Journal of Studies in Science and Mathematics Education
ISSN: 2630-662x
Published: 23.09.2026
Volume (issue): ***6*****(2)**
Page: 57 – 63

Gender-Stratified and Interaction Effects of Teaching Method on Senior Secondary Students' Acquisition of Communicating Skills in Biology

Festus Osasumwen Idiaghe*, Esther Ngozi Odafe

[1]Department of Curriculum and Instructional Technology, Faculty of Education, University of Benin, Benin City, Edo State, Nigeria

***Corresponding author:** festus.idiaghe@uniben.edu

Abstract

This study tested whether the ABI > GI > TM effectiveness ordering reported for a pooled sample of 127 senior secondary Biology students (Idiaghe & Odafe, 2026a) replicates for male and female students separately, and whether teaching method and gender interact on posttest communicating-skills scores. A quasi-experimental pretest-posttest design was used with students in Egor Local Government Area, Edo State, Nigeria, taught animal nutrition using Argument-Based Inquiry (ABI), Guided Inquiry (GI), or the Traditional Method (TM). Communicating skills were measured with the Science Written Communicating Skills Acquisition Test (SWCSAT, $\alpha = 0.74$). Among 59 male students, ANCOVA showed a significant method effect, $F(2, 55) = 434.98$, $p < .001$. Among 68 female students, ANOVA showed a significant method effect, $F(2, 65) = 169.02$, $p < .001$, with Tukey tests confirming all pairwise differences (ABI > GI > TM, $p < .001$). A two-way ANOVA on the full sample showed significant main effects of method ($\eta^2p = .89$) and gender ($\eta^2p = .05$), but a small, non-significant interaction, $F(2, 121) = 2.57$, $p = .081$, $\eta^2p = .04$. Given unbalanced cells (13-28 per group), this null interaction is read as absence of evidence, not evidence of absence.

Keywords: argument-based inquiry, guided inquiry, gender, method-by-gender interaction, two-way ANOVA, communicating skills, Biology education

1 Introduction

A pooled-sample finding that one instructional method outperforms others can mask an underlying gender-specific pattern - it is possible for a method to be highly effective for one sex and only mildly effective for the other, with the pooled average obscuring this divergence. Using a pooled sample of 127 Senior Secondary School One (SS1) Biology students in Egor Local Government Area, Edo State, Nigeria, Idiaghe and Odafe (2026a) established that Argument-Based Inquiry (ABI) produced significantly greater gains in written communicating skills than both Guided Inquiry (GI) and the Traditional lecture Method (TM). Analysing the same dataset disaggregated by gender, Idiaghe and Odafe (2026b) further reported a statistically significant, though modest, gender gap specifically within the ABI condition ($t(40) = 2.22$, $p = .03$; males scoring 9.59 points higher than females on average) and no significant gender gap within GI ($t(38) = 1.90$, $p = .06$). Neither of these prior findings, individually, can establish whether the overall method effect itself differs in size by gender - that is, whether method and gender interact - as opposed to method and gender each independently shifting communicating-skills scores by a separate, additive amount. This distinction matters practically: if ABI's advantage over GI and TM were substantially larger for one gender than the other, a uniform recommendation to adopt ABI could unintentionally widen an existing gender gap. The present study builds directly on Idiaghe and Odafe (2026a, 2026b) by (a) testing whether the ABI > GI > TM ordering they reported for the pooled sample replicates independently within male and female subgroups, and (b) testing directly, using a two-way analysis of variance not reported in either prior paper, whether a formal Method × Gender interaction exists. This is the specific, incremental contribution of the present manuscript relative to the two companion analyses.

