International Journal of Biology Education Towards Sustainable Development,
Vol.5, No.2, 2025, pp. 113-122
e-ISSN 2809-5073. DOI. 10.52889/ijbetsd.v5i2.827

The Correlation Between Test and Non-Test Assessments in the Structure and Function of Plants Course

Submitted 26 December 2025, Revised 31 December 2025, Accepted 31 December 2025

Lilit Rusyati[1*], Riana Nurismawati[1], Andi Isni Pujirana[2]

[1]Science Education Program, Faculty of Mathematics and Science Education,
Universitas Pendidikan Indonesia, Bandung, Indonesia

[2]Department of Physics, Nelson Mandela University, Gqeberha, South Africa
Corresponding Email: *lilitrusyati@upi

Abstract

The problem in this case is the requirement to examine the similarity between the students' theoretical knowledge (determined by the test results) and their competencies and engagement (determined by the results of the non-test assessment). The purpose of this study is to examine the relation between the test results and the non-test results among students in the Plant Structure and Function class. The data collected in this study came from 131 prospective junior high school teachers in the field of sciences. The results of the analysis in this case show that the Test results are positively associated and strongly significant with the results of the Non-Test results in this study with the value of the correlation coefficient ($r = 0.357$) and significance ($p = 0.000$). In fact, there is no causal relationship between the results in the two tests but there is an associative relation. The results in this study can be considered as a source for the teachers in constructing a comprehensive evaluation system.

Keywords: Correlation, Test scores, Non-test scores, Learning assessment, Structure and function of plants.

INTRODUCTION

Current situation (state of the art) that is, formative and summative assessment are very essential in assessing levels of student understanding in different contexts. Formative assessment, short-answer questions, are also used to provide information on conceptual student understanding and thinking processes (Carter & Prevost, 2023; Carter & Prevost, 2018). Finally, summative assessment, final exam and practical assignments, is also broadly used to assess student outcomes (Arnold & Marshall, 2024). Correlation in assessments has also been studied in different assessments, process-based assessments (practical assignments) and outcome-based assessments (final exam), indicating different levels of association based on the subject or students (Leppink, 2020). Moreover, studies have investigated the relationships between student evaluations of courses and instructors and the factors that affect these evaluations (Reverter et al., 2020). It has been found that the context in which questions are posed and the nature of the questions can affect student responses and their understanding. For instance, the inclusion of scenarios in the questions can yield different levels of knowledge of the students in comparison to questions without scenarios (Carter & Prevost, 2023; Carter & Prevost, 2018).

There is indeed a gap in research concerning the integration of non-test assessments, especially in the affective and psychomotor domains of science education. Such limitations bring about incomprehensive assessment processes. (Novitasari et al., 2024) Although question order might impact student response and understanding, few studies have addressed this factor. Preliminary data indicate that question order may serve as preparation for students and thereby affect their performance, but to understand this fully, further research is required. (Carter & Prevost, 2018) There is a lack of studies on how different demographics of students influence their performance and perceptions in plant structure and function courses, such as major and class rank. Current research tends to focus on general trends rather than specific subgroups.

Novelty proposed contribution: Developing a more integrated assessment framework that includes test and non-test methods across the cognitive, affective, and psychomotor domains. This is most likely to assure a more holistic evaluation of student learning and address existing gaps in the assessment of students without tests. Some research on how different contexts and question sets influence student understanding in courses on plant structure and function may lead to more adequate assessments that reveal the students' knowledge and ways of thinking (Carter & Prevost, 2023; Carter & Prevost, 2018). Application of correlation analysis to detail areas of student difficulty. Further, targeted teaching strategies to address those weaknesses can be developed, and thus, permit tailoring of teaching methods for Structure and Function of Plants course to best help students learn.

