E-ISSN: 1115 - 8425 P-ISSN: 3027-012X
MPRI-JAEE | INTERNATIONAL JOURNAL OF ASSESSMENT & EVALUATION IN EDU.
VOL. 13 NO. 8 – JULY. 2026

DEVELOPMENT AND VALIDATION OF THE EVOLUTIONARY MISCONCEPTION TEST FOR DIAGNOSING EVOLUTIONARY MISCONCEPTIONS AMONG SENIOR SECONDARY SCHOOL STUDENTS

AKILU ISMA'IL; & OLATUNBOSUN MUJIDAT LUKMAN

Department of Science Education, Faculty of Education, Federal University Gusau, Gusau, Zamfara State, Nigeria.
Corresponding Author: akilu@fugusau.edu.ng
DOI: https://doi.org/10.70382/mejaee.v13i8.0150

Abstract

Accurate identification of students' evolutionary misconceptions remains a major challenge in biology education because conventional assessments often fail to reveal the reasoning underlying students' responses. This study developed the Evolutionary Misconception Test (EMT) and established psychometric evidence for its use as a two-tier diagnostic instrument for assessing evolutionary conceptual difficulties among senior secondary school students. A multi-phase instrument development and validation design involving misconception identification, item development, expert review, pilot testing, and field validation was employed. An initial pool of 45 two-tier multiple-choice items was developed from the literature, curriculum content, biology textbooks, and documented student misconceptions. Following expert review and item analysis, 30 items covering six evolutionary concepts were retained. The final EMT was administered to 200 Senior Secondary II Biology students in Zamfara State, Nigeria. Data were analysed using exploratory factor analysis, item analysis, and Kuder–Richardson Formula 20 (KR-20) reliability estimation. Results provided evidence of satisfactory construct validity (KMO = 0.841; Bartlett's $\chi^2(435) = 2768.21$, $p < .001$), a one-factor structure explaining 39.21% of the variance, factor loadings ranging from 0.42 to 0.76, and satisfactory

Keywords: Biology Education; Conceptual Change; Evolution Education; Instrument Validation; Psychometric Properties; Two-Tier Diagnostic Test.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

internal consistency reliability (KR-20 = 0.84). The EMT also identified persistent misconceptions, particularly in natural selection and adaptation. These findings indicate that the EMT is a valid and reliable diagnostic instrument for identifying students' evolutionary misconceptions and supporting conceptual change instruction, while further validation with larger and more diverse samples is recommended.

INTRODUCTION

Evolution is a fundamental concept in Biology and serves as the unifying framework for explaining the diversity of life, biological processes, and contemporary issues related to health, agriculture, and environmental sustainability (Sá-Pinto et al., 2021; Brownell et al., 2014; Furrow & Hsu, 2019; Torkar & Gregorčič, 2026). It occupies a central position in Biology curricula because it provides the conceptual basis for understanding natural selection, adaptation, genetic variation, common ancestry, and speciation (Araújo, 2022; Matazu & Isma'il, 2024). Consequently, meaningful understanding of evolutionary concepts is essential for scientific literacy and informed biological reasoning (Torkar & Gregorčič, 2026; Woods & Scharmann, 2001). The teaching of evolution continues to be challenged by persistent misconceptions among secondary school students despite its recognized importance. Studies have shown that many learners hold scientifically inaccurate conceptions regarding natural selection, adaptation, common ancestry, and other evolutionary processes, often retaining these misconceptions even after formal instruction (Matazu & Isma'il, 2024; Yates & Marek, 2014; Archila et al., 2024; Hartelt & Martens, 2025; Nurdiansyah et al., 2022). Such misconceptions impede conceptual understanding and hinder students' ability to interpret biological evidence accurately (Matazu & Isma'il, 2024; Yates & Marek, 2014).

Effective remediation of misconceptions depends on accurate diagnosis before instruction. Diagnostic assessment enables teachers to identify learners' existing conceptions, distinguish misconceptions from lack of knowledge, and design appropriate conceptual change interventions (Archila et al., 2024; Desstya et al., 2025; Kuschmierz et al., 2020). Among available approaches, two-tier diagnostic tests have gained considerable acceptance because they assess both students' answer choices and the reasoning underlying those responses, thereby providing more valid evidence of misconceptions than conventional multiple-choice tests (Taşçi, 2024; Ningroom et al., 2025; Ristanto et al., 2023).

Although several evolution assessment instruments have been developed, many focus on specific evolutionary concepts, target undergraduate populations, or measure conceptual understanding rather than misconceptions, limiting their applicability to senior secondary school students (Furrow & Hsu, 2019; Kuschmierz et al., 2020; Goff,

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

2017). Furthermore, evidence supporting their psychometric properties for use in Nigerian secondary schools is scarce. This highlights the need for a valid and reliable diagnostic instrument capable of identifying evolutionary misconceptions across major evolution concepts among senior secondary school students. Therefore, this study developed and psychometrically validated the Evolutionary Misconception Test (EMT) as a two-tier diagnostic instrument for diagnosing evolutionary misconceptions among senior secondary school students.

Objective of the Study

The objective of this study was to develop and psychometrically validate the EMT as a two-tier diagnostic instrument for assessing evolutionary misconceptions among senior secondary school Biology students.