2 Literature Review/Theoretical Framework

2.1 Argument-Based Inquiry and the Basis for the Method Comparison

Argument-Based Inquiry embeds the claim-evidence-reasoning argumentation cycle that characterises scientific discourse (Duschl & Osborne, 2002) within a hands-on inquiry sequence built around group investigation and structured peer review of written reports (Sampson & Schleigh, 2013). Guided Inquiry retains the same underlying investigative activity but omits this explicit argumentation and peer-review layer, while the Traditional lecture Method omits both. This graded structure - TM, then GI, then ABI - is what motivates testing whether the resulting effectiveness ordering holds uniformly across genders rather than assuming it does.

2.2 Gender, Interaction Effects and the Rationale for Disaggregated Analysis

Hyde's (2005) gender similarities hypothesis, based on a synthesis of 46 meta-analyses of psychological and cognitive outcomes, shows that gender differences are typically small and highly context-dependent, varying by task, age, and setting rather than reflecting fixed, generalisable differences between the sexes. This context-dependence is precisely why an aggregate instructional-effectiveness finding should be checked for gender-specific patterns before being generalised: a genuine but context-specific gender effect could otherwise be masked, or a spurious one exaggerated, in a pooled analysis. Nigerian classroom-based studies of gender-by-method interaction in Biology have generally reported non-significant interactions. Nkok (2022), studying 178 SS1 students taught sexual reproduction in plants using computer simulation versus conventional teaching, found no significant gender-by-method interaction for interest ($F = 0.10$, $p = .963$), achievement ($F = 0.53$, $p = .466$), or retention ($F = 1.67$, $p = .199$), concluding that the instructional strategy examined benefited male and female students similarly. The present study extends this line of Nigerian evidence to a different instructional contrast (ABI, GI and TM) and a different outcome (written communicating skills), providing a second, topic-independent test of whether gender-by-method interactions are a general or a context-specific feature of Nigerian Biology instruction.

A statistically significant interaction effect between an instructional treatment and a demographic characteristic such as gender would imply that the treatment's benefit is not uniform across groups - for example, that ABI's advantage over TM is meaningfully larger for one gender than the other. The absence of such an interaction implies instead that the treatment effect is additive with any demographic main effect: method shifts everyone's scores by roughly the same amount, and gender shifts everyone's scores by a separate, smaller amount, independent of method (National Research Council, 2007).

2.3 Research questions

a) What is the difference in posttest mean communicating-skills scores among male students, and separately among female students, taught with ABI, GI and TM?

b) What is the interaction effect of teaching method and gender on posttest communicating-skills scores?

2.4 Null hypotheses

The following null hypotheses were tested at the 0.05 level of significance:

a) There is no significant difference in posttest mean communicating-skills scores among male students taught with ABI, GI and TM.

b) There is no significant difference in posttest mean communicating-skills scores among female students taught with ABI, GI and TM.

c) There is no significant interaction effect of teaching method and gender on posttest mean communicating-skills scores.

3 Materials and Methods

3.1 Design

The study adopted a quasi-experimental pretest-posttest, non-randomized equivalent control-group design. Three intact classes drawn from three different co-educational public senior secondary schools in Egor Local Government Area of Edo State, Nigeria, served as Experimental Group A (Argument-Based Inquiry, ABI), Experimental Group B (Guided

Inquiry, GI) and Control Group C (Traditional lecture Method, TM). Random assignment was applied at the level of schools and intact classes rather than individual students, consistent with standard practice in classroom-based quasi-experimental research. Because each condition was delivered in a single intact class, students within a condition share a common classroom and teacher context; this clustering is not statistically modelled in the ANCOVA/ANOVA analyses reported below (which treat students as independent observations) and is discussed as a limitation.

3.2 Population, Sample and Sampling Technique

The target population comprised 2,544 Senior Secondary School One (SS1) Biology students (1,297 male, 1,247 female) drawn from thirteen public secondary schools in Egor Local Government Area of Edo State. Three co-educational schools were selected at random from the eleven co-educational schools in the Local Government Area, and one intact SS1 class was randomly drawn from each school to represent the respective study group. The gender-stratified analyses used 59 male students (ABI $n = 18$, GI $n = 13$, TM $n = 28$) and 68 female students (ABI $n = 24$, GI $n = 27$, TM $n = 17$). The interaction analysis used the full sample of 127 students cross-classified by method and gender. These method-by-gender cell sizes are unbalanced (ranging from 13 to 28 students); this is acknowledged as a constraint on the equal-variance assumptions of ANOVA/ANCOVA and on the power of the interaction test, and is discussed further below.