METHOD

The quantitative method used in this study was a correlation design. The purpose of this design is to determine whether and to what extent two quantitative variables are related in a population (Creswell, 2014). By using a set score for each variable, researchers do not need to change the conditions of the research subjects to determine the strength and direction of the relationship. This study involved prospective junior high school science teachers taking the Structure and Function of Plants course. The study sample consisted of 131 students from five cohorts (2020–2024). Table 1 shows the distribution of participation per cohort.

Table 1. Research Participants

No	Prospective Science Teachers	Amount
1	Class of 2020	28
2	Class of 2021	21
3	Class of 2022	22
4	Class of 2023	30
5	Class of 2024	30
	Total	131

This study used a purposive sampling. Purposive sampling is a method in which researchers deliberately select people and locations to study a phenomenon because they can provide rich and relevant information about the research objectives (Creswell, 2014). To be included, an individual must meet two requirements: (1) have the status of a prospective junior high school science teacher; and (2) have a complete grade record for the course Plant Structure and Function. Any student from the 2020–2024 academic year who met these criteria was included. This was done so that the sample could be formed according to the research objectives and provide representative data that could be analyzed.

Data was collected from students' academic grade documents, which consist of two types of assessments:

1. Test: written evaluation results (mid-semester exams, final exams, and quizzes).
2. Non-Test: assessment results based on performance, projects, participation, assignments, and portfolios.

The IBM SPSS 25 statistical program was used to analyze the data. To examine the relationship between test and non-test scores, the Pearson Product Moment correlation test was used. A 99% confidence level ($\alpha = 0.01$) was used as the basis for interpreting significance.

RESULTS AND DISCUSSION

The results of the Pearson correlation analysis showed a positive and significant relationship between Test scores and Non-Test scores. This relationship was considered statistically significant with a significance value (p) of 0.000, which is lower than the alpha level of 0.01, and a correlation coefficient of 0.357 indicating a weak to moderate strength of relationship. Therefore, it can be concluded that Non-Test scores also tend to increase as Test scores increase. Although the relationship found between these variables is significant, this study found that there is no causal relationship between the two variables. The results of the correlation test with IBM SPSS Statistics 25 are presented in Table 2.

Table 2. Correlation Test Results

		Test	**Non-Test**
Test	Pearson Correlation	1	.357**
	Sig. (2-tailed)		.000
	N	131	131
Non-Test	Pearson Correlation	.357**	1
	Sig. (2-tailed)	.000	
	N	131	131

*** Correlation is significant at the 0.01 level (2-tailed).*

The results of the correlation analysis between the two variables are shown in Table 2. These variables are test and non-test assessment scores. The statistical test results can be explained as follows:

International Journal of Biology Education Towards Sustainable Development,
Vol.5, No.2, 2025, pp. 113-122
e-ISSN 2809-5073. DOI. 10.52889/ijbetsd.v5i2.827

1. A significant positive relationship was observed between test and non-test scores. The Pearson correlation coefficient was 0.357. Based on the general interpretation of correlation coefficients, this figure indicates a weak to moderate relationship. In other words, the tendency is that as the score on one variable increases, the score on the other variable also tends to increase.
2. With a significance value (Sig. 2-tailed) of 0.000, this relationship is considered highly statistically significant. This value is far below the general critical limit (α = 0.01). Consequently, the null hypothesis (H0), which indicates that there is no correlation between the two variables, is rejected. That this correlation is significant at the 99% confidence level (0.01 level) is also confirmed by the double asterisk (**) sign on the correlation value.
3. For this analysis, each variable had 131 respondents or data samples; this sample size was sufficient to conduct correlation tests and provide confidence in the results.
4. These results indicate that the abilities or performance measured by test and non-test instruments have a linear relationship. Although significant, the correlation coefficient is not very strong (0.357), indicating that the two evaluation categories are not entirely assessing the same thing. Each likely captures a different component or construct of the participant's abilities. In other words, tests and non-tests are not substitutes for each other in providing a picture of the participant. Visual pattern of scatter plot described in Figure 1.