LITERATURE REVIEW

Evolution is recognized as one of the most conceptually demanding topics in Biology because it requires learners to integrate abstract ideas such as natural selection, adaptation, genetic variation, common ancestry, and speciation into coherent scientific explanations. Despite its central role in Biology, research consistently shows that misconceptions about evolution remain widespread among secondary school students across different educational settings (Archila et al., 2024; Hartelt & Martens, 2025; Matazu & Isma'il, 2024). Rather than interpreting evolution as a population-level process driven by natural selection and genetic variation, students frequently attribute evolutionary change to need, effort, purposeful intention, or the inheritance of acquired characteristics (Yates & Marek, 2014; Nurdiansyah et al., 2022; Panagou et al., 2024). These misconceptions are often internally consistent from the learners' perspective, making them resistant to conventional instruction. The persistence of evolutionary misconceptions has been attributed to multiple interacting factors, including prior beliefs, inadequate classroom instruction, inaccurate textbook explanations, cultural influences, and teachers' own misconceptions about evolution (Yates & Marek, 2014; Nurdiansyah et al., 2022; Matazu & Isma'il, 2024). Consequently, students frequently retain scientifically inaccurate explanations despite completing formal Biology instruction. Hartelt and Martens (2025) further observed that students often overestimate the accuracy of their own understanding, making misconceptions difficult to identify without appropriate diagnostic assessment. Within the Nigerian situation, Matazu and Isma'il (2024) reported persistent misconceptions among secondary school students across major evolutionary concepts, highlighting the need for assessment instruments capable of accurately identifying specific conceptual difficulties before instructional intervention.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Effective remediation of misconceptions begins with accurate diagnosis because instructional interventions are unlikely to succeed when students' alternative conceptions remain unidentified. Unlike conventional achievement tests, diagnostic assessments are designed to reveal not only whether students answer correctly but also how they reason about scientific concepts (Archila et al., 2024; Desstya et al., 2025). This distinction is important because correct answers obtained through guessing or memorization may conceal underlying misconceptions that continue to influence students' conceptual understanding. Diagnostic assessment therefore provides teachers with detailed information about students' reasoning patterns, enabling them to identify specific misconceptions and design instructional strategies that promote conceptual change rather than factual recall (Hartelt & Martens, 2025). Such information is equally valuable for educational researchers because it provides empirical evidence regarding the nature and prevalence of misconceptions within particular populations. As conceptual change depends on identifying learners' existing conceptions before introducing scientifically acceptable explanations, diagnostic assessment has become an indispensable component of science education research (Archila et al., 2024; Desstya et al., 2025).

Several approaches have been employed to identify students' misconceptions, including interviews, open-ended questions, concept maps, and conventional multiple-choice tests. Although interviews provide rich diagnostic information, they are time-consuming and impractical for large-scale classroom use. Conventional multiple-choice tests are more efficient but cannot distinguish genuine conceptual understanding from correct responses obtained through guessing or rote memorization (Taşçi, 2024; Ningroom et al., 2025). These limitations have contributed to the growing use of two-tier diagnostic tests in science education. A two-tier diagnostic test consists of two related questions. The first-tier measures students' conceptual understanding by requiring them to select the scientifically correct answer, while the second tier examines the reasoning underlying that choice. Credit is awarded only when both responses are scientifically correct, thereby reducing the influence of guessing and improving diagnostic accuracy (Ristanto et al., 2023; Taşçi, 2024). By simultaneously assessing knowledge and reasoning, two-tier instruments provide more meaningful evidence of misconceptions than conventional achievement tests.

Recent studies have shown the effectiveness of two-tier diagnostic instruments across Biology, Chemistry, and general science. Ningroom et al. (2025), Taşçi (2024), and Ristanto et al. (2023) developed two-tier instruments capable of identifying specific misconception patterns with acceptable psychometric properties. Similarly, Desstya et al. (2025) and Soeharto and Csapó (2022) showed that systematically developed diagnostic instruments incorporating expert validation, pilot testing, and empirical item analysis can achieve

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

satisfactory validity and reliability. Collectively, these studies suggest that two-tier diagnostic tests provide a robust framework for diagnosing misconceptions when developed using appropriate psychometric procedures.

The quality of any diagnostic instrument depends on the rigor of its development and validation process. Instrument development begins with identifying common misconceptions from curriculum documents, previous research, classroom observations and expert knowledge before constructing items that represent both scientifically correct explanations and prevalent alternative conceptions (Taşçi, 2024; Arifah et al., 2022). This process ensures that distractors reflect authentic misconceptions rather than implausible alternatives. Validation is equally important because an instrument cannot be assumed to function effectively merely because it has been carefully constructed (Isma'il & Sodangi, 2025). Expert review provides evidence of face and content validity by evaluating the relevance, clarity, and representativeness of individual items. Pilot testing subsequently generates empirical evidence regarding item functioning, while item analysis determines whether each question adequately discriminates between students with strong and weak conceptual understanding (Desstya et al., 2025; Soeharto & Csapó, 2022). Indicators such as item difficulty, discrimination index, distractor effectiveness, and internal consistency reliability provide objective evidence of instrument quality. Together, these procedures establish the psychometric soundness required for educational assessment and research.

Although several instruments have been developed to assess evolution learning, important limitations remain. Existing instruments vary considerably in purpose, content coverage, target population, and psychometric evidence (Furrow & Hsu, 2019; Kuschmierz et al., 2020). Many assess students' general understanding of evolution or focus on isolated concepts such as natural selection or genetic drift rather than providing a comprehensive diagnosis of misconceptions across the major evolution concepts taught in secondary schools. Others were developed primarily for undergraduate students or measure acceptance of evolution rather than scientifically inaccurate conceptions (Goff, 2017; Kuschmierz et al., 2020). Furthermore, psychometric evidence obtained in one educational setting cannot automatically be generalized to another. Differences in curriculum, language, culture, and learners' prior experiences may substantially influence how students interpret diagnostic items and respond to distractors (Kuschmierz et al., 2020). Consequently, instruments validated elsewhere may not provide equally valid evidence when administered to Nigerian secondary school students without further validation. These limitations stress the importance of developing appropriate diagnostic instruments supported by empirical psychometric evidence.

The reviewed literature reveals that misconceptions about evolution remain widespread despite the central importance of evolution in Biology education. While previous studies have established the value of two-tier diagnostic instruments and documented

acceptable procedures for their development and validation, few instruments comprehensively assess misconceptions across the major evolutionary concepts prescribed for senior secondary school students. Existing evolution assessment instruments primarily target different educational levels, emphasize conceptual understanding or acceptance of evolution rather than misconceptions, or provide limited evidence regarding their psychometric suitability for Nigerian learners (Furrow & Hsu, 2019; Goff, 2017; Kuschmierz et al., 2020). Consequently, there is no adequately validated two-tier diagnostic instrument specifically designed to identify evolutionary misconceptions among Nigerian senior secondary school students. This gap necessitated the development and psychometric validation of the EMT to provide a valid and reliable instrument for diagnosing misconceptions across the major evolution concepts contained in the senior secondary Biology curriculum.

METHODOLOGY

Research Design

This study adopted a multi-phase instrument development and validation design to develop and establish the psychometric properties of the Evolutionary Misconception Test (EMT) for diagnosing evolutionary misconceptions among Senior Secondary School II (SS II) Biology students. The instrument development process comprised five phases: identification of students' evolutionary misconceptions, item generation, content validation, pilot testing, and psychometric validation.

