DIGITAL TRANSFORMATION IN EDUCATION: PERCEPTION, ACTION AND PROSPECT

Proceedings of the 3rd International Conference
on Innovation in Learning Instruction and Teacher Education – ILITE 3

ĐỔI MỚI SÁNG TẠO TRONG DẠY HỌC VÀ ĐÀO TẠO GIÁO VIÊN

Chủ đề: Chuyển đổi số trong giáo dục: nhận thức, hành động và triển vọng

NHÀ XUẤT BẢN ĐẠI HỌC SƯ PHẠM

USING AN ONLINE THREE-TIER DIAGNOSTIC TEST TO IDENTIFY HIGH SCHOOL STUDENTS' MISCONCEPTIONS ABOUT EVOLUTION

Pham Thi Phuong Anh*[*1], Phan Duc Duy[2], Nguyen Thi Dieu Phuong[2]

Abstract: *Evolution in the high school Biology program contains a wealth of knowledge closely related to real life, but it proves to be challenging for students to explain scientifically. This gives rise to alternative conceptions that differ from scientific concepts, called misconceptions. These misconceptions can impede students' understanding of Evolution. Therefore, teachers should implement measures to identify students' misconceptions before teaching Evolution to help students overcome these barriers. One valuable diagnostic tool is the three-tier diagnostic test, developed from two-tier tests, incorporating a certainty response index as a third tier for each item. This addition helps differentiate between a lack of knowledge and a misconception. The development of digital technology has led to the creation of numerous tools that teachers can employ for diagnostic assessments. This study proposed using an online three-tier diagnostic test to identify students' misconceptions about Evolution. The test was administered to 12th-grade students in Thua Thien Hue province, Vietnam. Item analysis used Cronbach's alpha reliability, item difficulty, and point biserial correlation coefficients. Consequently, ten misconceptions held by students about Evolution were identified. This research equips teachers with the means to recommend appropriate strategies for helping students rectify their misconceptions about Evolution. It also guides them in utilizing online three-tier diagnostic tests to identify other misconceptions students hold in Biology.*

Keywords: *Diagnostic assessments, three-tier diagnostic tests, misconceptions, Evolution*

INTRODUCTION

Educational researchers acknowledge that students' prior knowledge profoundly affects learning, even when this knowledge may be incorrect. Students' existing knowledge has been referred to in various terms such as misconceptions (Doran, 1972; Helm, 1980; Ola Adeniyi, E., 1985), preconceptions/prior knowledge (Novak, 1966), alternative conceptions (Krasner, 1984) or naïve conceptions (Champagne, 1983). Uzuntiryaki and Geban (2005) stated that science teachers should notice students' prior knowledge and misconceptions, examine how students learn and why misconceptions occur, and use appropriate teaching methods to eliminate misconceptions. They also suggested that science teaching should promote conceptual change. Conceptual change is one of the teaching approaches in science that is important in facilitating meaningful learning and preventing misconceptions. In this approach, students' ideas and views should be identified first. Then teachers create opportunities for students to explore their ideas, provide stimuli for students to develop, modify and, where necessary, change their ideas and views, and support their attempts to rethink and reconstruct them. The Biology general educational program of the Ministry of Education and Training in 2018 has also identified one of the indicators of the biological cognitive competency of high school students as "recognizing and correcting the mistakes; making critical statements regarding the topic in discussion" (MOET, 2018, p6). It proves that teaching Biology, besides focusing on forming new knowledge for students, also needs to pay attention to students' Biology misconceptions.

According to Morrison & Lederman (2003), teachers should know familiar students' misconceptions of science. Generally, senior teachers with many years of professional

[*] Corresponding Author: Pham Thi Phuong Anh; Email Address: ptpanh@hueuni.edu.vn
[1] University of Education
[2] Hue University

experience will likely list more students' misconceptions than new teachers. However, it is more beneficial if all teachers possess knowledge of student misconceptions from the outset and continue to collect misconceptions of individual students. Therefore, identifying and listing common scientific misconceptions is necessary to help teachers in teaching science.