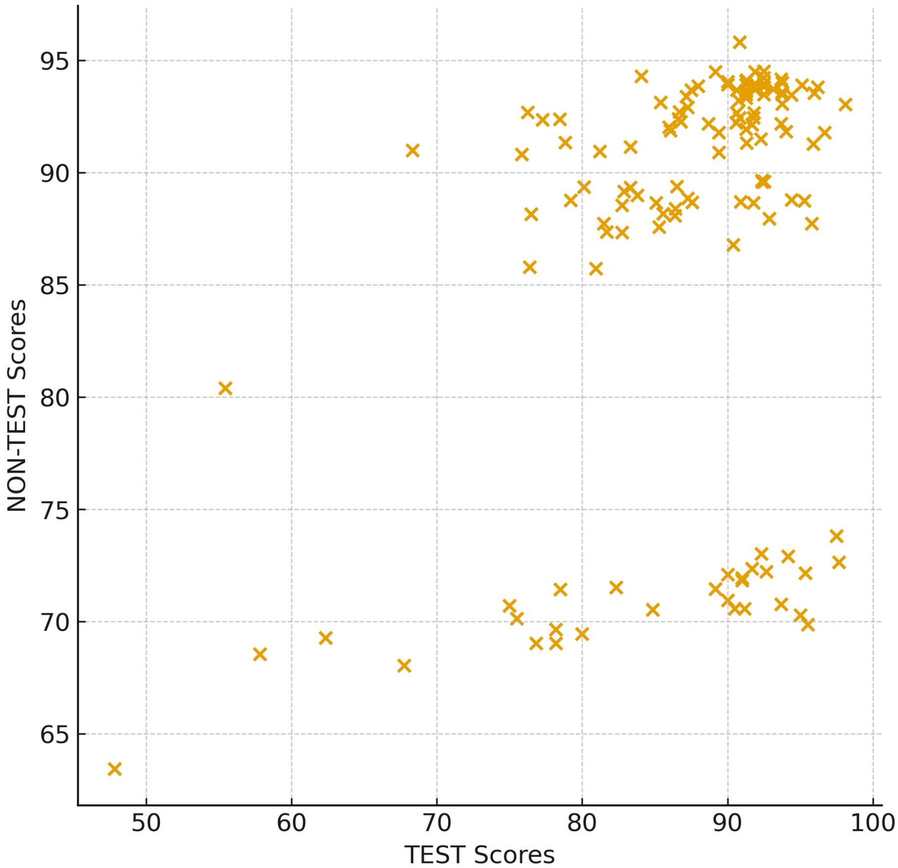

Figure 1. Scatter Plot of Test and Non-test

Figure 1 shows that the data points generally show a positive trend, meaning that the higher the test score, the higher the non-test score. This indicates a positive correlation, although the strength is not the same across the range of values. The group of values shown (*clustering*) that is:

1. High scores cluster (test range 85–100). Many points were clustered in the 85–100 test range and 88–95 non-test range. This indicates that the majority of students consistently performed high on both test and non-test assessments. This group is the most dense and therefore contributes the most to the overall correlation.
2. Medium scores cluster (test range 70–85). Within this range, non-test scores ranged from 88–92. The pattern tended to be more diffuse, indicating that some students had lower test scores but their non-test scores remained relatively stable.
3. Low score cluster (test < 70). Data points on the test 45–70 tend to fall within the non-test 63–72 range. In this group, the relationship remains positive, but the non-test variation is greater. This group indicates that when test scores drop significantly, non-test scores also drop.

There are a few points that deviate from the general trend, for example, tests are around 93, but non-tests drop into the 70s. Tests are very low (47–60), with non-tests also low (63–72). These outliers (deviant values) can be influenced by non-academic factors, can come from students with unstable performance, or can represent different assessment conditions between tests and non-tests. Visual patterns of scatter plot with the regression line shown in Figure 2.

