AJSTME, Volume. 12, Issue 1. 211-219, January, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)

Error Variance and Score Reliability of Facets in an Economics Multiple Choice-Test in Nsukka Education Zone of Enugu State, Nigeria

Ossai Elizabeth Ngozika
Department of Science Education, University of Nigeria, Nsukka
Email: elizabethngozika@gmail.com

Abstract

This study applied Generalizability Theory (G-Theory) to examine sources of measurement error and evaluate the reliability of scores from a 50-item Economics multiple-choice test taken by senior secondary school students. The study was guided by three research questions and one hypothesis and employed a fully crossed S × I × O design (Students × Items × Occasions). The population comprised 3,869 SS2 Economics students from 60 government-owned secondary schools in Nsukka Education Zone, Enugu State, Nigeria. Using proportionate stratified random sampling, 350 students were selected. The instrument was developed by the researcher and validated by experts, with reliability (KR-20) of 0.88. Variance estimates from EduG software answered the research questions, while a t-test tested the hypothesis at 0.05 significance level. Findings showed that the largest variance arose from the S × I × O interaction. Increasing test items improved the G-coefficient. The study recommends adopting G-Theory to enhance score reliability and interpretation.

Keywords: Economics, Error variance, Facets, Generalizability Theory, Score reliability.

Introduction

The identification of sources of measurement error and the estimation of score reliability are fundamental to the accurate interpretation of assessment outcomes in any academic discipline, including Economics. It has been defined as the study of how individuals, groups, or nations allocate limited resources to produce, exchange, and consume goods and services (American Economic Association, 2023; Smith & Okoye, 2024). Resources include land, labor, and capital goods such as machinery, vehicles, and buildings. Robbins (1935) described Economics as the study of human behavior in relation to ends and scarce means with alternative uses. This highlights the fact that human wants are numerous, resources are limited, and scarcity requires individuals to make choices among competing alternatives. From these perspectives, Economics examines human behavior in relation to scarce resources. According to the Nigerian Educational Research and Development Council (NERDC, 2023; Adewale & Uche, 2025), the objectives of teaching Economics in secondary schools include understanding economic principles for sound analysis, contributing meaningfully to discussions on economic reforms, understanding economic institutions and public policies, developing analytical and decision making skills, participating in national economic development through entrepreneurship and financial markets, understanding Nigeria's role in the global economy, and appreciating the challenges faced by developing countries in advancing economically. Obemeata (1991), supported by Nnadi and Okeke (2023), emphasized that Economics education equips citizens and leaders to understand fundamental concepts, appreciate economic realities, and seek improvements for social and economic wellbeing. Achieving these objectives requires effective learning, which is closely tied to proper assessment. Learning involves the transfer of knowledge, skills, and habits from one generation to the next, and assessment plays a central role in monitoring students' progress and providing feedback that enhances learning (Black & Wiliam, 2023; Johnson & Adeyemi, 2024). Consequently, appropriate and accurate assessment is critical for guiding students' learning in Economics. Educational assessment is a systematic process of gathering information from multiple sources to understand, describe, and improve learning outcomes (Popham, 2023). One common assessment tool is the test.

A test is a structured instrument comprising items designed to measure the extent to which an individual possesses specific traits or skills (Wang, 2022; Ali & Ahmed, 2025). While many instruments are used in education, not all instruments qualify as tests. In cognitive assessment, tests typically take two forms: essay tests and objective tests (Brookhart

& Nitko, 2022). Essay tests allow students to respond in their own words, whereas objective tests, such as multiple-choice tests, require students to select correct answers from a set of alternatives. Multiple choice tests are widely used in Economics assessment because they provide standardized, quantifiable measures of student performance (Chukwuemeka, 2025; Idris, 2024).