Among the various diagnostic tests to reveal students' misconceptions, the three-tier test is highly appreciated and increasingly employed. Since 2020, due to the impact of the COVID-19 pandemic, numerous high schools in Vietnam have invested in additional facilities to facilitate online teaching, adapting to the prevailing circumstances. Teachers have also received training and gained fresh insights into utilizing digital tools for instruction. Consequently, this presents an excellent opportunity to develop an online three-tier diagnostic test. This research aimed to develop an online three-tier test as a valid and reliable diagnostic tool to identify misconceptions about Evolution and assess the prevalence of these misconceptions among 12th-grade students in Thua Thien Hue province, Vietnam. From there, furnish teachers with instructional materials to restructure students' misconceptions in the light of their experience, ultimately enhancing the efficacy of Evolution teaching in high schools.

CONTENT

1. Literature review

1.1. Misconception

The idea that students develop misconceptions has been a focus of various research on learning mathematics and science since the 1970s. This focus emerged when educational researchers began to notice what students were saying and doing when they learned. Piaget's repeated demonstrations revealed that children perceive the world quite differently from adults. These studies unveiled that students did not approach instruction as blank slates; instead, they had developed durable conceptions with explanatory power, but most conflicted with the accepted mathematical and scientific concepts presented in instruction (Smith III, 1994). Allen (2014) supposed that students gained that knowledge from educational experiences or informal events.

Various definitions of misconceptions have been developed in educational research. Sanger and Greenbowe considered misconception as "student conceptual and propositional knowledge that is inconsistent with or different from the commonly accepted scientific consensus and is unable to adequately explain observable scientific phenomena." (Sanger and Greenbowe, 1997). This particular definition of misconception serves as the foundation for identifying students' misconceptions about Evolution in high schools.

1.2. An overview of diagnostic tests to identify students' misconceptions in science

The commonly used diagnostic tests are interviews, simple multiple-choice tests, multiple-tier tests (two-tier, three-tier, and four-tier), and open-ended tests. Each kind of test has benefits and limitations in assessing students' misconceptions.

Interviews and open-ended tests give students the freedom to express their ideas. Researchers may get detailed information about students' cognitive knowledge structures and get an unlimited range of answers given by students. However, both need significant time, and the researchers must have some assessment techniques to get better results. These methods are also only applied to a small number of students. Moreover, some students may need to be more confident and active to do the tests. Researchers also may need help in interpreting and analyzing student answers.

Multiple-choice tests come as the solution to assess students' misconceptions with many more participants. These tests enable researchers to cover various topics in a relatively short amount of time. They also provide quick and objective scoring of students' responses. Otherwise, constructing well-constructed multiple-choice items for these tests requires effort.

With this approach, students are unable to construct, organize, and interpret their understanding, making it more challenging to figure out other students' ideas to suggest the appropriate solutions for adjusting their misconceptions.

Within the realm of multiple-choice tests, simple multiple-choice tests exhibit several weaknesses. Firstly, guessing can cause errors in variances and compromise reliability. Secondly, students might occasionally provide correct answers for the wrong reasons. In other words, multiple-choice tests cannot distinguish the correct answers based on genuine understanding, and those arrived at through incorrect reasoning. This can lead to inaccuracies when evaluating student misconceptions.

Multiple-tier tests were developed as a response to the limitations of simple multiple-choice tests, and they include formats like two-tier, three-tier, and other variations of multiple-tier tests**.**

The two-tier tests are diagnostic tools, featuring a first tier that resembles simple multiple-choice questions and a second tier comprising reasons that align with the choices presented in the first tier. Students' responses are considered valid only when both tiers' choices are correct. This testing approach allows for measuring students' reasoning and its connection to answers linked to misconceptions. It could reduce guesses and give explanations regarding student reasoning. However, the presence of guessing may result due to overestimating the participants' knowledge levels as well as misconceptions, as these tests do not discriminate lack of knowledge from misconceptions.

Limitations appearing in two-tier tests encourage researchers to develop three-tier tests. Another additional tier, which contains a certainty of response index, has been proposed to compensate for the likely weakness of the diagnostic tests. Each item now has three tiers: the first tier is the simple multiple-choice, the second tier includes the possible reasons that correlate with the given answer in the first tier, and the third tier consists of confidence options for the first two tiers. Caleon and Subramaniam (2010) stated that "students tended to be poorly discriminating between what they know and what they do not know; confidence ratings may reflect the strength of students' conceptual understanding, as well as their alternative conceptions" (p. 941). Their study also indicated that students with high confidence in their incorrect responses represent tenacious misconceptions.