3.3 Instrument

Data were collected using the Science Written Communicating Skills Acquisition Test (SWCSAT), a researcher-developed instrument comprising eleven items built around three practical Biology tasks on the concept of animal nutrition, adapted from past West African Examinations Council (WAEC) practical Biology questions. Scoring followed a structured written-communicating-skills assessment guide and evaluated five components: content, organization, support for explanation, observing, and measuring. The instrument was subjected to face and content validation by the researchers' supervisors and three additional experts, and a pilot trial with 30 SS1 students outside the main sample yielded a Cronbach alpha reliability coefficient of 0.74. An alpha of .74 is generally regarded as acceptable, though modest, for an eleven-item, five-component rubric-scored instrument; some measurement error is therefore likely across the five components, which would tend to attenuate rather than inflate the group differences reported below.

3.4 Procedure and Data Analysis

The intervention lasted six weeks. The SWCSAT was administered as a pretest to all three groups before instruction began. Experimental Group A was thereafter taught the concept of animal nutrition using an eleven-step Argument-Based Inquiry sequence involving group investigation, evidence-based claim-making, whole-class argumentation, report writing and structured peer review. Experimental Group B received the same topic through an eight-step Guided Inquiry sequence in which the teacher posed guiding questions and structured tasks, but without the explicit argumentation and peer-review cycle. Control Group C was taught using the conventional lecture method. Participating teachers received a one-week training and were supervised throughout to ensure treatment fidelity. The SWCSAT was re-administered as a posttest in week six.

For male students, pretest scores differed significantly among groups, so ANCOVA with pretest as covariate was used. For female students, pretest scores did not differ significantly among groups, so one-way ANOVA with Tukey post-hoc comparisons was used. The interaction hypothesis was tested with a two-way (factorial) ANOVA on the full sample, examining the main effect of method, the main effect of gender, and the Method × Gender interaction term. Partial eta-squared ($\eta^2 p$) effect sizes were computed post hoc from the reported sums of squares to aid interpretation of practical, and not only statistical, significance. Homogeneity-of-variance diagnostics (e.g., Levene's test) were not computed in the original analysis; this omission is acknowledged as a limitation in the Discussion. All hypotheses were tested at the 0.05 level of significance. Ethical procedures including informed consent from school authorities, voluntary participation, and confidentiality of student data were observed throughout the study, consistent with standard practice in classroom-based educational research in Nigeria.

4 Results

Table 1 presents posttest composite means by method within each gender stratum. Table 2 presents the ANCOVA for male students. Table 3 presents the one-way ANOVA and Tukey post-hoc comparisons for female students. Table 4 presents the two-way ANOVA testing the Method × Gender interaction. Table 5 summarises partial eta-squared effect sizes for all three analyses. Figure 1 displays the interaction plot.

Table **1**

Posttest Composite Communicating Skills Scores by Method, Stratified by Gender

Gender	Method	N	Posttest M (SD)	Pretest-Posttest Gain
Male	ABI	18	156.39 (14.12)	106.28
Male	GI	13	114.23 (12.24)	67.92
Male	TM	28	79.69 (9.28)	34.11
Female	ABI	24	146.80 (17.69)	99.14
Female	GI	27	107.88 (12.54)	62.85
Female	TM	17	80.41 (9.00)	34.71

M = Mean, SD = Standard Deviation.

