



Peer Review Report Form

Manuscript Title:	Integration of Authentic Assessment in Teaching Science and Academic Performance among Students
Type of Article:	Original Article

FIRST ROUND

Editor

<u>General Comments and Recommendations</u>
The reviewers acknowledge the significance of your manuscript and recommend several revisions to further improve its clarity, coherence, organization, and overall scholarly quality. In addition, you are advised to thoroughly review and validate all references included in the manuscript. Please ensure that all cited sources are accurate, accessible, and retrievable through Google Scholar to uphold the credibility and reliability of the study. Furthermore, only sources that are cited within the manuscript should appear in the References section, and all entries listed in the References section must also be cited in the text. For example, if 20 unique sources are cited in the manuscript, then only those 20 sources should be included in the References section. Please also note that repeated citations of the same source within the text are counted as a single reference entry; therefore, the total number of references may be lower than the total number of in-text citations.

Reviewer 1

<u>General Comments and Recommendations</u>
The paper discusses a timely and useful topic in science education, especially on the use of authentic assessment in public junior high school science classes. The study is relevant and has practical value for teachers and school leaders.

<u>Specific Comments and Recommendations</u>
The study wants to determine whether authentic assessment is related to students' academic performance, but the way the variables were measured may not be strong enough to support this claim. The integration of authentic assessment was based on teachers' self-report, while academic performance was based on students' first-quarter grades. These two data sources may not directly match each other because grades can be affected by many factors such as prior knowledge, grading practices, written works, performance tasks, attendance, teacher leniency, and school-level policies. It is also not clear if the students' grades used in the analysis were specifically linked to the same teachers who answered the survey, and whether the grades actually reflected authentic assessment outputs or general quarterly grades only. Because of this, the finding of "no significant relationship" may not really mean that authentic assessment has no connection with performance; it may only show that the study design and measurement were not precise enough to detect the relationship.



Please indicate your recommendation by checking the appropriate box below.

<u>Decision</u>	
	Accept the manuscript for publication.
/	Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.
	Reject the manuscript, as it is not suitable for publication.

Reviewer 2

<u>General Comments and Recommendations</u>
The paper presents a very timely investigation. In the context of our shift towards outcomes-based education and the demands of the K to 12 curriculum in the Philippines, looking into how our JHS teachers in the local districts actually implement authentic assessment (AA) is highly commendable.
<u>Specific Comments and Recommendations</u>
The most critical issue in your study is how you measured the dependent variable. You utilized the students' first-quarter grades as the metric for academic performance , which were then averaged per teacher respondent to pair with the teacher's perceived integration of authentic assessment. This approach is highly circular. The quarterly grades given by the teachers are the direct result of their own assessment methods. A teacher who heavily uses traditional objective tests can still give a class average of 85, just as a teacher who uses performance tasks can give an 85. Correlating a teacher's pedagogical strategy with the subjective grades they themselves generate for their students does not give us a true picture of academic achievement. You even noted in your discussion that there could be a "ceiling effect". This methodological choice is very likely the reason why your Spearman rank-order correlation yielded a coefficient of -0.006 and a p-value of 0.971. To truly measure if AA improves learning, you need an independent, objective measure of students' cognitive outcomes, such as a standardized division-wide test or a validated achievement test, rather than teacher-generated grades. Since the data is already gathered, you must heavily discuss this as a major limitation of your study that impacted your results.
You mentioned that you used universal sampling which resulted in 40 JHS science teacher respondents. While N=40 covers your target population for those districts, it is quite small for establishing robust correlational relationships. The bigger issue appears in Table 5, where you used the Kruskal-Wallis H Test to find significant differences when respondents are grouped by demographic variables. Looking back at Table 1, your sub-groups are very uneven and small. For example, you only have 4 teachers with 1-3 years of teaching experience, and only 2 teachers with 4-6 years of experience. Running statistical tests to compare such small cell sizes severely lacks statistical power, making your non-significant findings highly vulnerable to Type II error (failing to detect a difference that might actually exist). I suggest collapsing these demographic groups into broader, more balanced categories (e.g., early-career vs. late-career teachers) before running the test, or explicitly stating this lack of statistical power as a limitation.



There is a significant mismatch between your statistical results and your narrative conclusions. In your abstract, you claimed that "authentic assessment helps promote meaningful learning". However, your actual data proved the opposite for this specific cohort: it showed absolutely no significant relationship between AA and academic performance. Claiming that it promotes meaningful learning without empirical backing from your own results is an overreach. Furthermore, in the discussion and conclusion, you attributed the non-significant relationship to "external factors such as student grit and prior knowledge" and students' "lack of grit and self-motivation". While these are valid concepts supported by literature, your study did not measure student grit, motivation, or prior knowledge. You cannot make definitive conclusions about variables you did not test. Please revise your discussion and conclusion to reflect only what your data actually proved, and avoid making speculative claims to justify the non-significant relationship.

Please indicate your recommendation by checking the appropriate box below.

<u>Decision</u>	
	Accept the manuscript for publication.
/	Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.
	Reject the manuscript, as it is not suitable for publication.

SECOND ROUND

<u>Decision</u>	<u>Editor</u>	<u>Reviewer 1</u>	<u>Reviewer 2</u>
Accept the manuscript for publication.	/	/	/
Reconsider the manuscript after the authors have satisfactorily addressed and complied with the reviewers' comments and recommendations.			
Reject the manuscript, as it is not suitable for publication.			