborative testing strategies for learning besides the regular use of testing for assessment

purposes. Also, the strategy should be integrated into the chemistry teacher education training programme to acquaint would-be teachers with the operational use of the strategy before graduation.

Despite the improvement in students' achievement recorded with collaborative testing, the researchers observed that few students did not perform as expected. However, this study did not put a mechanism to address those "at-risk" learners. Based on this, collaborative testing suggests further studies with a mechanism to deal with students in this category. Similarly, the findings of this study may not be generalized due to the small sample. Therefore, there should be further studies with more schools, and students should extend their knowledge on the effectiveness of collaborative testing.

References

Acharya, K. P., Acharya, M., & Shrestha, M. K. (2020). Collaborative learning practices by gender: A case of a community school in Nepal. *International Education Studies*, *13*(4), 75–83. doi: 10.5539/ies.v13n4p75

Breane-Boove, D. C. (2016). The impact of collaboration testing on test anxiety. *Chiropractic Journal of Australia, 44*(3), 215-221.

Caboral-Stevens, M., & Fox, D. P. (2020). The use of collaborative testing with Baccalaureate nursing students. *Teaching and Learning in Nursing, 15*(1), 37-41. doi: 10.1016/j.teln.2019.09.007

Chabari, K. K., Wachanga, S. W., & Anditi, Z. O. (2018). Effects of graphic organizers experiential teaching approach on secondary school students' achievement in Chemistry in Nakuru North Sub-County, Kenya. *IOSR Journal of Humanities and Social Sciences*, *23*(2), 79–91. doi: 10.4236/ce.2018.912137

Cheng, H. Y. F. (2019). The role of gender in collaborative learning. *EC Psychology and Psychiatry, 8*(7), 721-725.

Conejo, R., Barros, B., Guzman, E., & Garcia-Vinas, J.I. (2013). A web based collaborative testing environment. *Computers & Education*, *68*, 440–457. doi: 10.1016/j.compedu.2013.06.001

Eastridge, J. A., & Benson, W. L. (2020). Comparing two models of collaborative testing for teaching statistics. *Teaching of Psychology*, *47*(1), 68–73. doi: 10.1177/0098628319888113

Eastwood, J. L., Kleinberg, K. A., & Rodenbaugh, D. W. (2020). Collaborative testing in medical education: Students perceptions and long-term retention. *Medical Science Education, 30,* 737-747. doi: 10.1007/s40670-020-00944-x

Efu, S. I. (2019). Exam as learning tools: A comparison of traditional and collaborative assessment in higher education. *College Teaching, 67*(1), 73-83. doi: 10.1080/87567555.2018.1531282

Ejedegbe, O. (2016). Gender differences in secondary schools students' achievement in Chemistry. *International Journal of Trend in Research and Development*, *3*(3), 55–61.

Ekpen, M. (2020). *Influence of scientific attitudes on student's academic performance in chemistry in Bekwarra Local Government Area of Cross River State.* (SSRN Scholarly Paper No. ID 3554301). Rochester, NY: Social Science Research Network. doi: 10.2139/SSRN.3554301

Elschami, M., & Kummerer, K. (2020). Design of a master of science sustainable chemistry. *Sustainable Chemistry and Pharmacy, 17* (2020). doi: 10.1016/j.scp.2020.100270

Federal Ministry of Education. (2018). *Science and technology education policy*. Retrieved from https://education.gov.ng/national-policy-on-science-and-technology-education/

Galal, S., Vyas, D., Hackett, R. K., Rogan, E., & Nguyen, C. (2021). Effectiveness of Music Interventions to Reduce Test Anxiety in Pharmacy Students. *Pharmacy*, *9*(10), 1–10. doi: 10.3390/pharmacy 9010010

Helms, K. D., Keith, L. A., & Walker, L. P. (2019). Evaluating student learning through collaborative testing in a psychiatric mental health course. *Journal of Doctoral Nursing Practice*, *12*(1), 3–9. doi: 10.1891/2380-9418.12.1.3

Juntunen, M. (2016). *Sustainable Development in Science Education*. Retrieved from https://www.luma.fi/en/files/2018/04/Juntunen_KeKe_ENG.pdf

Kapitanoff, S., & Pandey, C. (2019). Collaborative testing in statistics: Group interaction, anxiety, and class performance. *Statistics Education Research Journal, 17*(2), 51-67.

Kleinberg, K., Eastwood, J. L., & Rodernbough, D. W. (2018). Efficacy of collaborative testing for long-term retention of medical knowledge. *The FASEB Journal, 32*(1), supplement, 21.6. doi: 10.1096/fasebj.2018. 32.1_supplement.21.6

Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrollment introductory biology class. *CBE—Life Sciences Education. 11*, 392–401.

Lerman, Z. M. (2014). The challenges for chemistry education in Africa. *African Journal of Chemical Education, 4*(2), 80-90.