Despite the structured use of assessments in Economics, evidence from the West African Senior School Certificate Examination (WAEC, 2021–2025) indicates fluctuations and, in some cases, declines in student performance. Chief Examiners' reports from the West African Examinations Council (WAEC, 2021–2025) highlighted difficulties with graphical analysis, application of economic concepts, and the use of correct terminology. From 2021 to 2025, these trends persisted: the overall WASSCE pass rate increased to 81.7 % in 2021 but declined to 76.36 % in 2022, rose slightly to 79.81 % in 2023, and dropped sharply to 72.12 % in 2024. The 2025 examination initially recorded a low pass rate of 38.32 %, largely attributed to the introduction of stricter anti-malpractice measures, including paper serialization and computer-based testing. Subsequent corrections raised the overall pass rate to 62.9 %, with Economics-specific credit achievement around 58 %. The persistence of similar weaknesses across multiple examination years suggests that performance variations may not be solely attributable to student ability, but may also be influenced by measurement-related factors such as facets, including items, raters, occasions, and test forms. These errors can affect the consistency and stability of scores, which is referred to as reliability (Cohen, Swerdlik & Sturman, 2023; Akande & Bello, 2025). Traditional methods of reliability estimation under Classical Test Theory (CTT) consider only a single undifferentiated error term, limiting the ability to examine multiple sources of error simultaneously (Zhang & Liu, 2023; Adeyemi & Ojugo, 2024).

CTT treats error as random and is unable to distinguish systematic from random error, nor can it adequately provide information for absolute decisions about individual performance (Brown & Smith, 2023; Mohsen & Brennan, 2023). To overcome these limitations, Generalizability Theory (G Theory) provides a more comprehensive framework for evaluating reliability. G Theory extends CTT by allowing simultaneous estimation of multiple sources of error and their interactions through analysis of variance procedures (Song, Lee & Jiao, 2025; Smith, 2024). It identifies variance attributable to facets such as students, items, occasions, and raters, and evaluates their contribution to measurement error (Sanz-Fernández et al., 2024; Özder & Takır, 2024). Facets can be fixed or random, crossed or nested, depending on the study design. A generalizability study (G Study) estimates the variance associated with each facet and their interactions, while a decision study (D Study) uses these estimates to optimize test design, minimize error, and maximize score reliability for both relative and absolute decisions (GeneralizIT, 2024; Child & Medvedev, 2024). By applying G-Theory to Economics multiple-choice tests, researchers can quantify how facet errors such as items and occasions affect score reliability, disentangle multiple sources of error, and enhance the accuracy and fairness of assessment outcomes. This study is therefore designed to apply Generalizability Theory to estimate error variance and score reliability in Economics multiple choice tests, with a focus on facets such as items, occasions, and their interactions.

Conceptual Framework

The conceptual framework of this study is anchored on Generalizability Theory (G-theory), with Classical Test Theory (CTT) serving as a conceptual contrast. In CTT, an observed test score (X) is considered the sum of a true score (T) and a single, undifferentiated error (E) (Brown, 2022). CTT provides reliability estimates based on one source of error at a time, such as test-retest reliability, parallel forms, or inter-rater agreement. While useful, CTT cannot simultaneously account for multiple sources of error, interactions between them, or distinguish between random and systematic error (Bachman, 2023; Mohsen & Brennan, 2022). G-theory extends CTT by allowing simultaneous estimation of multiple sources of error, called facets, such as items and occasions, and their interactions (Brennan, 2022; Kim & Wilson, 2023).

Statement of the Problem

It is an observed fact that teachers assess students' performance in every examination. Teachers use instrument such as multiple-choice test to assess the knowledge, skills and abilities of the students in Economics at senior secondary schools. Despite the adequate assessment of students' learning objectives in Economics, there are impressions of a decline in students' performance in Economics in Nigeria West African Senior School Examination (WASSCE). Regrettably, obtainable evidences have shown that there are poor academic performances in Economics by the West African Examinations Council Chief Examiners' Reports of

2021-2025. The reasons for the poor academic performances of the students could be attributed to sources of error or facets such as items, invigilators, raters, occasions, and test forms among others. Since, the scores attained by our secondary school students in Economics multiple choice test are influenced by multiple sources of error, the use of classical test theory to estimate the errors simultaneously with their interaction effect are not available. The inability to analyze more than one source of error at a time with their interaction effect limits classical test theory as a psychometric technique. CTT treats error as random and cannot be used to distinguish systematic measurement error from random measurement error; CTT considers only the relative decision where individual's score is compared to a reference group and did not consider absolute decisions. These limitations of CTT therefore necessitated the need for a better measurement theory such as generalizability theory. A significant feature of the G-theory is based on the simultaneous analysis of multiple sources of error variances. Equally important, generalizability theory considers interaction among the sources of measurement error. G-theory also distinguishes systematic measurement error from random measurement error; these G-theory procedures also help in making relative and absolute decisions which allow individual's score to be compared to a reference group and which also focus on the level of performance concerning students. Hence, it is pertinent to apply generalizability theory for the estimation of error variance and score reliability of facets in Economics multiple choice test. Based on this premise, lies the rational for this study. The study applied Generalizability Theory to estimate error variance and score reliability associated with multiple measurement facets in Economics multiple choice tests.