Participants

The study involved 352 SS II Biology students drawn from secondary schools in Zamfara State, Nigeria. Schools offering Senior Secondary Biology were selected purposively to ensure the availability of the target population. Within the selected schools, participating students were selected through simple random sampling. Participants were involved across three phases of instrument development: misconception identification, pilot testing, and psychometric validation. Students who participated in the pilot testing were excluded from the final validation phase to avoid testing effects.

Table 1: Distribution of Participants Across the EMT Development Phases

Phase	Purpose	Participants
Phase I	Identification of evolutionary misconceptions	50
Phase II	Pilot testing of the preliminary EMT	102
Phase III	Psychometric validation of the final EMT	200
Total		352

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Development of the Evolutionary Misconception Test

The EMT was developed based on the SS II Biology curriculum, literature on evolutionary misconceptions, Biology textbooks, and students' alternative conceptions identified during the initial phase. The instrument covered six major evolutionary concepts: natural selection, adaptation, genetic variation, common ancestry, speciation, and evidence for evolution. An initial pool of 45 two-tier multiple-choice items was generated. Each item consisted of two components. The first tier required students to select the scientifically correct response, while the second tier required them to select the reason that best explained their answer. The second tier enabled the identification of misconception-based reasoning rather than relying solely on answer selection.

The EMT employed dichotomous scoring. An item received 1 mark only when both Tier 1 and Tier 2 responses were correct; all other responses received 0. Thus, each of the 30 items contributed a single 0/1 score to the psychometric analysis. The maximum score was 30, with higher scores indicating greater evolutionary conceptual understanding and fewer misconceptions.

Content Validation of the EMT

The preliminary EMT was evaluated by three experts in Biology Education, Measurement and Evaluation, and secondary school Biology teaching. The experts examined each item for scientific accuracy, curriculum alignment, clarity of expression, relevance to the intended construct, and appropriateness of the distractors. Their recommendations guided the revision, modification, and refinement of the instrument before pilot administration. Quantitative evidence of content validity was established using expert ratings and appropriate content validity indices.

Pilot Testing and Psychometric Validation

The revised EMT was pilot tested using 102 SS II Biology students from schools that did not participate in the final validation phase. The pilot study assessed item clarity, response patterns, item difficulty, discrimination, distractor functioning, and the overall suitability of the instrument.

Following the pilot study, poorly performing items were revised or removed, resulting in a final 30-item EMT. The instrument was subsequently administered to 200 SS II Biology students for psychometric validation. This sample size was considered adequate for exploratory factor analysis and item analysis in instrument development studies.

Construct validity was examined using Exploratory Factor Analysis (EFA) with Principal Axis Factoring after assessing sampling adequacy using the KMO measure and Bartlett's test of sphericity. Because the EMT items were dichotomously scored (0/1), EFA was based on a Pearson correlation matrix. Although tetrachoric correlations are often

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

recommended for binary data, Pearson correlations were used in the present study; this distinction is acknowledged as a methodological limitation because Pearson correlations may produce conservative factor loadings. Although the 30 items covered six related concepts, the EMT was designed to measure an overarching construct of evolutionary conceptual understanding. The single-factor solution was therefore retained, while future studies should examine competing multifactor models using confirmatory factor analysis with larger independent samples. Internal consistency reliability was estimated using KR-20.

Data Analysis

Data obtained during the misconception identification phase were analysed using content analysis to identify students' alternative conceptions, which informed item development and distractor construction. Data from the validation phase were analysed using descriptive statistics and psychometric procedures. Construct validity was evaluated using the KMO measure, Bartlett's test of sphericity, and Exploratory Factor Analysis (Principal Axis Factoring). Item quality was assessed using item difficulty indices, discrimination indices, point-biserial correlation coefficients, and distractor effectiveness analysis. Internal consistency reliability was determined using the KR-20.

Ethical Considerations

Ethical approval was obtained from the relevant school authorities. Participation was voluntary, and informed consent was obtained from the participants and their schools.

RESULTS

Item Refinement and Final Structure of the Evolutionary Misconception Test

The development of the EMT began with an initial pool of 45 two-tier diagnostic multiple-choice items generated from evolutionary misconception literature, SS II Biology curriculum content, Biology textbooks, and alternative conceptions identified during the preliminary misconception identification phase. Each item consisted of a content tier assessing students' selected answers and a reasoning tier assessing explanations for their responses. The initial items were subjected to expert validation and pilot testing to determine scientific accuracy, curriculum relevance, clarity, distractor effectiveness, and psychometric quality. Following expert recommendations, items with unclear wording, inadequate conceptual representation, or inappropriate distractors were revised or removed. During pilot testing, item analysis was conducted using difficulty index, discrimination index, point-biserial correlation, and distractor functioning. Items with weak discrimination, low item-total relationships, or ineffective distractors were eliminated. The refinement process as shown in table 2 resulted in a final 30-item EMT.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Table 2: Item Refinement Process of the Evolutionary Misconception Test

Development Stage	Number of Items	Decision
Initial item pool	45	Items generated from literature, curriculum, and identified misconceptions
Expert validation	40	Five items revised or removed based on expert recommendations
Pilot testing and item analysis	30	Ten items removed due to low discrimination indices, weak item-total correlations, or ineffective distractors
Final EMT	30	Items retained for psychometric evaluation

Table 2 shows the systematic refinement of the EMT from 45 initial items to the final 30-item instrument. Expert validation reduced the item pool to 40 items after five items were revised or removed. Subsequent pilot testing and psychometric analysis resulted in the removal of 10 additional items based on item quality criteria, producing a psychometrically evaluated instrument.

Table 3: Distribution of Final EMT Items by Evolutionary Concept

Evolutionary Concept	Number of Items	Percentage (%)
Natural selection	6	20.0
Adaptation	5	16.7
Genetic variation	5	16.7
Common ancestry	5	16.7
Speciation	4	13.3
Evidence for evolution	5	16.7
Total	**30**	**100**

Table 3 presents the distribution of the final 30 EMT items across six evolutionary concepts. Natural selection contained the highest number of items with 6 items (20.0%), while adaptation, genetic variation, common ancestry, and evidence for evolution each contained 5 items (16.7%). Speciation was represented by 4 items (13.3%), indicating coverage of all targeted evolutionary concepts.

