www.ijarp.com ISSN 2456-9992 Page: 01-12

International Journal Advanced Research Publications

Review Article

Volume: 02
Issue: 05

REVISITING TEACHERS' MISCONCEPTIONS: THE RELATIONSHIP BETWEEN VOCABULARY COMPLEXITY AND ITEM DIFFICULTY IN ITEM ANALYSIS

***Aloyce Nyoni**

Tanzania.

Article Received: 10 April 2026, Article Revised: 30 April 2026, Published on: 20 May 2026

***Corresponding Author: Aloyce Nyoni**

Tanzania.

DOI: https://doi-doi.org/101555/ijarp.4138

ABSTRACT

This study examined the relationship between vocabulary complexity and item difficulty and assessed whether teachers' perceptions align with empirical evidence derived from statistical item analysis. Anchored in the field of Educational Measurement, the study adopted a mixed-methods design involving quantitative analysis of 12 test items and questionnaire responses from 38 teachers complemented by qualitative interpretation of teachers' perceptions. Item difficulty was computed using Item Analysis, while vocabulary complexity was measured through a structured rubric focusing on lexical demand and linguistic features. The findings from Pearson correlation analysis revealed a weak and non-significant relationship between vocabulary complexity and item difficulty **(**$r = 0.24$, $p = 0.420$**),** indicating that vocabulary complexity is not a strong predictor of item difficulty. However, teachers strongly perceived vocabulary complexity and related linguistic features as major determinants of item difficulty, revealing a discrepancy between perception and empirical evidence. The study concludes that item difficulty is influenced more by cognitive demand, conceptual understanding and item structure than by vocabulary factor alone, the overreliance on vocabulary complexity may threaten Construct Validity in assessment practices.

KEYWORDS: Vocabulary complexity, item difficulty, item analysis, teachers' misconceptions, educational measurement

INTRODUCTION

Assessment is a central component of teaching and learning as it provides systematic evidence for evaluating student achievement and instructional effectiveness. Within educational assessment, item difficulty is a key psychometric indicator used to determine the quality of test items and their ability to differentiate between high and low performing students. In classical test theory, item difficulty is operationalized using students' performance data rather than subjective judgment in ensuring that interpretations are grounded in empirical evidence and measurement principles (Crocker & Algina, 1986; Nitko & Brookhart, 2014).

Despite this, in many educational contexts, teachers tend to assume that vocabulary complexity is the primary determinant factor of item difficulty. This assumption may lead to misconceptions in test construction and interpretation of student performance as it places irrelevant emphasis on linguistic features while underestimating cognitive and conceptual demands embedded in assessment tasks. Such misconceptions may also threaten the validity of assessment instruments by introducing construct-irrelevant variance, particularly when language complexity interferes with the accurate measurement of intended learning outcomes (Messick, 1995; Osterlind, 2002).

This study therefore investigates whether vocabulary complexity is a significant predictor of item difficulty and examines whether teachers' perceptions are consistent with empirical evidence generated through statistical item analysis within the framework of Educational Measurement.

Theoretical Literature Review

This study is grounded in classical and modern theories of educational assessment particularly those that explain how test item's function and how difficulty is determined. Central to this study is the framework of Item Response Theory and classical test theory both of which conceptualize item difficulty as a statistical property derived from examinee performance rather than subjective judgment. In classical test theory, item difficulty (p-value) represents the proportion of students who answer an item correctly meaning that difficulty is empirically determined by test-taker ability distribution rather than linguistic or surface features alone (Crocker & Algina, 1986).

The concept of Construct Validity provides an important theoretical lens for this study. Construct validity emphasizes that a test should measure the intended construct without contamination from irrelevant factors such as language complexity when the construct is not

language-related. According to Messick's unified validity framework, construct-irrelevant variance occurs when extraneous factors such as vocabulary difficulty, influence test performance and distort interpretation of scores (Messick, 1995). This theory is particularly relevant to the present study as it questions whether vocabulary complexity should meaningfully affect item difficulty in subject based assessments.