Specifically, the study determined the differences in the magnitude of error variances due to students, items, occasions, and their interactions; the relative and absolute error variances associated with the facets students, items, occasions, and their interactions in making relative and absolute decisions about students; and the differences in the generalizability coefficients of an Economics multiple choice test by increasing the number of items and occasions.

Research Questions

The following research questions guided the study;

1. What are the differences in the magnitude of error variances due to students, items, occasions and their interactions?
2. What are the relative and absolute error variances associated with the facets: students, items, occasions and their interactions in making relative and absolute decisions about the students?
3. What are the differences in the generalizability coefficients of Economics multiple choice test by increasing the number of items and occasions?

Hypothesis

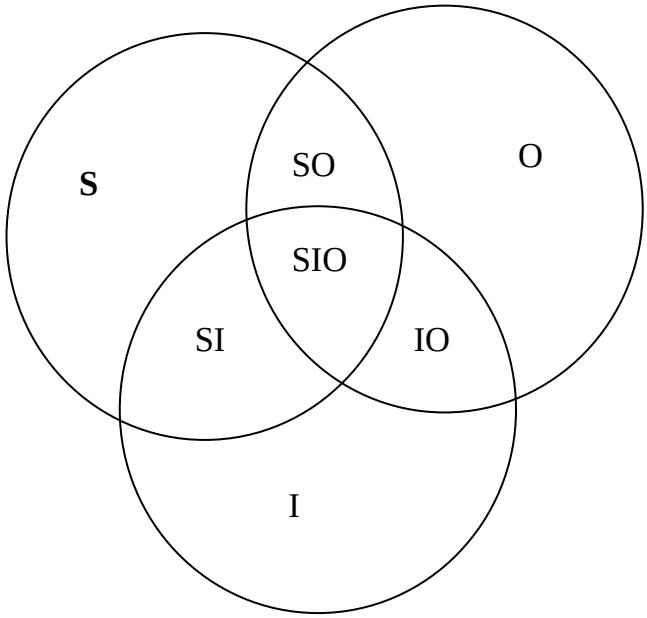
The following null hypothesis was formulated and tested at the 0.05 level of significance:

H_{01}: There are no significant differences in the generalizability coefficients of Economics multiple choice test by increasing the number of items and occasions.

Method

A fully crossed S × I × O G study design was adopted for this study. According to Shavelson and Webb (2005), a fully crossed design is one in which every condition of each facet is observed for every condition of the other facets. In this design, students (S), items (I), and occasions (O) are crossed such that all students respond to the same set of items across the same testing occasions. The design was considered appropriate for this study because it provides maximal information about the components of variation in observed scores attributable to students, items, occasions, and their interactions. The design is illustrated in the Venn diagram below, where each circle represents one of the facets: students (S), items (I), and occasions (O). The areas of overlap represent the interactions among the facets, highlighting how variations in one facet may combine with variations in another to influence observed scores.

AJSTME, Volume. 12, Issue 1. 211-219, January, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)



This symbolizes the facets; students (S), item (I), and occasions (O) whereas their intersections represent the interactions of the facets.

The population of the study comprised 3,869 Senior Secondary School Two Economics students in the 60-government owned secondary schools in Nsukka Education Zone. Specifically, 2,618 students were from Nsukka Local Government Area, 840 from Igbo Etiti Local Government Area, and 411 from Uzo Uwani Local Government Area.

The sample size of 350 SS2 Economics students was obtained through a multistage sampling procedure involving proportionate stratified allocation of students from the total population of 3,869 across the three Local Government Areas.

In the first stage, simple random sampling technique was used to select schools from the three Local Government Areas. A sampling fraction of 20 percent was applied to select 12 schools out of the 60 schools in the zone. Specifically, six schools were selected from the 31 schools in Nsukka Local Government Area, three schools from the 15 schools in Igbo Etiti Local Government Area, and three schools from the 14 schools in Uzo Uwani Local Government Area.