Construct Validity of the EMT

EFA was conducted using the combined EMT scores to examine the underlying structure of the instrument. The sample consisted of 200 students responding to 30 items, producing a participant-to-item ratio of approximately 6.7:1, which was considered adequate for exploratory factor analysis. The suitability of the data was examined using

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity. The KMO values and significant Bartlett's tests indicated that the correlation matrices were appropriate for factor analysis.

Table 4: Kaiser–Meyer–Olkin Measure and Bartlett's Test of Sphericity for EMT Responses

EMT Component	KMO Value	Bartlett's χ^2	df	*p*-value
Tier 1 Responses	0.812	2146.73	435	< .001
Tier 2 Responses	0.836	2389.54	435	< .001
Combined EMT Scores	0.841	2768.21	435	< .001

Table 4 indicates that the EMT responses were suitable for factor analysis. The KMO values ranged from 0.812 to 0.841, exceeding the recommended minimum value of 0.70. Bartlett's test of sphericity was significant for Tier 1 responses (χ^2 = 2146.73, df = 435, p < .001), Tier 2 responses (χ^2 = 2389.54, df = 435, p < .001), and combined EMT scores (χ^2 = 2768.21, df = 435, p < .001). Principal axis factoring was performed on the combined EMT scores. Factor retention was determined using eigenvalues greater than one, examination of the scree plot, and theoretical interpretability. The analysis produced a single-factor structure, with all retained items contributing meaningfully to the measured construct.

The construct validity of the Evolutionary Misconception Test (EMT) was further examined using Exploratory Factor Analysis (Principal Axis Factoring). Figure 1 presents the resulting factor structure of the instrument.

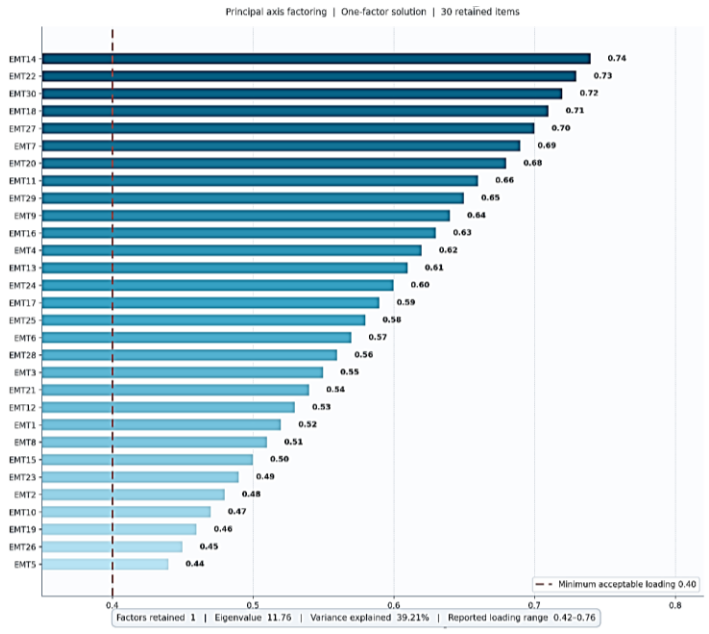

Figure 1: Factor Loadings and Factor Characteristics of the Final EMT

Figure 1 presents the factor structure of the EMT. Principal axis factoring produced a one-factor solution with an eigenvalue of 11.76, explaining 39.21% of the total variance. Factor loadings ranged from 0.42 to 0.76 across the 30 retained items, with all items exceeding the minimum acceptable loading criterion of 0.40.

Reliability and Item Quality Analysis

The internal consistency reliability of the EMT was determined using the Kuder–Richardson Formula 20 (KR-20) because items were scored dichotomously. Item retention was based on acceptable difficulty indices (0.20–0.80), discrimination indices above 0.30, and satisfactory point-biserial correlations as shown in Table 5.

Table 5: Reliability and Item Characteristics of the EMT

Parameter	Minimum	Maximum	Mean
Item difficulty index	0.31	0.76	0.54
Discrimination index	0.32	0.68	0.47
Point-biserial correlation	0.34	0.71	0.51
KR-20 reliability coefficient			0.84

Table 5 presents the reliability and item characteristics of the EMT. Difficulty indices ranged from 0.31 to 0.76 (M = 0.54), discrimination indices ranged from 0.32 to 0.68 (M = 0.47), and point-biserial correlations ranged from 0.34 to 0.71 (M = 0.51). The KR-20 coefficient of 0.84 indicated satisfactory internal consistency reliability.

Descriptive Statistics of EMT Scores

Table 6: Descriptive Statistics of EMT Scores (n = 200)

Variable	Mean	SD	Minimum	Maximum
EMT total score	15.84	5.72	3	28

Table 6 presents students' performance on the EMT. The mean score was 15.84 (SD = 5.72) out of a possible 30 points. Individual scores ranged from 3 to 28, showing variation in students' responses. The results indicate differences in students' levels of evolutionary conceptual understanding as measured by the two-tier diagnostic instrument.

Diagnostic Classification of Students' Conceptual Understanding

Students were classified based on their two-tier responses. Scientific understanding was assigned when both the answer and explanation were scientifically correct. Partial understanding occurred when only one tier was correct. Misconceptions were identified

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

when incorrect responses reflected documented alternative conceptions, while lack of knowledge represented incorrect responses without identifiable reasoning.

Table 7: Distribution of Students' Conceptual Categories Based on EMT Response Patterns (n = 200)

Conceptual Category	Frequency	Percentage (%)
Scientific understanding	64	32.0
Partial understanding	55	27.5
Evolutionary misconception	49	24.5
Lack of knowledge	32	16.0
Total	200	100

Table 7 shows that 64 students (32.0%) indicated scientific understanding, while 55 students (27.5%) showed partial understanding. Evolutionary misconceptions were identified among 49 students (24.5%), whereas 32 students (16.0%) revealed lack of knowledge.