Table **2**

Summary of ANCOVA of Posttest Communicating Skills Scores of Male Students by Method

Source	Type III SS	df	Mean Square	F	Sig.	η^2p
Corrected Model	66398.23	3	22132.74	389.13	.000	-
Method	49481.66	2	24740.83	434.98	.000	.94
Pretest (covariate)	1716.41	1	1716.41	30.18	.000	-
Error	3128.28	55	56.88			
Corrected Total	69526.51	58				

$p < .05$ is significant. η^2p = partial eta-squared, computed as SSeffect / (SSeffect + SSerror).

Table **3**

One-Way ANOVA and Tukey Post-Hoc Comparisons of Posttest Communicating Skills Scores of Female Students by Method

Source / Comparison	SS / Mean Diff.	df	F	Sig.	η^2p
Between Groups	45956.01	2	169.02	.000	.84
Within Groups	8836.74	65	-	-	
ABI vs GI (post-hoc)	+38.90	-	-	.000	
ABI vs TM (post-hoc)	+66.38	-	-	.000	
GI vs TM (post-hoc)	+27.48	-	-	.000	

$p < .05$ is significant. Positive mean differences indicate the first-named method scored higher.

Table **4**

Summary of Two-Way ANOVA of Posttest Communicating Skills Scores by Method and Gender

Source	Type III SS	df	Mean Square	F	Sig.	η^2p
Corrected Model	111157.29	5	22231.46	196.62	.000	-
Method	107220.87	2	53610.44	474.14	.000	.89
Gender	756.31	1	756.31	6.69	.011	.05
Method × Gender	580.79	2	290.39	2.57	.081	.04
Error	13681.44	121	113.07			
Total (Corrected)	124838.72	126				

$p < .05$ is significant. η^2p = partial eta-squared.

Table **5**

Summary of Partial Eta-Squared Effect Sizes Across the Three Analyses

Analysis	Effect	η^2p	Interpretation
Male ANCOVA	Method	.94	Very large
Female ANOVA	Method	.84	Very large
Two-way ANOVA (full sample)	Method	.89	Very large
Two-way ANOVA (full sample)	Gender	.05	Small-to-medium
Two-way ANOVA (full sample)	Method × Gender	.04	Small

Effect sizes computed by the authors from the sums of squares reported in Tables 2-4; not directly output by the original SPSS analysis. Benchmarks follow Cohen's (1988) conventions for partial eta-squared (.01 small, .06 medium, .14 large).

Figure **1**

Interaction Plot of Teaching Method and Gender on Posttest Communicating Skills Scores