Theories of cognitive processing in assessment suggest that item difficulty is influenced by multiple interacting dimensions, including cognitive demand, reasoning processes, and familiarity with content, rather than vocabulary alone. Therefore, relying solely on linguistic features may oversimplify the multidimensional nature of item functioning in educational measurement (Nitko & Brookhart, 2014).

Empirical Literature Review

Empirical studies in the field of educational assessment have produced mixed findings regarding the relationship between vocabulary complexity and item difficulty. Some studies suggest that linguistic complexity can influence student performance particularly in contexts where reading ability is closely tied to content understanding. However, a larger body of research indicates that vocabulary complexity alone is not a strong predictor of item difficulty, as cognitive demand and conceptual understanding often play a more dominant role (Haladyna & Rodriguez, 2013).

Research in Educational Measurement has consistently shown that item difficulty is best explained through statistical item analysis rather than surface-level linguistic features. For example, studies using Item Analysis demonstrate that items with similar vocabulary complexity can vary significantly in difficulty depending on the reasoning required and the structure of the question.

Moreover, studies on teacher assessment practices reveal that educators often overestimate the influence of vocabulary complexity when judging item difficulty. This tendency reflects a reliance on intuition rather than empirical data, leading to potential misinterpretation of student performance and weak alignment between perceived and actual item functioning (Osterlind, 2002). Such findings highlight a persistent gap between assessment theory and classroom practice.

Overall, the empirical literature suggests that while vocabulary complexity may contribute to item difficulty in some cases, it is neither a sufficient nor a consistent predictor. Instead, item difficulty emerges from a combination of linguistic, cognitive, and content-related factors, reinforcing the need for evidence-based approaches in test development and evaluation

METHODOLOGY

The study adopted a mixed-methods research design to provide a comprehensive examination of the relationship between vocabulary complexity and item difficulty, as well as teachers' perceptions of test item construction. The quantitative component involved the analysis of 12 test items and responses from 29 teachers, while the qualitative component explored teachers' views on the factors influencing item difficulty. Data were collected using a structured Likert-scale questionnaire to capture teachers' perceptions, student test scripts to compute item difficulty indices, and a vocabulary complexity rubric designed to assess linguistic demand based on word difficulty, use of technical terminology, and sentence structure. Item difficulty was determined through Item Analysis by calculating p-values derived from student performance data. Quantitative data were analysed using descriptive statistics and Pearson correlation analysis to determine the relationship between vocabulary complexity and item difficulty, while qualitative responses were subjected to thematic analysis to identify recurring perceptions and possible misconceptions among teachers regarding the influence of vocabulary on item difficulty.

Data Presentation and Interpretation

This section presents the findings of the study on the relationship between vocabulary complexity and item difficulty, as well as teachers' perceptions regarding item difficulty. The results are organized into item difficulty analysis, vocabulary complexity analysis, correlation results, and interpretation.

Item Difficulty Results

Table 1: Item Difficulty Indices.

Item	DIFFICULTY INDEX P	INTERPRETATION
Q1	0.80	Easy
Q2	0.45	Moderate
Q3	0.20	Difficult
Q4	0.35	Moderate
Q5	0.75	Easy
Q6	0.50	Moderate
Q7	0.60	Moderate
Q8	0.30	Moderate
Q9	0.85	Easy
Q10	0.25	Difficult
Q11	0.70	Easy
Q12	0.65	Moderate

Interpretation

The comparison between item difficulty and vocabulary complexity reveals no consistent or direct relationship between the two variables. While it might be expected that items with more complex vocabulary would be more difficult, the data does not support this assumption. For instance, items such as Q9 and Q11 were classified as easy despite having relatively high vocabulary complexity, indicating that students were able to understand and respond correctly even when the language was more advanced. Conversely, items like Q3 and Q10 were found to be difficult despite having lower vocabulary complexity, suggesting that factors such as cognitive demand, conceptual difficulty, or item structure played a more significant role in determining difficulty. Similarly, items of moderate difficulty exhibited mixed levels of vocabulary complexity, further demonstrating the absence of a clear pattern. These findings indicate that vocabulary complexity is not a reliable predictor of item difficulty and support the broader statistical evidence of a weak relationship between the two variables. This has important implications for Construct Validity, as overemphasis on vocabulary may lead to incorrect assumptions about what test items measure, potentially introducing construct-irrelevant factors and undermining the validity of assessment outcomes.