Distribution of Identified Evolutionary Misconceptions

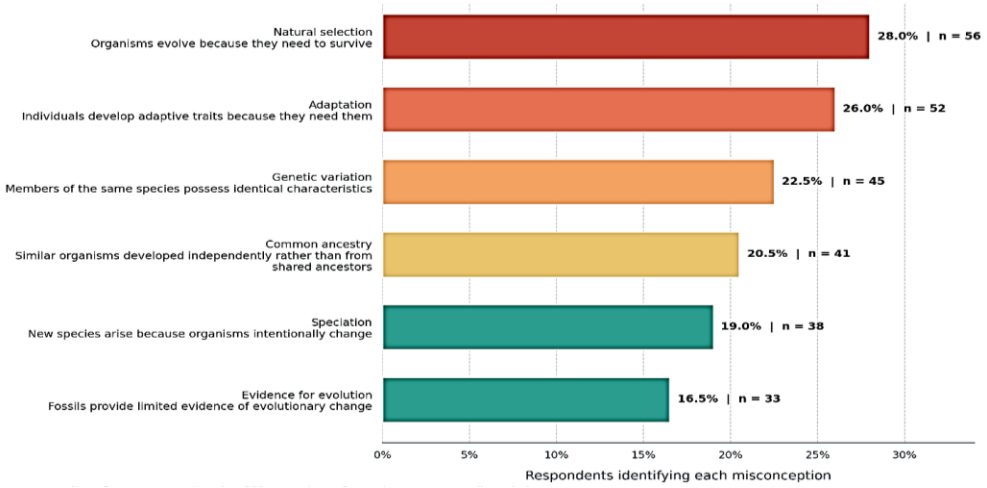

Figure 2: Evolutionary Misconceptions Identified Through the EMT (n = 200)

Figure 2 presents the specific evolutionary misconceptions identified among students. Natural selection recorded the highest misconception frequency, with 56 students (28.0%) demonstrating this misconception, followed by adaptation with 52 students (26.0%). Other misconceptions occurred in genetic variation (45, 22.5%), common ancestry (41, 20.5%), speciation (38, 19.0%), and evidence for evolution (33, 16.5%). Percentages were calculated independently because students could show multiple misconceptions.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

DISCUSSION

The purpose of this study was to develop and establish psychometric evidence for the EMT as a diagnostic instrument for identifying students' conceptions and misconceptions of evolution among senior secondary school students. The findings provide evidence supporting the validity, reliability, and diagnostic usefulness of the EMT. Unlike conventional achievement tests that primarily determine whether students select correct answers, the EMT examines both students' responses and the reasoning underlying those responses. This two-tier structure offers a more comprehensive representation of students' conceptual understanding by distinguishing scientific understanding, partial understanding, and misconception-based reasoning.

The construct validity evidence indicates that the EMT possesses a coherent internal structure for assessing students' evolutionary conceptual understanding. Although the instrument covers multiple concepts, including natural selection, adaptation, genetic variation, common ancestry, speciation, and evidence for evolution, these concepts collectively reflect a common underlying construct of evolutionary conceptual understanding. This finding is consistent with Sá-Pinto et al. (2021), who argued that evolution should be viewed as an integrated biological framework rather than a collection of isolated concepts. Similarly, Torkar and Gregorčič (2026) emphasized that meaningful understanding of evolution depends on integrating concepts such as variation, inheritance, natural selection, and common ancestry. The emergence of a single underlying construct therefore suggests that students' understanding of individual evolutionary mechanisms contributes to their overall understanding of evolution.

The reliability evidence further demonstrates that the EMT provides consistent measurement of students' evolutionary conceptual understanding. This consistency may be attributed to the systematic development of the instrument, including misconception identification, expert review, pilot testing, and empirical item evaluation. The findings support previous research showing that diagnostic instruments developed from documented alternative conceptions generally produce dependable measures of conceptual understanding. For example, Taşçi (2024) noted that multi-tier diagnostic assessments achieve stronger measurement quality when distractors are derived from scientifically established misconceptions rather than factual errors. Likewise, Desstya et al. (2025) and Ningroom et al. (2025) reported that carefully developed two-tier instruments provide reliable evidence of students' conceptual difficulties in science.

The item analysis further supports the quality of the EMT by demonstrating that the retained items adequately differentiate students with stronger conceptual understanding from those experiencing conceptual difficulties. These findings highlight the importance of constructing diagnostic assessments around students' documented misconceptions. When distractors represent common alternative conceptions, they provide meaningful

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

insight into students' reasoning processes rather than merely identifying incorrect responses. Consequently, the EMT extends beyond traditional achievement testing by revealing the nature of students' misconceptions.

The diagnostic findings indicate that misconceptions about evolution remain widespread among senior secondary school students. While a proportion of students demonstrated scientific understanding, many exhibited either partial understanding or persistent misconceptions, particularly regarding natural selection and adaptation. These findings suggest that many learners continue to interpret evolutionary processes through purpose-based or teleological explanations rather than scientifically accepted evolutionary mechanisms. This pattern is consistent with conceptual change theory, which posits that learners' prior conceptions are often resistant to conventional instruction and require explicit conceptual restructuring.

The persistence of teleological reasoning has been widely documented in evolution education research. Yates and Marek (2014) observed that students frequently explain evolutionary change by assuming that organisms intentionally develop traits needed for survival. Similarly, Nurdiansyah et al. (2022) found that students commonly interpret adaptation as purposeful adjustment by individual organisms, while Panagou et al. (2024) identified misconceptions about natural selection as one of the most persistent barriers to understanding evolution. The present findings extend this body of evidence by demonstrating that similar misconceptions are evident among Nigerian senior secondary school students.

The predominance of misconceptions related to natural selection and adaptation may reflect the abstract and counterintuitive nature of evolutionary processes. Unlike many observable biological phenomena, evolutionary change occurs gradually across generations through interactions among genetic variation, inheritance, environmental pressures, and differential reproductive success. As a result, students often rely on intuitive explanations derived from everyday experiences that conflict with scientific explanations of evolutionary change. Kuschmierz et al. (2020) similarly reported that learners experience considerable difficulty understanding population-level evolutionary processes, while Hartelt and Martens (2025) emphasized that students often struggle to move beyond individual-level reasoning toward population-based explanations.

A major contribution of the EMT lies in its two-tier diagnostic structure, which enables the identification of students whose selected answers do not accurately reflect their underlying reasoning. This overcomes an important limitation of conventional multiple-choice tests, where correct responses may result from guessing or incomplete understanding. By combining answer selection with justification, the EMT provides a more accurate representation of students' conceptual understanding and misconceptions. This finding agrees with Ristanto et al. (2023), who argued that multi-tier

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

diagnostic assessments generate richer evidence of students' conceptual understanding because they reveal the reasoning underlying learners' responses. Similarly, Taşçi (2024) emphasized that diagnostic assessments become more informative when they evaluate not only what students know but also why they hold particular conceptions.