Maciejewski, M. L. (2020). Quasi-experimental design. *Biostatistics & Epidemiology*, *4*(1), 38–47. doi: 10.1080/24709360.2018.1477468

Mahoney, J. W. & Harris-Reeves, B. (2019). The effects of collaborative testing on higher-order thinking: Do the bright get brighter? *Active Learning in Higher Education, 20*(1), 25-37. doi: 10.1177/1469787417 723243.

Nbina, J. B. (2012). Analysis of poor performance of senior secondary students in chemistry in Nigeria. *An International Multidisciplinary Journal, 6*(4), 324-334

Ojukwu, M. O. (2016). Perception of students on causes of poor performance in chemistry in external examinations in Umuahia North Local Government of Abia State. *International Journal of Education & Literacy Studies, 4*(1), 68-73. doi: 10.7575/a1ac.ijels.v.4n/p.67.

Okeke, O. J. (2018). Effect of gender on mean achievement score of chemistry students taught using mind mapping strategy (MMTS) in Enugu. *International Journal in Physical and Applied Sciences*, *5*(08), 19.

Pate, A. N., Neely, S., Malcom, D. R., Daugherty, K. K., Zagar, M., & Medina, M. S. (2021). Multisite Study Assessing the Effect of Cognitive Test Anxiety on Academic and Standardized Test Performance. *American Journal of Pharmaceutical Education*, *85*(1). doi: 10.5688/ajpe8041

Phillips, T. A., Munn, A. C., & George, T. P. (2019). The Impact of Collaborative Testing in Graduate Nursing Education. *Journal of Nursing Education*, *58*(6), 357–359. doi: 10.3928/01484834-20190521-07

Popoola, A. R., & Solomon, O., A. (2017). The use of problem-based instruction strategy pbis in gas laws class: A gender academic performance difference. *International Journal of Trend in Scientific Research and Development*, *1*(4), 458–464. doi: 10.31142/ijtsrd171

Popoola, A. R., & Solomon, O., A. (2017). The use of problem-based instruction strategy pbis in gas laws class: A gender academic performance difference. *International Journal of Trend in Scientific Research and Development*, *1*(4), 458–464. doi: 10.31142/ijtsrd171

Retnowati, E., Ayres, P. & Sweller, J. (2018). Collaborative learning effects when students have complete or incomplete knowledge. *Applied Cognitive Psychology*, *32*(6). doi: 10.1002/acp.3444

Rosa, W. (Ed.). (2017). Transforming Our World: The 2030 Agenda for Sustainable Development. In *A New Era in Global Health*. New York, NY: Springer Publishing Company. doi: 10.1891/9780826190123.ap02

Shiqiqi, A. S., Permanasari, A., & Hernani, H. (2020). Review on education for sustainable chemistry education curriculum. *J. Phys: Conf. Ser. 1521*(4), 042080. doi.10.1088/1742-6596/1521/4/042080

Slusser, S. R., & Erickson, R. J. (2006). Group quizzes: An extension of the collaborative learning process. *Teaching Sociology*, *34*(3), 249–262.

Takeda, S., & Homberg, F. (2014). The effects of gender on group work process and achievement: An analysis through self-and peer-assessment. *British Educational Research Journal*, 40 (2), 373-396. doi: 10.1002/berj.3088.

Thomas, C. L. (2021). Predicting test anxiety using the 3x2 achievement goal model. *International Journal of School & Educational Psychology*, *1*, 1–11. doi: 10.1080/21683603.2020.1816237

Vazquez-Garcia, M. (2018). Collaborative-group testing improves learning and knowledge retention of human physiology topics in second-year medical students. *Advances in Physiology Education, 42,* 232-239. doi: 10.1152/advan.00113.2017

Vogler, J. S., & Robinson, D. H. (2016). Team-based testing improves individual learning. *The Journal of Experimental Education*, *84*(4), 787–803. doi: 10.1080/00220973.2015.1134420.

West Africa Examinations Council Chief Examiners' Report. (2014). *General comment, weakness/remedies and candidates' strengths*. Retrieved from https://www.waeconline.org.ng/e-learning/Chemistry/chem223mc.html

West Africa Examinations Council Chief Examiners' Report. (2015). *General comment, weakness/remedies and candidates' strengths*. Retrieved from https://www.waeconline.org.ng/e-learning/Chemistry/chem324mc.html

West Africa Examinations Council Chief Examiners' Report. (2016). *General comment, weakness/remedies and candidates' strengths*. Retrieved from https://www.waeconline.org.ng/e-learning/Chemistry/chem225mc.html

West Africa Examinations Council Chief Examiners' Report. (2017). *General comment, weakness/remedies and candidates' strength*. Retrieved from https://www.waeconline.org.ng/e-learning/Chemistry/chem226mc.html

West Africa Examinations Council Chief Examiners' Report. (2018). *General comment, weakness/remedies and candidates' strength*. Retrieved from https://www.waeconline.org.ng/e-learning/Chemistry/chem227mc.html