Vocabulary Complexity Results

Table 2: Vocabulary Complexity Scores.

ITEM	VOCABULARY SCORE	LEVEL
Q1	3.9	Moderate
Q2	3.9	Moderate
Q3	3.2	Low
Q4	4.4	High
Q5	3.9	Moderate
Q6	3.9	Moderate
Q7	3.8	Moderate
Q8	3.2	Low
Q9	4.1	High
Q10	2.4	Low
Q11	4.4	High
Q12	4.5	High

Interpretation

The vocabulary complexity results indicate a varied distribution of linguistic demand across the test items, ranging from low to high levels. A majority of the items (Q1, Q2, Q5, Q6, and Q7) fall within the moderate complexity range, suggesting a generally balanced use of language that is neither overly simple nor excessively complex. Four items (Q4, Q9, Q11,

and Q12) were classified as having high vocabulary complexity, reflecting the presence of more advanced or technical language features, while three items (Q3, Q8, and Q10) exhibited low vocabulary complexity, indicating relatively simple wording. Despite this variation, the distribution does not follow a clear pattern when compared with item difficulty, as some high-complexity items corresponded with high student performance, while some low-complexity items were associated with lower performance. This inconsistency suggests that vocabulary complexity alone does not determine how students respond to test items, and that other factors such as cognitive demand and conceptual understanding play a more substantial role. These findings reinforce concerns related to Construct Validity, highlighting that language complexity should be carefully managed to avoid introducing construct-irrelevant difficulty that may distort the intended measurement of students' knowledge and skills.

Correlation Analysis

Table 3: Pearson Correlation Results.

Variables	r	p-value
Vocabulary Complexity vs Item Difficulty	0.24	0.420

Interpretation

The Pearson correlation analysis revealed a weak positive relationship between vocabulary complexity and item difficulty ($r = .24$), indicating only a slight tendency for items with higher vocabulary complexity to be associated with increased difficulty. However, this relationship was not statistically significant ($p = 0.420 > 0.05$), suggesting that the observed association may have occurred by chance and cannot be generalized beyond the sample. This finding demonstrates that vocabulary complexity is not a strong or reliable predictor of item difficulty and supports the view that other factors such as cognitive demand, conceptual clarity, and item structure play a more substantial role in determining how difficult an item is for students. The result reinforces evidence from Item Analysis and raises important considerations for Construct Validity as overreliance on vocabulary as an indicator of difficulty may lead to misinterpretation of student performance and compromise the accuracy of assessment outcomes.

Teachers' Perceptions

Table 4: Teachers' Perceptions on Vocabulary Complexity and Item Difficulty.

Statement	SD	D	N	A	SA
Items with complex vocabulary are usually more difficult	2	0	0	14	20
Long sentences increase item difficulty	2	4	1	21	9
Students often fail items due to unfamiliar vocabulary	1	2	2	13	20
Simplifying vocabulary reduces item difficulty	1	0	1	16	20
Technical terms make items more difficult	3	4	4	12	15
Complex wording can mislead students	1	0	5	13	19
Teachers can judge item difficulty based on wording alone	7	13	8	6	4

Key: SD = Strongly Disagree; D = Disagree; N = Neutral; A = Agree; SA = Strongly Agree

The findings on teachers' perceptions revealed a strong belief that vocabulary complexity plays a significant role in determining item difficulty. The majority of respondents perceived linguistic features as key factors affecting students' performance on assessment items. Specifically, most teachers agreed that items containing complex vocabulary tend to be more difficult for students, with 55.6% strongly agreeing and 38.9% agreeing that vocabulary complexity increases item difficulty. Likewise, 52.6%of respondents strongly agreed and 34.2% agreed that students often fail test items because of unfamiliar vocabulary. The findings further indicated that teachers associated simplified language with reduced item difficulty. A large proportion of respondents about 52.6% strongly agreed and 42.1% believed that simplifying vocabulary reduces the difficulty level of test items. Similarly, regarding sentence structure, **5**6.8% of respondents agreed and 24.3% strongly agreed that long sentences increase the difficulty of assessment items.