The present study also addresses limitations identified in previous evolution assessment research. Existing instruments have frequently focused on evolutionary acceptance, isolated concepts, or university-level populations rather than diagnosing misconceptions among secondary school learners. Furrow and Hsu (2019) and Goff (2017) emphasized the need for assessment tools that evaluate specific dimensions of students' evolutionary understanding, while Kuschmierz et al. (2020) highlighted the importance of identifying learners' alternative conceptions rather than simply measuring achievement. By integrating curriculum-aligned evolutionary concepts within a two-tier diagnostic framework, the EMT provides a valid and reliable instrument for identifying misconceptions and supporting conceptual change instruction. Its application can assist biology teachers and researchers in designing targeted instructional interventions that promote meaningful understanding of evolution.

Limitations and Directions for Future Research

The study was limited to SS II Biology students in Zamfara State, Nigeria, which may restrict the generalizability of the findings. Future studies should validate the EMT with larger and more diverse samples across different educational settings. Confirmatory factor analysis is also recommended to verify the instrument's factor structure.

CONCLUSION

This study developed and validated the EMT as a two-tier diagnostic instrument for senior secondary school students. The findings showed satisfactory construct validity, good internal consistency (KR-20 = 0.84), and acceptable item characteristics, with misconceptions most prevalent in natural selection and adaptation. The EMT is therefore a valid, reliable, and diagnostically useful instrument for identifying evolutionary misconceptions. Its two-tier structure is effectively distinguishing scientific understanding from misconception-based reasoning and supports conceptual change instruction.

Recommendations

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

1. Biology teachers should use EMT as a diagnostic tool to identify students' misconceptions and guide instruction on evolution.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

2. Researchers should further validate the EMT across different geographical regions, educational levels, and student populations to establish its psychometric stability and generalizability.
3. Biology education researchers and curriculum developers should incorporate two-tier diagnostic assessments into biology instruction and assessment to improve the identification of students' misconceptions and support conceptual understanding.

ACKNOWLEDGEMENTS

The authors acknowledge the Tertiary Education Trust Fund (TETFund), Nigeria, for funding this study through the Institutional Based Research (IBR) Grant awarded to Federal University Gusau. Appreciation is also extended to the Zamfara State Ministry of Education and the participating secondary schools for their support during data collection.

Funding



This study was funded by the TETFund Nigeria, under the IBR Grant through Federal University Gusau (Grant No. TETF/DR&D/CE/UNI/GUSAU/IBR/2026/VOL1).

Conflicts of Interest

The authors declare no conflict of interest.

REFERENCES

Araújo, L. A. L. (2022). The centrality of evolution in biology teaching: Towards a pluralistic perspective. *Journal of Biological Education, 56*(1), 109–120. https://doi.org/10.1080/00219266.2020.1757486

Archila, P. A., Restrepo, S., Truscott de Mejía, A., & Rodríguez, J. M. (2024). STEM and non-STEM misconceptions about evolution: Findings from 5 years of data. *Science & Education, 33*, 1211–1229. https://doi.org/10.1007/s11191-023-00428-5

Arifah, M., Jusniar, J., & Anwar, M. (2022). Pengembangan instrumen tes diagnostik tiga tingkat untuk mengidentifikasi miskonsepsi peserta didik pada materi pokok laju reaksi. *Chemistry Education Review (CER), 6*(1). https://doi.org/10.26858/cer.v6i1.39494

Brownell, S. E., Freeman, S., Wenderoth, M. P., & Crowe, A. J. (2014). BioCore guide: A tool for interpreting the core concepts of *Vision and Change* for biology majors. *CBE—Life Sciences Education, 13*(2), 200–211. https://doi.org/10.1187/cbe.13-12-0233

Desstya, A., Sayekti, I. C., Abduh, M., & Sukartono. (2025). Development of a four-tier diagnostic test for misconceptions in natural science of primary school pupils. *Journal of Turkish Science Education, 22*(2), 338–353. https://doi.org/10.36681/tused.2025.017

Furrow, R. E., & Hsu, J. L. (2019). Concept inventories as a resource for teaching evolution. *Evolution: Education and Outreach, 12*, Article 2. https://doi.org/10.1186/s12052-018-0092-8

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Goff, K. D. (2017). *Validation of an instrument for assessing conceptual change with respect to the theory of evolution by secondary biology students* (Doctoral dissertation, William & Mary). https://doi.org/10.25774/w4-g3gv-m405

Hartelt, T., & Martens, H. (2025). How accurate are students in self-assessing their conceptions of evolution? *Science Education, 109*, 965–989. https://doi.org/10.1002/sce.21945

Isma'il, A. & Sodangi, U. (2025). Development of VARK Learning Styles Assessment Instrument for Secondary School Students. *Indonesian Journal of Education and Psychological Science, 3*(2), 183-200. https://doi.org/10.59890/ijeps.v3i2

Kuschmierz, P., Beniermann, A., & Graf, D. (2020). Development and evaluation of the knowledge about evolution 2.0 instrument (KAEVO 2.0). *International Journal of Science Education, 42*(15), 2601–2629. https://doi.org/10.1080/09500693.2020.1822561

Matazu, S. S., & Isma'il, A. (2024). Misconceptions about evolution in biology among secondary school students in Nigeria. *International Journal of Educational and Life Sciences, 2*(7), 789–806. https://doi.org/10.59890/ijels.v2i7.1642

Ningroom, R. A. A., Yamtinah, S., & Riyadi, R. (2025). A two-tier multiple-choice diagnostic test to find student misconceptions about the change of matter. *International Journal of Evaluation and Research in Education, 19*(2), 1160–1169. https://doi.org/10.11591/edulearn.v19i2.21478

Nurdiansyah, I., Hartati, S., & Paujiah, E. (2022). Analisis miskonsepsi siswa pada materi evolusi sub-bab adaptasi menggunakan metode *three-tier test*. *Jurnal Literasi Pendidikan Dasar, 3*(2), 29–39. https://doi.org/10.36928/jlpd.v3i2.2068