With the three-tier tests, researchers can conclude students' perceptions of the posed questions based on combinations of students' responses across all three tiers, as shown in Table 1.

Table 1. The conclusions from the specified combinations of students' responses across all three tiers (Adapted from Aslan, 2012)

Situation	First tier	Second tier	Third tier	Conclusion
1	Correct	Correct	Yes	Scientific knowledge
2	Incorrect	Incorrect	Yes	Misconception
3	Correct	Incorrect	Yes	Misconception (false positive)
4	Incorrect	Correct	Yes	Misconception (false negative)
5	Correct	Incorrect	No	Lack of knowledge
	Incorrect	Correct		
	Incorrect	Incorrect		
6	Correct	Correct	No	Lucky guess or lack of confidence

False negatives and false positives are terms used to describe assessment errors in scientific research. A false positive occurs when a correct answer is accompanied by an incorrect reason, suggesting the presence of an effect that is actually absent. Conversely, a false negative entails an incorrect answer accompanied by a valid reason, failing to identify a present effect. As shown in Table 1, students can be deemed to possess scientific knowledge only when both their responses in the first and second tiers are accurate and exhibit confidence in the third tier.

Similarly, students indeed hold misconceptions if they select both incorrect answers in the first tier and second tier and exhibit confidence in the third tier. Other combinations of responses across all three tiers (correct/incorrect/uncertain, incorrect/correct/uncertain, and incorrect/incorrect/uncertain) for each item were treated as a "lack of knowledge." By adding a confidence tier in each item, the three-tier test can distinguish a lack of understanding from a misconception. This differentiation also helps determine whether students possess genuine scientific knowledge or if it is just a lucky guess. Therefore, three-tier tests are considered to provide more accurate results compared to two-tier tests.

Only a few studies employed more than three tiers within each item, such as four or five tiers, with specific enhancements. Although these additional tiers can address the issues encountered in three-tier tests, it is essential to note that the testing process might require significant time (Soeharto, 2019). Soeharto (2019) also clarified that the utilization of multiple-choice and multi-tier tests for assessing misconceptions has increased since 2015. Multiple-tier tests remain the most commonly used tool in the present study for identifying misconceptions. However, employing four-tier multiple-choice tests to diagnose misconceptions remains relatively rare in research. Moreover, other researchers have adopted a combination of diverse diagnostic methods to gather data related to student misconceptions, including open-ended tests, interviews, and multiple-tier tests, all aimed at achieving more comprehensive and accurate results.

2. The development of an online three-tier test to identify students' misconceptions about Evolution

Research using three-tier tests in teaching Biology is still relatively uncommon compared to other science subjects. Based on 111 articles published from 2015 to 2019, Soeharto (2019) concluded that multiple-choice biology tests were less frequent compared to chemistry and physics subjects. In Vietnam, there has been limited attention from researchers toward studying students' misconceptions and employing diagnostic tests to identify these misconceptions in learning. Initial research into misconceptions has been conducted on the subject of Physics by authors such as Trinh et al. (2017) and Vo et al. (2019), but even these studies did not explore the use of three-tier diagnostic tests for identifying misconceptions.

The topic of Evolution poses challenges within the high school Biology program. Numerous studies have investigated learners' misconceptions regarding Evolution. Noteworthy among these studies are the works of Nehm (2007), Cunningham (2009), and Yates and Marek (2015). These authors have employed various assessment methods to identify misconceptions, including Likert scale questionnaires and open-response instrument essay questions. Regrettably, there are only a few rare instances of research that have employed three-tier tests to identify students' misconceptions about Evolution, with the research conducted by Putri (2017) being one such example.