Additionally, teachers perceived technical terminology and wording structure as important contributors to item difficulty. Specifically, 39.5% of respondents strongly agreed and 31.6% agreed that technical terms increase item difficulty. Likewise, 50% strongly agreed and 34.2% agreed that complex wording can mislead students even when they understand the intended content.

However, perceptions concerning the determination of item difficulty through empirical approaches produced relatively mixed responses. A considerable proportion of respondents questioned the adequacy of relying solely on wording when determining item difficulty, with 34.2% of respondents disagreeing and 18.4% strongly disagreeing that teachers can accurately judge item difficulty based only on wording characteristics.

Overall Interpretation

The overall findings reveal a noticeable discrepancy between teachers' perceptions and empirical evidence regarding the determinants of item difficulty. Teachers strongly perceived vocabulary complexity, unfamiliar terms, sentence structure, and wording characteristics as major factors influencing item difficulty and students' performance in assessments. However, the statistical findings from Item Analysis and Pearson correlation analysis demonstrated only a weak and non-significant relationship between vocabulary complexity and item difficulty ($r = .24$, $p = 0.420$), suggesting that vocabulary alone does not sufficiently explain variations in item difficulty.

These findings indicate that item difficulty is influenced by multiple interacting factors including cognitive demand, conceptual understanding, reasoning requirements, and item structure rather than linguistic features alone. The strong emphasis teachers place on vocabulary characteristics reflects a perception-based interpretation that is not fully supported by empirical evidence. This has important implications for Construct Validity, as excessive reliance on vocabulary complexity during test development and evaluation may introduce construct-irrelevant variance and potentially shift assessment from measuring intended subject knowledge toward unintended constructs such as language proficiency.

DISCUSSION OF THE FINDINGS

The findings of this study revealed a discrepancy between teachers' perceptions and empirical evidence regarding the determinants of item difficulty. Teachers strongly believed that vocabulary complexity, unfamiliar words, technical terminology, sentence structure, and wording characteristics substantially influence the difficulty of assessment items. However, the findings from Item Analysis and Pearson correlation analysis indicated only a weak and non-significant relationship between vocabulary complexity and item difficulty ($r = .24$, $p = 0.420$). These findings can be interpreted within the framework of Classical Test Theory (CTT) and Construct Validity theory, which guided this study. According to Classical Test Theory, item difficulty is considered a statistical property determined through the proportion of students who answer an item correctly rather than through subjective judgments about the appearance or wording of test items (Crocker & Algina, 1986). Therefore, the strong perceptions expressed by teachers are not fully consistent with theoretical expectations, as item difficulty should be established through empirical performance evidence rather than assumptions regarding vocabulary complexity.

The findings are further supported by Construct Validity theory, which emphasizes that assessment instruments should accurately measure the intended construct without contamination from irrelevant factors (Messick, 1995). Teachers in this study strongly emphasized vocabulary characteristics as indicators of item difficulty, suggesting a tendency to associate language complexity with students' ability to answer assessment items. However, the weak relationship identified between vocabulary complexity and item difficulty suggests that language characteristics alone do not adequately explain differences in item performance. Overreliance on vocabulary complexity may introduce construct-irrelevant variance, where students' performance becomes influenced by language-related factors instead of the intended subject knowledge or skills being measured (Messick, 1995). Such situations may threaten the validity and fairness of assessment outcomes.

The findings also align with theories of cognitive processing in assessment, which suggest that item difficulty emerges from multiple interacting factors including cognitive demand, reasoning requirements, prior knowledge, and conceptual understanding rather than linguistic characteristics alone (Nitko & Brookhart, 2014). This interpretation is further supported by the item analysis results of the present study, where some items with relatively high vocabulary complexity were classified as easy while some items with lower vocabulary complexity were identified as difficult. Such findings indicate that deeper cognitive processes and task requirements may have a stronger influence on item difficulty than vocabulary itself. Similar observations have been reported in educational assessment research, where educators frequently rely on intuitive judgments and equate reading difficulty with overall item difficulty despite empirical evidence suggesting that the two constructs are not necessarily equivalent (Osterlind, 2002; Haladyna & Rodriguez, 2013).