Panagou, D., Kostara, C., Dimos, E., Stylos, G., & Kotsis, K. (2024). Honors high school graduates students' misconceptions regarding evolutionary theory of biology. *EIKI Journal of Effective Teaching Methods, 2*(3). https://doi.org/10.59652/jetm.v2i3.188

Ristanto, R. H., Suryanda, A., & Indraswari, L. A. (2023). The development of ecosystem misconception diagnostic test. *International Journal of Evaluation and Research in Education, 12*(4), 2029–2038. https://doi.org/10.11591/ijere.v12i4.25200

Sá-Pinto, X., Realdon, G., Torkar, G., Sousa, B., Georgiou, M., Jeffries, A., Korfiatis, K., Paolucci, S., Pessoa, P., Rocha, J., Stasinakis, P. K., Cavadas, B., Crottini, A., Gnidovec, T., Nogueira, T., Papadopoulou, P., Piccoli, C., Barstad, J., Dufour, H. D., ... Mavrikaki, E. (2021). Development and validation of a framework for the assessment of school curricula on the presence of evolutionary concepts (FACE). *Evolution: Education and Outreach, 14*, Article 3. https://doi.org/10.1186/s12052-021-00142-2

Soeharto, S., & Csapó, B. (2022). Exploring Indonesian student misconceptions in science concepts. *Heliyon, 8*(9), e10720. https://doi.org/10.1016/j.heliyon.2022.e10720

Taşçi, G. (2024). Development of a two-tier diagnostic test to assess misconceptions about biology concepts at primary school. *Science Insights Education Frontiers, 25*(2). https://doi.org/10.15354/sief.24.or652

Torkar, G., & Gregorčič, T. (2026). Evolution education in Slovenia: Using the framework for the assessment of school curricula on the presence of evolutionary concepts (FACE) in curriculum reform. *EvoDevo, 17*, Article 6. https://doi.org/10.1186/s13227-026-00260-2

Woods, C. S., & Scharmann, L. C. (2001). High school students' perceptions of evolutionary theory. *The Electronic Journal for Research in Science & Mathematics Education, 6*(2). https://ejrsme.icrsme.com/article/view/729

Yates, T. B., & Marek, E. A. (2014). Teachers teaching misconceptions: A study of factors contributing to high school biology students' acquisition of biological evolution-related misconceptions. *Evolution: Education and Outreach, 7*, Article 7. https://doi.org/10.1186/s12052-014-0007-2

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

Appendix A

Evolutionary Misconception Test (EMT)

Instruction: For each item**, circle one answer in Tier 1 and one reason in Tier 2**. A score of **1** is awarded only when both responses are correct.

No.	Concept	Tier 1	Tier 2
1	Natural selection	***Natural selection occurs when:*** **A.** Traits develop by need **B.** Favourable individuals reproduce **C.** Organisms choose traits **D.** Environment creates traits	**A.** Environment produces traits **B.** Effort changes traits **C.** Need creates traits **D.** Inherited variation improves survival
2	Natural selection	***Antibiotic resistance occurs because:*** **A.** Need **B.** Antibiotics create resistance **C.** Resistant bacteria survive **D.** Bacteria choose resistance	**A.** Antibiotics change genes **B.** Bacteria learn resistance **C.** Resistance already exists and is inherited **D.** All bacteria become resistant
3	Evolution	***Evolution involves changes in:*** **A.** Individuals **B.** Populations **C.** Needs **D.** Environment	**A.** Environment changes organisms directly **B.** Individuals evolve during life **C.** Need causes evolution **D.** Populations change genetically over generations
4	Natural selection	***Giraffes have long necks because:*** **A.** Stretching **B.** Longer-necked individuals reproduced **C.** Need **D.** Environment lengthened necks	**A.** Acquired traits are inherited **B.** Selection favoured inherited variation **C.** Organisms choose traits **D.** Environment creates traits
5	Genetic variation	***Variation exists because:*** **A.** Need **B.** Genetic differences **C.** Environment **D.** Choice	**A.** Mutation and reproduction generate variation **B.** Traits are created by need **C.** Environment makes genes **D.** Individuals control genes
6	Natural selection	***Natural selection acts on:*** **A.** Needs **B.** Individual variation **C.** Evolutionary goals **D.** Environmental change	**A.** Individuals with favourable traits reproduce more **B.** Traits develop intentionally **C.** Nature creates traits **D.** Individuals choose traits
7	Adaptation	***An adaptation is:*** **A.** Inherited survival trait **B.** Lifetime change **C.** Need-based trait **D.** Temporary response	**A.** Environment creates traits **B.** Traits develop by effort **C.** Adaptations evolve through selection **D.** Acquired traits are inherited
8	Adaptation	***Desert plant thick stems are:*** **A.** Need-based **B.** Drought-induced **C.** Inherited adaptation **D.** Chosen by plants	**A.** Selection favours inherited traits **B.** Plants change themselves **C.** Environment creates traits **D.** Plants control adaptation
9	Adaptation	***Adaptation occurs through:*** **A.** Effort **B.** Environmental demand **C.** Natural selection **D.** Individual choice	**A.** Individuals alter genes **B.** Organisms develop needed traits **C.** Species choose traits **D.** Beneficial variation accumulates
10	Adaptation	***Different bird beaks evolved because of:*** **A.** Need **B.** Environment changed all birds **C.** Use **D.** Suitable beaks improved reproduction.	**A.** Traits are acquired **B.** Selection favoured inherited beak variation **C.** Environment creates beaks **D.** Birds direct evolution
11	Adaptation	***Traits disappear because:*** **A.** No longer needed **B.** They reduce reproductive success **C.** Organisms remove them **D.** Environment destroys them	**A.** Selection changes trait frequency **B.** Organisms decide traits **C.** Traits disappear intentionally **D.** Environment removes traits
12	Genetic variation	***Variation results from:*** **A.** Need **B.** Choice **C.** Environment **D.** Mutation and reproduction	**A.** Environment produces traits **B.** Organisms create genes **C.** Variation enables selection **D.** Individuals control genes