In the second stage, proportionate stratified random sampling technique was adopted to select students from the sampled schools in proportion to the population of each Local Government Area. Consequently, 250 students were selected from Nsukka Local Government Area out of its population of 2,618 students, 54 students were selected from Igbo Etiti Local Government Area out of its population of 840 students, and 46 students were selected from Uzo Uwani Local Government Area out of its population of 411 students, making a total of 350 students. This procedure ensured that the sample reflected the population distribution across the zone and enhanced its representativeness. This is in line with Nworgu's (2015) recommendation that proportionate stratified random sampling ensures greater representativeness of the sample relative to the population and guarantees that minority constituents of the population are represented in the sample.

The instrument for data collection was Economics Multiple Choice Test developed by the researcher. The test was constructed based on selected Senior Secondary School Two Economics topics including Demand and Supply, Financial Institutions, Public Finance, Labour Force, Alternative Economic Systems, Theory of Cost, and Inflation. The instrument consisted of 50 multiple choice items with four response options A to D. Each correct response attracted one mark, while incorrect responses attracted zero. The items were developed in line with the six levels of the cognitive domain of Wilson (2022) Revised Bloom Taxonomy of Education and were constructed using a table of specification to ensure adequate coverage of content and objectives. A scoring key was also prepared to guide scoring. These topics were selected because they represent core areas of the Senior Secondary School Two Economics curriculum and provide a balanced coverage of fundamental economic concepts required for assessing students' cognitive achievement

To ascertain the validity of the instrument, face and content validation were carried out. The instrument was given to three experts, two from Science Education Measurement and Evaluation Unit and one from Social Science Education Economics Unit, Faculty of Education, University of Nigeria, Nsukka. The experts examined the clarity, relevance, and adequacy of the items in addressing the objectives of the study. Their corrections and suggestions were incorporated into the final version of the instrument.

The reliability of the instrument was established through trial testing conducted on 25 Senior Secondary School Two Economics students in Community Secondary School Nimbo, Uzo Uwani Local Government Area, who were part of the study population but not included in the main study. The responses were analyzed using Kuder Richardson Formula 20 to determine the internal consistency reliability of the instrument. A reliability coefficient of 0.88 was obtained, indicating high reliability. The choice of Kuder Richardson 20 was appropriate because the items were dichotomously scored and administered once.

For data collection, the researcher visited the sampled schools and administered the instrument to the students with the assistance of the Economics teachers. The test was conducted under proper examination conditions and lasted for one hour. The instrument was administered twice at an interval of two weeks to accommodate the occasion's facet in the fully crossed design. The completed scripts were retrieved immediately for scoring and analysis.

The research questions were answered using variance estimate in Edu-G software programe while the hypothesis were tested using t-test at 0.05 level of significance. The use of variance estimates enabled the estimation of error components associated with students, items, occasions, and their interaction. Decisions were based on the magnitude and comparison of the estimates. Larger variance components and error variances indicated greater contribution to measurement error, while smaller values indicated greater measurement precision. Higher generalizability and dependability coefficients indicated more reliable and dependable test scores, and any increase in these coefficients with increased items and occasions was considered an improvement in measurement quality. For the hypothesis, the decision rule was: reject the null hypothesis if $p \leq 0.05$, and do not reject the null hypothesis if $p > 0.05$.

Result

Table 1: Estimated variance components in a Generalizability study of Economics multiple choice test (p × i × o design).

Source of Variance	SS	Df	MS	Symbole of source of variance	Estimated Variance Component	% of Total variability
Students (S)	2998.25323	349	7.76750	σ^2_s	0.07516	26.6
Items (I)	122.78359	49	2.50579	σ^2_i	0.00330	1.2
Occasions (O)	0.00747	1	0.00747	σ^2_o	-0.00001	0.0
S x I	2508.96641	18914	0.13265	σ^2_{si}	-0.03429	0.0
S x O	123.54253	386	0.32006	σ^2_{so}	0.00238	0.8
I x O	0.88013	49	0.01796	σ^2_{io}	-0.00047	0.0
S x I x O	3806.06987	18914	0.20123	$\sigma^2_{sio,e}$	0.20123	71.3
Total	**9560.50323**	**38663**				**100%**

Keys: S = students, I = items, O = occasions, SI= students by items, IO = items by occasions and SIO, e = Students by items by occasions (interactions).