Some researchers have proposed procedures for developing diagnostic three-tier tests, including both traditional and online formats. Notable examples are the procedures introduced by Saat et al. (2016) and Hasyim et al. (2018). Saat's procedure for developing the online diagnostic tool to identify misconceptions in cellular respiration consisted of five phases: (1) construction of items, (2) pilot study, (3) validation of the diagnostic instrument, (4) transformation of the three-tier diagnostic instrument into an online assessment tool, and (5) administration of the online diagnostic tool. Meanwhile, Hasyim's procedure, used to identify misconceptions in the Human Reproduction System section, consists of only three phases: (1) defining content, (2) obtaining information, and (3) developing a diagnostic test. Based on Saat's and Hasyim's procedures, the procedure to build the Online Evolutionary Diagnostic Test of Knowledge (OEDTK) was developed with five main phases: (1) making a list of possible evolutionary misconceptions, (2) constructing items, (3) embedding the three-tier multiple-

choice items into an online tool, (4) conducting a pilot study, (5) revising and completing the OEDTK.

Phase 1: Make a list of possible Evolutionary misconceptions

Common evolutionary misconceptions were collected through teaching experiences and various sources, including reference books and online teaching materials. Based on this information, a list of potential evolutionary misconceptions was compiled. This list was subsequently shared with six teachers with extensive experience teaching evolution in high schools in Thua Thien Hue to assess its appropriateness. Out of the list, ten misconceptions that were deemed most relevant were then selected to construct three-tier multiple-choice items. These misconceptions correlate with scientific concepts from the Biology 12 textbook (Nguyen et al., 2012) and can be found in Table 2.

Table 2. A list of possible Evolutionary misconceptions

Conceptions	Scientific conceptions	Misconception
1. Homology	Homology describes organs in different animals with a common evolutionary origin but may have various functions.	Homology describes organs in different animals that have various functions.
2. Antibiotic resistance	Penicillin-resistant mutations existed at low frequency in the bacteria population, and penicillin changed the frequency of Penicillin-resistant mutations.	Penicillin caused new Penicillin-resistant mutations in the bacteria population.
3. Polyploid Speciation	A polyploid mutant population is considered a new species.	An individual polyploid mutant is considered a new species.
4. The founder effect	The founder effect is considered a genetic drift.	The founder effect is considered the gene flow.
5. Fossils	Fossils are preserved remains, impressions, or traces of once-living things from a past geological age.	Fossils are the petrifications of once-living things from a past geological age.
6. Raw materials for evolution	Inheritable variations	All kinds of variations.
7. Adaptive trait	An adaptive trait is a genetic trait that cannot change during life.	Any trait that helps an organism maximize its survival success.
8. Geographical isolation	Geographical isolation is geographical obstacles that prevent individuals of populations of the same species from interbreeding.	Geographical isolation includes ecological isolation.
9. The rate of adaptation	The rate of adaptation depends on the generation and accumulation of mutations, the rate of reproduction in each species, and the pressure of natural selection.	The higher the evolutionary level of the species, the faster the rate of adaption formation.
10. Human evolution	Both cultural and biological factors play a significant role in human evolution.	Cultural factors create all changes in human evolution.

Phase 2: Constructing of items

From the list of potential misconceptions, three-tier multiple-choice items were developed. It became evident that many students struggled to distinguish between two certain concepts. Thus, in numerous cases, including more than two options appeared unnecessary. Instead, we fashioned items featuring two options in the first two tiers and added a blank space for students to provide answers and explanations. This approach allowed us to delve deeper into students' reasoning and interpretations of evolutionary misconceptions. Our approach aligns with the recommendations of other researchers who have employed three-tier tests (Eryılmaz, 2010; Pesman & Eryilmaz, 2010; Arslan, 2012; Kanli, 2015).

With these new enhancements, each item in our three-tier test can be illustrated as the sample item in Figure 1.

Question 10:
[First tier] The human brain is more developed than other animals as a result of
A. The process of labor and thinking helps perfect the brain's structure.
B. the process of accumulation of variations under the influence of natural selection.
C. Other (please specify):
[Second tier] The reason I chose this option is
A. the process of labor and thinking sets humans apart from other animals.
B. Brain development is an adaptive feature of humans to the environment.
C. Other (please specify):
[Third tier] Are you confident about the option you have chosen?
A. Yes.
B. No.