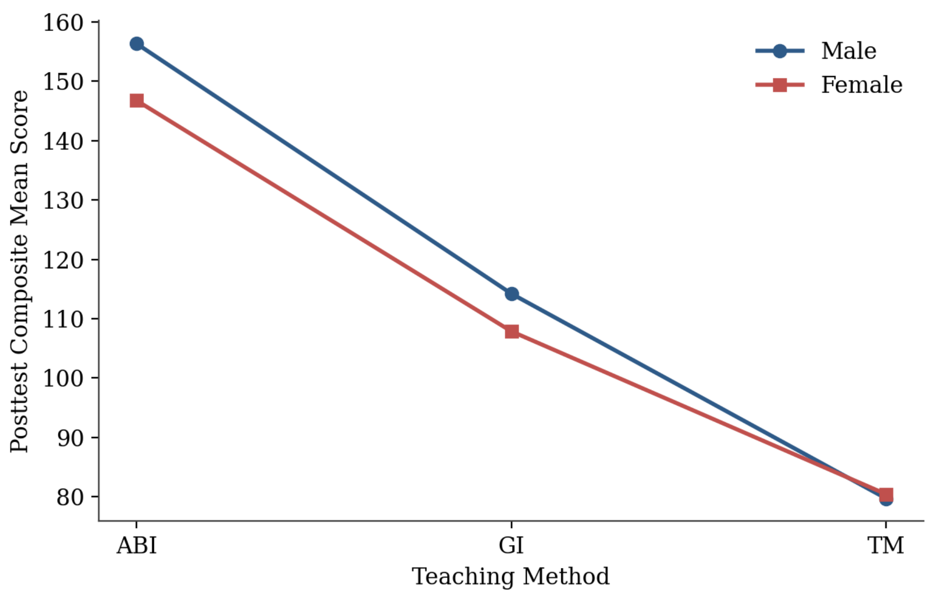

Table 1 shows the same ABI > GI > TM ordering of posttest means within both gender strata. For male students, the ANCOVA in Table 2 shows a statistically significant, very large effect of method, $F(2, 55) = 434.98$, $p < .001$, $\eta^2p = .94$ (H1 rejected). For female students, the one-way ANOVA in Table 3 shows a statistically significant, very large effect of method, $F(2, 65) = 169.02$, $p < .001$, $\eta^2p = .84$ (H2 rejected), with Tukey post-hoc comparisons confirming all three pairwise differences among ABI, GI and TM are statistically significant (all $p < .001$). The two-way ANOVA in Table 4 shows significant main effects of method, $F(2, 121) = 474.14$, $p < .001$, $\eta^2p = .89$, and gender, $F(1, 121) = 6.69$, $p = .011$, $\eta^2p = .05$, but a small, non-significant Method × Gender interaction, $F(2, 121) = 2.57$, $p = .081$, $\eta^2p = .04$ (H3 retained). Table 5 summarises these effect sizes: the method effect is very large in every analysis, the gender main effect is small-to-medium, and the interaction effect is small. Figure 1 illustrates this pattern: the male and female lines run broadly parallel across ABI, GI and TM, converging only at TM, without a statistically significant crossing or divergence in slope.

5 Discussion

The replication of the ABI > GI > TM ordering within both the male and female subsamples, using different but statistically appropriate adjustments for each, provides evidence that the overall effectiveness ranking reported by Idiaghe and Odafe (2026a) for the pooled sample is not an artefact of the sample's gender composition. This is consistent with

Nkok's (2022) finding, in a different Nigerian Biology topic and instructional contrast, that gender-by-method interactions were non-significant for interest, achievement and retention alike, suggesting that additive rather than interactive method-and-gender effects may be a broader pattern in Nigerian Biology classrooms rather than a feature specific to the present dataset. The combination of significant main effects with a non-significant interaction indicates that method and gender each independently influence communicating-skills scores but do so additively rather than multiplicatively: ABI's advantage over GI and TM is not statistically distinguishable in magnitude between male and female students in the present two-way analysis, even though Idiaghe and Odafe (2026b) found a significant male advantage specifically within the ABI condition ($t(40) = 2.22$, $p = .03$, a 9.59-point mean difference) when that condition was examined in isolation. The more conservative, full-sample interaction test reported here does not confirm this ABI-specific gender pattern as a formal statistical interaction once all three methods and both genders are modelled jointly.

Several limitations qualify these conclusions and temper the interaction-null finding specifically. First, random assignment occurred at the level of intact classes rather than individual students, and each condition was delivered in a single class; the ANCOVA and ANOVA models reported here treat students as independent observations and do not correct for this clustering. With only one class per condition, a fully specified multilevel model was not feasible in the present design, so residual within-class dependence may make the reported standard errors somewhat anti-conservative; this is a structural limitation of the design rather than an analytical choice that could be corrected post hoc. Second, the method-by-gender cells were unbalanced (ranging from 13 to 28 students) and homogeneity-of-variance diagnostics such as Levene's test were not computed for the original analysis; unequal variances combined with unequal cell sizes can distort both Type I error rates and statistical power in ANOVA designs, and this should be reported and checked in any replication. Third, and most directly relevant to the central interaction claim, the partial eta-squared for the Method × Gender interaction ($\eta^2 p = .04$) was small, and statistical power to detect an interaction of this size with the present cell sizes was limited. The non-significant interaction p-value of .081 should therefore be read as an absence of evidence for an interaction, not as evidence of its absence; a larger, gender-balanced replication is needed before concluding with confidence that teaching method and gender do not interact in this population.