The result in Table 1 shows that there are seven variance components in G study conducted with s x i x o design. The variance components of the residual had the highest error variance accounted for 71.3% of the total variance. This was followed by the students which accounted for 26.6% of the

total variance. The third variance component was attributed to the items with the total variance of 1.2%. The fourth variance component was the students by occasions with the total variance of 0.8%.

Finally, occasions, students by items and items by occasions had 0.0% respectively, meaning that there was absent of error variance amongst them.

Table 2: G-study Estimate of Facets and their interaction on the relative and absolute decisions.

Source of variance	Differ-entiation variance	Relative error variance	% Relative	Absolute error variance	% Absolute
S	0.51760	.....	…..	.....	
I	.....	.....	….	0.02000	2.0
O	.....	.....	…..	(0.0000)	0.0
SI	.....	0.01196	37.1	0.01196	36.4
SO	….	(0.00000)	0.0	(0.00000)	0.0
IO	.....	.....	…..	(0.00000)	0.0
SIO, e	.....	0.02013	62.9	0.02013	61.6
Sum of Variances	0.51760 0.71944	0.03209	100%	0.05209	100%
Standard Deviation		Relative SE: 0.17913		Absolute SE: 0.22823	
Coef_G relative	0.78				
Coef_G absolute	0.74				

Keys: S = students, I = items, O = occasions, SI= students by items, IO = items by occasions and SIO, e = Students by items by occasions (interactions).

Result in Table 2 reveals the relative and absolute error variance associated to the facets and their interactions. The result indicated that items comprised an absolute error variance of 0.02. Student-by-item interaction had same value for both relative and absolute error variance, which recorded 0.01. Absolute error variance of 0.0 was attained for student by occasion interaction. Residual interaction recorded a relative and absolute error variance of 0.020 each. The two generalizability coefficients associated with making both relative and absolute decision were 0.78 and 0.74 respectively. This suggests that the measurement precision with the numbers of students, items and occasions certainly was not satisfactory.

Table 3: Estimated Generalizability coefficients for Economics multiple choice test, by increasing the number of items and occasions.

	Primary Conditions	**Optimization**				
Number of Items	**50**	**51**	**52**	**53**	**54**	**55**
Generalibility Coefficient (G or $E\rho^2$)	0.78	0.82	0.86	0.90	0.92	0.96
Number of Occasions	**2**	**3**	**4**	**5**	**6**	7
Generalibility Coefficient (G or $E\rho^2$)	0.78	0.80	0.82	0.85	0.87	0.90

Keys: S = students, I = items, O = occasions, SI= students by items, IO = items by occasions and SIO, e = Students by items by occasions (interactions).

Result in table 3 reveals the D-study and generalizability coefficients ($E\rho^2$) by increasing the number of items and occasions. The result indicated that increasing the number of Items had G-coefficients of 0.78, 0.82, 0.86, 0.90, 0.92 and 0.96 as the items increased from 50-55 respectively while occasions had G-coefficient of 0.78, 0.80, 0.82, 0.85, 0.87 and 0.90 as the occasions increased from 2-7 respectively. The result implies that increasing the number of items generates better generalizability coefficients than increasing the number of occasions. However, increasing the conditions of each facet increases the generalizability-coefficients.

H_{01}: There are no significant differences in the generalizability coefficients of Economics multiple choice test by increasing the number of items and occasions.

Table 4: t-test analysis of the influence of generalizability coefficients for Economics multiple choice test by increasing the number of items and occasions at 0.05 level of significant

Conditions	N	Mean	SD	df	t-value	P
Items	6	.87	.07	10	1.118	.290
Occasions	6	.84	.05			

The result in Table 4 shows independent sample t-test, assuming equal variance. From the result the t-value of 1.118 and probability value of .290 was obtained. Since the probability value of .290 is greater than the 0.05 level of significance, thus the null hypothesis was accepted. Therefore, there was no significant difference in the generalizability coefficients of Economics multiple choice test by increasing the number of conditions in each facets.