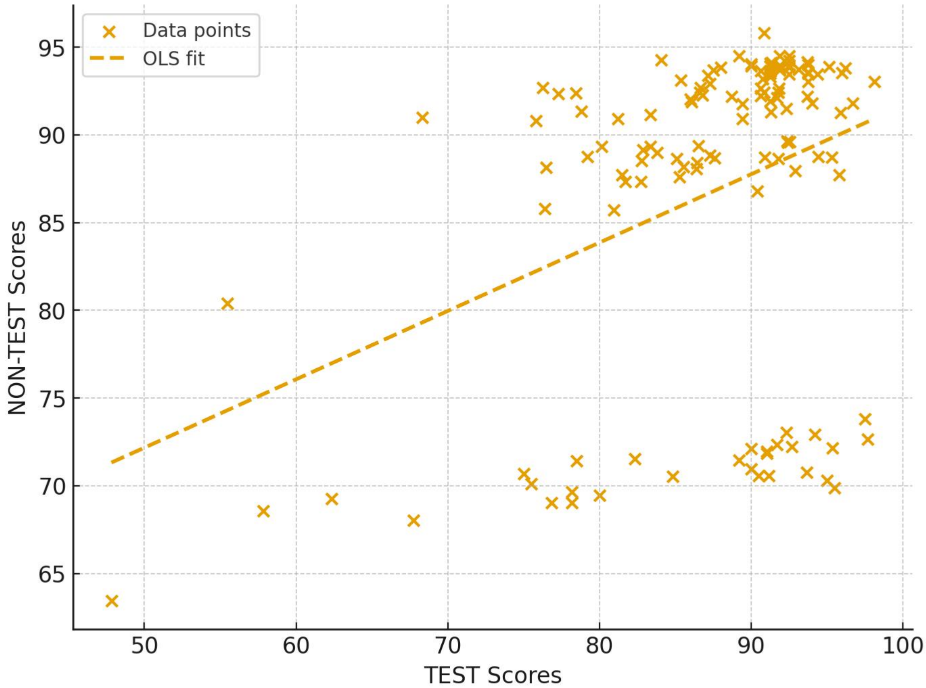

Figure 2. Test and Non-test Regression

Figure 2 shows a linear relationship between test scores (x-axis) and non-test scores (y-axis) for a number of students. Here is the interpretation:

1. Each orange dot represents a student, with horizontal position = test score and vertical position = non-test score. There are clusters of scores that appear to be clustered: many test scores are in the 85–95 range, some are in the 70–80 range, and some are in the 90–95 range. Outlier at low scores (e.g. test around 50–60).
2. The dashed line represents a linear regression that attempts to predict non-test scores from the test. This means the regression line has a positive slope, meaning that as test scores increase, non-test scores tend to increase. This relationship is linear on average, not an absolute relationship for each point. However, the line does not appear to follow closely in the high-score group; this indicates a relatively weak to moderate relationship.
3. The distribution of the points shows a moderate positive correlation. The points generally rise to the right, but are not tightly clustered around the line. This indicates that the test is

related to the non-test, but not the sole factor. This condition aligns with previous statistical results. Pearson's correlation coefficient value is 0.357(moderate correlation).

4. Visually, there are groups with test scores of 85–95 but non-test scores ranging from 88–96. There are groups with test scores of 85–95 but non-test scores of only 70–75. There are low-scoring groups of 50–70 with non-test scores of 63–80. This indicates the possibility of heterogeneity in the student groups, or that a simple linear regression model is insufficient to capture the true data pattern, or the possibility of moderator variables or different subpopulations.
5. There are several points far below the regression line (e.g., test 45–60), indicating students with low test and very low non-test scores. This outlier weakens the power of the regression model and contributes to nonlinearity.