Figure 1. The sample of an item in the OEDTK

Phase 3: Embedding the three-tier multiple-choice items into an online tool

The items constructed in Phase 2 need to be integrated into an online tool. Google Forms was chosen as a platform due to its numerous advantages. This user-friendly platform offers unrestricted access to all users, enabling students to easily take the test. Furthermore, the responses are automatically stored in a shareable spreadsheet, streamlining the data collection process. Additionally, Google Forms provides rapid statistical insights into students' answer choices. This feature allows researchers to promptly calculate results using spreadsheet data and subsequently analyze the collected information. Moreover, Google Forms accommodates the inclusion of an alternative blank response, aligning well with the modifications we made to our three-tier test.

Phase 4: Conducting a pilot study

The pilot version of the OEDTK was developed and tested on a sample of 78 students in a high school in Thua Thien Hue. The data were then input into the SPSS program and dichotomized using an answer key. Alternative answers were categorized according to the provided options for the first two tiers. Responses that were unclear or vague were coded as incorrect answers. Eight scores were calculated, namely the first tier score, second tier score, third tier score, both tiers score, all tiers score, misconception first tier score (M-first tier score), misconception both tiers score (M-both tiers score), and misconception all tiers score (M-all tiers score), following the approach described by Pesman and Eryilmaz (2010). Percentages were computed for each of these scores. Table 3 provides an overview of the methodology used to calculate these scores.

Table 3. The calculation of the score and M-score (Adapted from Pesman and Eryilmaz, 2010)

Tier of item	Description	Score	M-score
First-tier [FT]	Only the first-tier scores are considered.	Correct: 1; Incorrect: 0.	Incorrect: 1; Correct: 0.
Second-tier [ST]	Only the second-tier scores are considered.	Correct: 1; Incorrect: 0.	
Third-tier [TT]	Only the third-tier score is considered.	Yes: 1; No: 0.	
Both tiers [BT]	The first- and second-tier scores are considered.	Both tiers are correct: 1; Other: 0.	Both tiers are incorrect: 1; Other: 0.
All tiers [AT]	All three tiers' scores are considered.	Both first tiers are correct, and "yes" is in the third tier: 1; Other: 0.	Both tiers are incorrect, and "yes" is in the third tier: 1; Other: 0.

Cronbach's alpha reliability, item difficulties, and point biserial correlation coefficients were employed for item analysis. The Cronbach's alpha coefficient for achievements measured by the AT score must exceed 0.7. Item difficulty represents the percentage of respondents who answered the item correctly. Difficulty levels are indicated on a scale of 0 to 1, where an index value > 0.90 corresponds to straightforward questions, and an index value < 0.30 indicates challenging questions. The assessment of item discrimination, achieved by comparing correct responses with overall scores, involved the use of the point biserial correlation coefficient. Only questions having a discrimination index > 0.20 are considered (Milenkovic et al., 2016; Hasyim et al., 2018).

The pilot study's results indicated that Cronbach's alpha is 0.681. Although all the questions are acceptable, Question 1 shows a relatively low difficulty level (with a difficulty score of 0.795) and an item discrimination value of 0.278. Adjusting this question's difficulty to a moderate level is necessary for better results.

Additionally, student interviews were conducted to evaluate the usability and user-friendliness of the OEDTK after the pilot test. Most students agreed that the instructions in the OEDTK were clear and comprehensible. Regarding user-friendliness, most students found the OEDTK easy to navigate. In general, students provided positive feedback about the OEDTK.

Phase 5: Revising and completing the OEDTK

Based on the pilot study's findings, specific alternative answers provided by students in the first two tiers were noted and considered to create improved distractors for three items. Furthermore, one item underwent refinement based on item analysis results. The second version of the OEDTK was developed after a comprehensive review process and several revisions. This final version comprised ten three-tier multiple-choice items used for the official testing.

3. Methodology

3.1. Sample

Testing was performed at six high schools in April 2023 in Thua Thien Hue province, Viet Nam. All students are in grade 12 and finished learning Evolution three months ago. The sample total has 391 students who had accepted to participate in the study voluntarily. Population by gender was 52,43% females and 47,57% males.

3.2. Data collection and analysis

Students finished the test in class on their smartphones within 20 minutes with the teacher's supervision. Students' results were collected rapidly from the spreadsheet. The data was analyzed by Microsoft Office Excel and IBM SPSS Statistics v.25. All the statistical calculations similar to the pilot study were performed, including the reliability and the items difficulty analysis.