6 Conclusion

The superiority of Argument-Based Inquiry over Guided Inquiry and the Traditional lecture Method for communicating-skills acquisition, reported by Idiaghe and Odafe (2026a) for a pooled sample, replicates independently within both male and female student subgroups. Teaching method and student gender each independently and significantly affect communicating-skills acquisition, and the present two-way analysis does not detect a statistically significant interaction between them, though the small effect size and unbalanced, clustered design mean this should be treated as a provisional rather than a definitive finding. Overall, the results support a largely unified instructional strategy - adopting ABI for its method-level benefit - paired with continued, gender-disaggregated outcome monitoring rather than a gender-differentiated pedagogical redesign. As a brief direction for future work, any large-scale digital or AI-assisted implementation of ABI should build gender-disaggregated outcome monitoring into its analytics from the outset, given the significant, independent gender main effect identified here (Zawacki-Richter, Marin, Bond & Gouverneur, 2019).

7 Recommendations

1. Method-choice recommendations favouring ABI over GI and TM can be applied to male and female students alike, based on the gender-stratified replication reported here.
2. Schools and researchers should continue to monitor gender-disaggregated outcomes within methods, given the significant, independent gender main effect identified.
3. Future replications should use larger, gender-balanced method-by-gender cells, report homogeneity-of-variance diagnostics, and where feasible use multiple classes per condition to permit multilevel modelling of the school/class-level clustering inherent in this design.
4. Future replications should extend the gender-stratified and interaction analyses to additional Biology topics before generalising the present non-significant interaction finding.

8 Declarations

Ethical statements: Ethical procedures including informed consent from school authorities, voluntary participation, and confidentiality of student data were observed throughout the study, consistent with standard practice in classroom-based educational research in Nigeria.

Conflict of Interest: The authors declare neither financial nor non-financial conflict of interest.

Author Contributions: F.O.I. conceived the study, collected the data, and conducted the primary statistical analyses. E.N.O. contributed to the literature review, interpretation of findings, and critical revision of the manuscript. Both authors read and approved the final manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors acknowledge the Biology teachers and students of the participating schools in Egor Local Government Area, Edo State.

Data Availability Statement: The dataset analysed is available from the corresponding author on reasonable request.

9 Reference

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, *38*(1), 39–72. https://doi.org/10.1080/03057260208560187

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066X.60.6.581

Idiaghe, F. O., & Odafe, E. N. (2026a). Comparative and pairwise effects of Argument-Based Inquiry and Guided Inquiry on senior secondary students' acquisition of communicating skills in Biology. Manuscript in preparation, University of Benin.

Idiaghe, F. O., & Odafe, E. N. (2026b). Gender and instructional approach: Differential effects of Argument-Based Inquiry and Guided Inquiry on male and female students' acquisition of communicating skills in Biology. Manuscript in preparation, University of Benin.

National Research Council. (2007). Taking science to school: Learning and teaching science in grades K–8. National Academies Press.

Nkok, E. A. E. (2022). Interaction effect of gender and teaching methods on students' interest, achievement and retention in sexual reproduction in plant. *International Journal of Innovative Social & Science Education Research*, *10*(4), 56–65.

Sampson, V., & Schleigh, S. (2013). Scientific argumentation in biology: 30 classroom activities. NSTA Press.

Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, *16*, 39. https://doi.org/10.1186/s41239-019-0171-0.

Publisher's Note

The views, interpretations, and conclusions expressed in this article are solely those of the author(s) and do not necessarily reflect the views or policies of the Journal of Studies in Science and Mathematics Education (JOSSME), its publisher, editorial board, editors, reviewers, or the authors' affiliated institutions. The publisher remains neutral with respect to jurisdictional claims in published maps, institutional affiliations, and other representations. Any products, materials, methods, or manufacturers mentioned or evaluated in this article are presented solely for scholarly purposes and do not constitute endorsement, recommendation, or guarantee by the publisher.