Combining standardized-test and non-standardized-non-test-based assessments could have a considerable improvement in knowledge assessment, showing both interdependence and complementarity between the two. Non-standardized tests can provide far more accurate measures of students' learning or their ability to apply knowledge. This approach will allow the development of creativity and scientific thinking. Non-test assessments are a type of performance assessment that can serve as a helpful diagnostic tool in formative evaluation in order to gain insight into how the skills of students are developing and to support educators in improving the instruction of science. In a comparative study between project-based assessment as a form of non-test assessment and conventional assessment, the outcomes showed that project-based assessment significantly enhanced the scientific attitude and the science process skills of the students. These findings indicate that non-test assessments are effective in science education.

One such investigation revealed that the interactive oral assessments performed better than the traditional written exams, indicating that students did better on the final assessments and earned higher grades for the course. This suggests that non-test assessments may help students learn better. As an example, cognitive and non-cognitive assessments aim to measure students' understanding and application of knowledge; both often produce positive correlations. These include cognitive and non-cognitive assessments, which are meant to determine how a student understands and applies knowledge. Indeed, these mostly produce a positive correlation. In some cases, educator assessments may be subjected to biases or perceptions correlating with student test scores. For instance, educators may overestimate the performance of particular students, such as female students, in mathematics, compared to their test scores, which could lead to the notion of a correlation between the two types of assessment.

With high levels of motivation and effort, students are capable of affecting their test performance as well as performance in non-test assessments. High motivation and effort have positive outcomes in different assessment areas, leading to positive correlations (Reeve & Lam, 2007). In order to ensure this positive correlation is maintained, it is important to discuss issues regarding test as well as non-test assessment validity and fairness. This is because it is essential that these assessments are representative as well as free from bias (Marcenaro-Gutierrez & Vignoles, 2015; Lane, 2020). The results of diagnosis in education are used to make specific attempts to correct misconceptions, especially when carried out in contextual forms. The concept of test performance is also essential as it is recommended to be used as pre-instructional diagnosis (Amelia et al., 2023). Apart from similarity in test scores regarding core skills, similarity in assessment forms in relation to educational outcomes, as well as influences due to high student motivation and effort, it is also possible to ensure positive correlations are maintained in test and non-test performance through implementation of a wide or broader approach.

CONCLUSION

Based on the results of the data analysis, it can be concluded that for prospective junior high school science teachers, there is a statistically significant positive relationship between test-based assessment and non-test assessment in the Plant Structure and Function course. Although the relationship is considered weak to moderate, the Pearson correlation coefficient of 0.357 and a significance level of 0.000 indicate that the relationship is significant. The results indicate that an increase in one assessment category tends to be followed by an increase in the other category. However, it should be emphasized that these results only indicate a relationship between the two assessment variables, not a causal relationship.

SUGGESTIONS

Some recommendations have been put forward based on the findings from the research. It is recommended that the instructor goes ahead and uses a balanced process that entails both test and non-test assessment. This will pave way for them to carry out a deeper analysis that entails the optimization process for non-test assessments. For them to have a clearer understanding regarding test dynamics, they should consider increasing their sample size. They can also have a deeper analysis of other variables.

ACKNOWLEDGEMENT

The authors would like to thank the 131 volunteer prospective junior high school science teacher students in this research, especially in the Structure and Function of Plants course. Their grades were used as the main basis for the correlational analysis between the test and non-test assessments. Without their active participation, this research would not have been possible.

REFERENCES

Amelia, M., Thomas, M. M., & Rusyati, L. (2023). Analysis of students misconceptions about reproduction in plants using a four-tier diagnostic test. *Jurnal Pendidikan Matematika dan IPA*, *15*(2), 287-301. https://doi.org/10.26418/jpmipa.v15i2.68888

Arnold, M. A., & Marshall, M. L. (2024). Student attendance, study activities, major, and midterm grades as predictors of final course grades of undergraduate and graduate plant materials courses during the eras before and after COVID-19. *HortTechnology*, *34*(6), 687-695. https://doi.org/10.21273/horttech05503-24

Carter, K. P., & Prevost, L. B. (2023). Formative assessment and student understanding of structure-function. *Advances in Physiology Education*, *47*(3), 615-624. https://doi.org/10.1152/advan.00215.2022

Carter, K. P., & Prevost, L. B. (2018). Question order and student understanding of structure and function. *Advances in physiology education*, *42*(4), 576-585. https://doi.org/10.1152/advan.00182.2017

Creswell, J. W. (2014). *Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. New York: Pearson Education.