Discussion

The findings of the study revealed that the largest source of variation in the Economics multiple-choice test was the student–item–occasion interaction (residual variance), followed by students, while items and occasions contributed minimally. This may be due to inconsistent response behaviour by students across items and testing occasions, differences in attention, and variations in item interpretation. The substantial student variance may be due to differences in ability and prior knowledge in Economics, while the low item and occasion variances may be because the items were relatively similar in difficulty and testing conditions were stable. This agrees with Ikeh (2023) and Egbulefu (2022), who reported that residual variance and student differences are the major sources of variation in achievement tests. However, it disagrees with Shavelson and Webb (2023), who emphasized that improved measurement designs and better facet control should reduce residual error and improve score precision.

The findings further showed that student-by-item and student-by-occasion interactions contributed to both relative and absolute error variance, while item-by-occasion interaction had no contribution. This may be because students perform inconsistently across items and occasions due to differences in understanding, engagement, and testing conditions. This agrees with Shavelson and Webb (2023) and Ikeh (2023), who noted that person–facet interactions are major sources of measurement error in Generalizability Theory studies. However, it disagrees with Lee (2019), who found that item–occasion interactions can become important when learning or fatigue effects occur across repeated testing situations.

The findings also showed that the generalizability coefficients for both relative (0.78) and absolute (0.74) decisions were below the acceptable benchmark of 0.80. This may be due to insufficient number of items and occasions, which reduced measurement stability and increased error variance. This agrees with Egbulefu (2022), who reported that low facet sampling reduces reliability and generalizability. However, it disagrees with Shavelson and Webb (2023), who suggested that well-constructed measurement designs with adequate optimization should achieve acceptable reliability levels above 0.80.

The study further revealed that increasing the number of items produced higher generalizability coefficients than increasing the number of occasions. This may be because increasing items improves content coverage and better representation of the Economics achievement domain, while increasing occasions only repeats measurement under similar conditions. This agrees with Ikeh (2023) and Kim and Park (2020), who found that increasing item sampling improves reliability more effectively than increasing occasions. However, it disagrees with Shavelson and Webb (2023), who noted that increasing occasions can improve reliability when temporal variation in performance is substantial.

Finally, the hypothesis test revealed that there was no significant difference in the generalizability coefficients when increasing the number of items and occasions ($p > 0.05$). This may be because the improvement in reliability from increasing either facet was not strong enough to produce a statistically significant effect due to dominant residual error. This agrees with Ikeh (2023) and Egbulefu (2022), who reported that increases in measurement facets do not always lead to statistically significant improvements in generalizability coefficients.

Conclusion

This study concludes that the main source of variation in the Economics multiple-choice test was the student–item–occasion interaction, followed by differences among students, while items and occasions contributed minimally. This indicates that inconsistencies in students' responses largely influenced measurement error, whereas item and occasion effects were relatively stable. The generalizability coefficients obtained were below the acceptable level, showing that the test had limited measurement precision. However, increasing the number of items improved reliability more than increasing occasions. Overall, the study concludes that the test requires improvement in item sampling to enhance its reliability and generalizability for accurate assessment of students' performance.

Recommendations

Based on the findings of this study, the following recommendations are made:

1. Economics teachers and test developers should add more items to multiple-choice tests to boost reliability and generalizability.
2. Teachers should undergo training to design assessments that reduce error and improve measurement accuracy.
3. Future assessments should account for multiple facets, such as items and occasions, to enhance test design and fairness.

References

Adewale, T., & Uche, O. (2025). *Economics education and student performance in Nigerian secondary schools*. Enugu: Academic Press.

Akande, O., & Bello, S. (2025). *Reliability and validity in educational assessment: A Nigerian perspective*. Lagos: Education Research Publications.

Ali, M., & Ahmed, H. (2025). *Test construction and psychometrics in secondary education*. Cairo: Learning Insights.

American Economic Association. (2023). *Principles of economics*. Washington, DC: AEA Publishing.

Bachman, L. F. (2023). *Statistical methods for language testing*. New York, NY: Routledge.

Black, P., & Wiliam, D. (2023). *Assessment and learning in schools: Contemporary perspectives*. London: Sage.

Brennan, R. L. (2022). *Generalizability theory*. New York, NY: Springer.