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

No.	Concept	Tier 1	Tier 2
13	Genetic variation	***Siblings differ because:*** **A.** Need **B.** Environment **C.** Different gene combinations **D.** Choice	**A.** Sexual reproduction reshuffles genes **B.** Individuals modify inheritance **C.** Environment changes genes **D.** Organisms choose genes
14	Genetic variation	***Mutation contributes by:*** **A.** Producing useful traits only **B.** Creating variation **C.** Occurring by need **D.** Occurring only in stress	**A.** Mutations introduce new variation **B.** Mutations are intentional **C.** Environment causes mutations **D.** All mutations are beneficial
15	Genetic variation	***Genetic variation:*** **A.** Appears after change **B.** Exists before selection **C.** Is created by need **D.** Does not exist	**A.** Individuals control genes **B.** Organisms create variation **C.** Environment creates identical traits **D.** Selection acts on existing variation
16	Genetic variation	***Genetic differences matter because:*** **A.** Organisms choose traits **B.** They provide material for selection **C.** They occur only by need **D.** They make organisms identical	**A.** Environment makes traits identical **B.** Traits develop by need **C.** Selection acts on inherited differences **D.** Individuals control genes
17	Common ancestry	***Species similarities suggest:*** **A.** Independent creation **B.** Common ancestry **C.** Survival needs **D.** Identical genes	**A.** Species choose evolution **B.** Species copy each other **C.** Environment creates identical forms **D.** Shared traits can be inherited
18	Common ancestry	***Humans and chimpanzees are similar because:*** **A.** Humans evolved from chimpanzees **B.** Shared ancestor **C.** Chimpanzees became humans **D.** They have always been identical	**A.** Related species inherit traits from ancestors **B.** One modern species becomes another **C.** Species stop evolving **D.** Need drives change
19	Common ancestry	***Fossils show that:*** **A.** Species changed over time **B.** Species never changed **C.** Species appeared suddenly **D.** Organisms changed intentionally	**A.** Fossils reveal historical change **B.** Fossils show no change **C.** Fossils show only living species **D.** Fossils prove intention
20	Common ancestry	***Homologous structures indicate:*** **A.** Same function **B.** Common ancestry **C.** Need-based change **D.** Independent creation	**A.** Effort produces similarity **B.** Environment creates similarity **C.** Shared structures are inherited **D.** Structures lack evolutionary history
21	Common ancestry	***DNA similarity indicates:*** **A.** Common ancestry **B.** Choice **C.** No relationship **D.** Environmental control	**A.** Organisms modify DNA by need **B.** Shared DNA reflects inheritance **C.** Environment creates DNA **D.** Species control DNA
22	Speciation	***A new species forms when:*** **A.** Organisms need change **B.** Populations become reproductively isolated **C.** Individuals decide **D.** Environment creates species	**A.** Organisms intentionally change **B.** Isolation allows genetic divergence **C.** Environment produces species **D.** Individuals evolve during life
23	Speciation	***Geographic isolation may cause speciation because:*** **A.** Populations evolve separately **B.** Separation changes organisms directly **C.** Isolation makes populations identical D. Geography creates traits	**A.** Restricted gene flow allows divergence **B.** Organisms choose adaptation **C.** Environment forces identical change **D.** Inheritance is unnecessary
24	Speciation	***Two populations become different species when they:*** **A.** Have different needs **B.** Cannot interbreed successfully **C.** Live apart **D.** Look different	**A.** Environment stops reproduction **B.** Appearance defines species **C.** Species decide to separate **D.** Reproductive isolation prevents gene exchange

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]

No.	Concept	Tier 1	Tier 2
25	Speciation	***New species evolve through changes in:*** **A.** Individuals **B.** Populations over generations **C.** Intentions **D.** Environment alone	**A.** Species choose evolution **B.** Individuals transform into species **C.** Genetic change accumulates in populations **D.** Environment alone creates species
26	Evidence for evolution	***Fossils support evolution because they show:*** **A.** Sudden appearance **B.** No change **C.** Change through time **D.** Intentional change	**A.** Fossils represent only living species **B.** Fossils show constancy **C.** Fossils document historical change **D.** Fossils prove intention
27	Evidence for evolution	***Similar body structures indicate:*** **A.** Environment alone **B.** Copying **C.** Choice **D.** Common ancestry	**A.** Organisms copy traits **B.** Shared structures reflect inheritance **C.** Environment creates structures **D.** Species are unrelated
28	Evidence for evolution	***Similar DNA sequences indicate:*** **A.** No inheritance **B.** Choice **C.** Shared ancestry **D.** Environment creates DNA	**A.** DNA similarity reflects common ancestry **B.** Species alter DNA intentionally **C.** Genes develop by need **D.** Habitat creates genes
29	Evidence for evolution	***Species distribution supports evolution because:*** **A.** It reflects evolutionary history **B.** Organisms choose habitats **C.** All species originated together **D.** Environment creates identical species	**A.** Environments produce identical species **B.** Organisms choose locations **C.** Species never change **D.** Isolation influences evolution
30	Evidence for evolution	***Evidence for evolution includes:*** **A.** Environment alone **B.** Need and intention **C.** Lifetime changes **D.** Fossils, DNA and comparative anatomy	**A.** Multiple independent lines of evidence support evolution **B.** Organisms consciously change **C.** Acquired traits are inherited **D.** Evolution depends only on environment

Appendix B

Answer Key for the Evolutionary Misconception Test (EMT)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Tier 1	B	C	B	B	B	B	A	C	C	D	B	D	C	B	B
Tier 2	D	C	D	B	A	A	C	A	D	B	A	C	A	A	D

Item	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Tier 1	B	B	B	A	B	A	B	A	B	B	C	D	C	A	D
Tier 2	C	D	A	A	C	B	B	A	D	C	C	B	A	D	A

Appendix C

Scoring Guide for the Evolutionary Misconception Test (EMT)

1. **Award 1 mark** only when **both Tier 1 (T1)** and **Tier 2 (T2)** responses are correct.
2. **Award 0 mark** if:
   - Tier 1 is correct but Tier 2 is incorrect;
   - Tier 1 is incorrect but Tier 2 is correct; or
   - Both Tier 1 and Tier 2 are incorrect.
4. **No partial credit** is awarded.
5. The **maximum obtainable score** is **30**, while the **minimum obtainable score** is **0**.
6. **Higher scores** indicate better conceptual understanding of evolution and fewer misconceptions, whereas **lower scores** indicate a greater prevalence of evolutionary misconceptions.

EDITOR-IN-CHIEF
Prof. A.I. Mteiye [MALAWI]