Generally, the findings suggest that teachers may rely heavily on perception-based judgments when evaluating assessment items, while theoretical and empirical evidence indicate that item difficulty is multidimensional and should be determined through evidence-based approaches. These findings reinforce the importance of strengthening teachers' assessment literacy and promoting the use of statistical item analysis in educational practice to improve the validity and quality of assessment decisions.

CONCLUSION OF THE STUDY

The study concludes that although teachers strongly perceive vocabulary complexity as an important determinant of item difficulty, empirical findings do not provide sufficient statistical support for this assumption. The questionnaire findings indicated that most teachers

believed that complex vocabulary, unfamiliar terms, technical words, and sentence structure increase the difficulty of assessment items and influence students' performance. However, the findings from Item Analysis and Pearson correlation analysis revealed only a weak and non-significant relationship between vocabulary complexity and item difficulty **(**r = .24, p = 0.420**),** suggesting that vocabulary complexity alone is not a strong predictor of item difficulty.

The findings further suggest that item difficulty is influenced by multiple interacting factors, including cognitive demand, conceptual understanding, reasoning processes, and item structure rather than linguistic characteristics alone. The discrepancy observed between teachers' perceptions and empirical evidence indicates the presence of misconceptions regarding how assessment items function in practice. From an Educational Measurement perspective, such misconceptions may affect the quality of assessment and threaten Construct Validity, particularly when vocabulary complexity is overemphasized during test construction and evaluation. Therefore, assessment practices should be guided by empirical evidence and statistical item analysis rather than relying solely on subjective perceptions and linguistic assumptions.

Recommendations of the Study

- **Strengthen teachers' training on evidence-based determination of item difficulty**.

The findings of this study revealed that teachers strongly perceived vocabulary complexity, unfamiliar words, sentence structure, and technical terms as major determinants of item difficulty despite empirical findings showing only a weak and non-significant relationship between vocabulary complexity and item difficulty. This indicates the need for structured professional development programs that equip teachers with knowledge and practical skills in Item Analysis and Educational Measurement. Such training should emphasize the interpretation of item difficulty indices and the use of statistical evidence in evaluating assessment items to reduce reliance on subjective perceptions.

- **Promote evidence-based assessment practices in schools and educational institutions**.

The study identified a discrepancy between teachers' perceptions and empirical findings regarding the factors determining item difficulty. Schools and educational institutions should therefore encourage the use of student performance data and empirical evidence in making assessment decisions. Greater use of evidence-based approaches will support more accurate

evaluation of test items and minimize misconceptions regarding the influence of vocabulary complexity on assessment outcomes.

- **Strengthen assessment literacy in teacher education and professional development programs**.

The findings suggest that teachers may place excessive emphasis on linguistic characteristics while underestimating the role of cognitive demand, conceptual understanding and item structure in determining item difficulty. Teacher education institutions should therefore strengthen assessment literacy by integrating comprehensive content on measurement theory, test construction principles, Construct Validity and interpretation of item statistics. Improving assessment literacy will enable teachers to make more informed and scientifically grounded decisions during test development and evaluation.

- **Institutionalize the routine use of Item Analysis in classroom and institutional assessments**.

To improve assessment quality and validity, schools and other educational institutions should ensure that Item Analysis becomes a routine component of both formative and summative assessment practices. Regular use of Item Analysis can help identify poorly functioning items, determine actual levels of difficulty and ensure that assessment practices reflect intended learning outcomes. Consistent use of empirical item evaluation procedures will enhance assessment quality and reduce the risk of introducing construct-irrelevant factors into educational measurement.

REFERENCES

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
2. Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.
3. Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
4. Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items* (3rd ed.). Routledge. https://doi.org/10.4324/9780203850381
5. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

6. Nitko, A. J., & Brookhart, S. M. (2014). *Educational assessment of students* (7th ed.). Pearson Education.
7. Osterlind, S. J. (2002). *Constructing test items: multiple-choice, constructed-response, performance, and other formats* (2nd ed.). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-010-0241-6