Brookhart, S., & Nitko, A. (2022). *Educational assessment of students* (8th ed.). Boston, MA: Pearson.

Brown, J., & Smith, K. (2023). *Classical test theory and measurement errors in education*. New York, NY: Routledge.

Chukwuemeka, E. (2025). *Multiple-choice tests in Nigerian secondary schools: Issues and prospects*. Nsukka: University of Nigeria Press.

Child, D., & Medvedev, A. (2024). *Decision studies in educational measurement*. London: Routledge.

Cohen, R., Swerdlik, M., & Sturman, E. (2023). *Psychological testing and assessment* (10th ed.). New York, NY: McGraw-Hill.

Egbulefu, P. (2022). *Sources of error in multiple-choice testing: A Nigerian study*. Enugu: Nsukka Educational Press.

GeneralizIT. (2024). *Generalizability theory software manual*. London: GeneralizIT Press.

Hossein, M. (2022). *Standards of reliability in educational assessment*. New York, NY: Academic Press.

Idris, H. (2024). *Economics assessment in secondary schools: A Nigerian context*. Lagos: University Press.

Ikeh, F. E., Ani, M. F., Kalu, I. A., Iketeaku, R. I., Eze, B. A., Madu, B. C., & Ene, C. U. (2021). *Application of generalizability theory in estimating score dependability of economics essay test. International Journal of Mechanical and Production Engineering Research and Development, 11*(3), 221–228.

Kim, S., & Park, M. Y. (2020). Applications of generalizability theory in educational assessment: A multivariate approach. *Educational and Psychological Measurement*, 80(4), 789–805.

Kim, S., & Wilson, M. (2023). *Advanced applications of generalizability theory*. London: Routledge.

Lee, Y. W. (2019). Applications of generalizability theory in educational assessment.

Educational Measurement: Issues and Practice, 38(2), 50–58.

Mohsen, A., & Brennan, R. L. (2023). *Measurement errors and classical test theory*. New York, NY: Springer.

NERDC. (2023). *Curriculum standards for Economics in Nigerian secondary schools*. Abuja: Nigerian Educational Research and Development Council.

Nnadi, C., & Okeke, F. (2023). *Economics education for sustainable development*. Enugu: Nsukka Academic Press.

Nworgu, B. (2015). *Educational research: Basic issues and methodology* (3rd ed.). Nsukka: University Trust Publishers.

Obemeata, J. (1991). *The teaching of Economics in Nigerian secondary schools*. Lagos: Macmillan Nigeria.

Özder, M., & Takır, A. (2024). *Generalizability theory and psychometric applications*. Istanbul: Academic Press.

Popham, W. J. (2023). *Classroom assessment: Principles and practice* (9th ed.). Boston, MA: Pearson.

Robbins, L. (1935). *An essay on the nature and significance of economic science*. London: Macmillan.

Sanz-Fernández, L., Pérez, R., & González, M. (2024). *Modern applications of G-theory in education*. Madrid: Springer.

Shavelson, R. J., & Webb, N. M. (2023). *Generalizability theory for educational measurement* (2nd ed.). New York, NY: Springer.

Smith, J. (2024). *Measurement and evaluation in education*. London: Routledge.

Smith, J., & Okoye, A. (2024). *Economics principles and applications*. Enugu: Academic Press.

Song, L., Lee, H., & Jiao, H. (2025). *Variance components in educational assessment: A G-theory approach*. Singapore: Springer.

Wang, Y. (2022). *Principles of educational testing and measurement*. Beijing: Education Press.

Wilson, L. (2022). *Revised Bloom's taxonomy: Cognitive domain for educational assessment*. New York, NY: Routledge.

WAEC. (2021). *Chief examiners' reports: Economics*. Lagos: West African Examinations Council.

WAEC. (2022). *Chief examiners' reports: Economics*. Lagos: West African Examinations Council.

WAEC. (2023). *Chief examiners' reports: Economics*. Lagos: West African Examinations Council.

WAEC. (2024). *Chief examiners' reports: Economics*. Lagos: West African Examinations Council.

WAEC. (2025). *Chief examiners' reports: Economics*. Lagos: West African Examinations Council.

Zhang, Y., & Liu, P. (2023). *Limitations of classical test theory in modern assessment*. Beijing: Higher Education Press.