Davey, S. K., Birbeck, D., Nallaya, S., Sallows, G., & Della Vedova, C. B. (2025). Utilising one-on-one interactive oral assessments as the major final assessment within a bioscience course. *Assessment & Evaluation in Higher Education*, *5*(7), 1-18. https://doi.org/10.1080/02602938.2025.2502577

Davidovitch, N., Gerkerova, A., & Kyselyova, O. (2024). Comparative analysis of knowledge control and evaluation methods in higher education. *Education Sciences*, *14*(5), 505-515. https://doi.org/10.3390/educsci14050505

Doxey, J. S., Waliczek, T. M., & Zajicek, J. M. (2009). The impact of interior plants in university classrooms on student course performance and on student perceptions of the course and instructor. *HortScience*, *44*(2), 384-391. https://doi.org/10.21273/hortsci.44.2.384

Hongling, Y. E., Qingsheng, Y. A. N. G., Jiazheng, D. U., Xia, L. I. U., & Junjun, S. H. A. N. G. (2022). Exploration and practice of non-standardized assessment model in teaching of fundamental mechanics. *Mechanics in Engineering*, *44*(5), 1206-1212. https://doi: 10.6052/1000-0879-22-105

Kruit, P., Oostdam, R., van den Berg, E., & Schuitema, J. (2020). Performance assessment as a diagnostic tool for science teachers. *Research in Science Education*, *50*(3), 1093-1117. https://doi.org/10.1007/s11165-018-9724-9

Lane, S. (2020). Test-based accountability systems: The importance of paying attention to consequences. *ETS Research Report Series*, *2020*(1), 1-22. https://doi.org/10.1002/ets2.12283

Lee, J., McArthur, W., & Ellis, N. J. (2019). NAPLAN versus in-school assessment: how similar or different are students' results?. *Curriculum and Teaching*, *34*(2), 5-25. https://doi.org/10.7459/ct/34.2.02

Leppink, J. (2020). Assessment programs and their components: a network approach. *Scientia Medica*, *30*(1), 27-37. https://doi.org/10.15448/1980-6108.2020.1.37124

Marcenaro-Gutierrez, O., & Vignoles, A. (2015). A comparison of teacher and test-based assessment for Spanish primary and secondary students. *Educational Research*, *57*(1), 1-21. https://doi.org/10.1080/00131881.2014.983720

Novitasari, A., Mujianti, A., Solviana, M. D., Supriyadi, S., & Kesuma, A. J. (2024). Analysis of difficulties faced by high school biology teachers in Pringsewu regency in implementing non-test assessments. *E3S Web of Conferences*, *482*(1), 1-20. https://doi.org/10.1051/e3sconf/202448204021

Parmiti, D. P., Rediani, N. N., Antara, I. G. W. S., & Jayadiningrat, M. G. (2021). The effectiveness of local culture-integrated science learning through project-based assessment on scientific attitudes and science process skills of elementary school students. *Jurnal Pendidikan IPA Indonesia*, *10*(3), 439-446. https://doi.org/10.15294/jpii.v10i3.31301

Reeve, C. L., & Lam, H. (2007). Consideration of g as a common antecedent for cognitive ability test performance, test motivation, and perceived fairness. *Intelligence*, *35*(4), 347-358. https://doi.org/10.1016/j.intell.2006.08.006

Reverter, A., Martinez, C., Currey, P., van Bommel, S., & Hudson, N. J. (2020). Unravelling student evaluations of courses and teachers. *Cogent Education*, *7*(1), 1-20. https://doi.org/10.1080/2331186x.2020.1771830