4. Result and discussion

The descriptive statistical analysis results for the all-tier score (AT) reveal that the maximum score attained was 10, while the minimum was 0. The calculated mean value stands at 4.15 out of 10, signifying an average score within the moderate category. The reliability analysis, utilizing Cronbach's alpha coefficient, yields a value of 0.743, indicating a relatively high level of internal consistency among the items.

Regarding the AT scores, the item difficulty analysis has indicated two items with a difficulty index below 0.30 (Item 8 and Item 10), characterizing them as challenging tasks. The remaining tasks all fall within the range of 0.30 to 0.60, representing tasks of moderate difficulty. The outcomes of this analysis demonstrate that all ten items exhibit a discrimination index in the range of 0.477 to 0.626, surpassing the threshold of 0.20. This result shows that the test effectively distinguishes between students with a solid and weak conceptual grasp. Further details are provided in Table 4.

Table 4. Index of Difficulties and Index of Discrimination of AT score

Items	1	2	3	4	5	6	7	8	9	10
Index of Difficulties	0.537	0.432	0.534	0.312	0.481	0.468	0.304	0.248	0.570	0.263
Index of Discrimination	0.554	0.589	0.572	0.477	0.626	0.577	0.547	0.518	0.511	0.517

The FT, BT, and AT scores were computed using the data from Table 3. The analysis of correct answers (tiers 1 and 2) and affirmative responses (tier 3) for all items involved comparing student achievements observed in the FT, BT, and AT scores. Figure 2 demonstrates the percentages of correct/yes responses across tiers. Across all items, the difficulty index increases as tiers increase. The average achievement values gradually decrease as tiers increase (FT: 5.23; BT: 4.53; AT: 4.15). The difference between the scores obtained for FT and BT is approximately 0.7, which could be attributed to false positives. On the other hand, the disparity in the mean value achieved between BT and AT was slightly under 0.4. This difference could be explained by a lucky guess or a lack of confidence. Consequently, it becomes apparent that the proportion of students possessing genuine scientific knowledge and confidence in their answers is significantly lower than those who merely select the correct answer in a simple multiple-choice question.

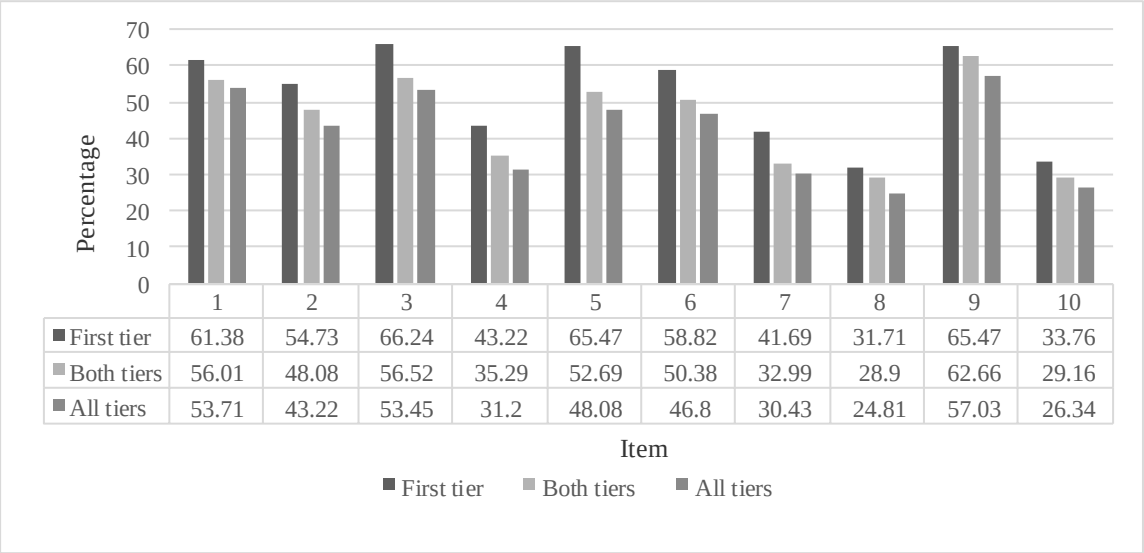
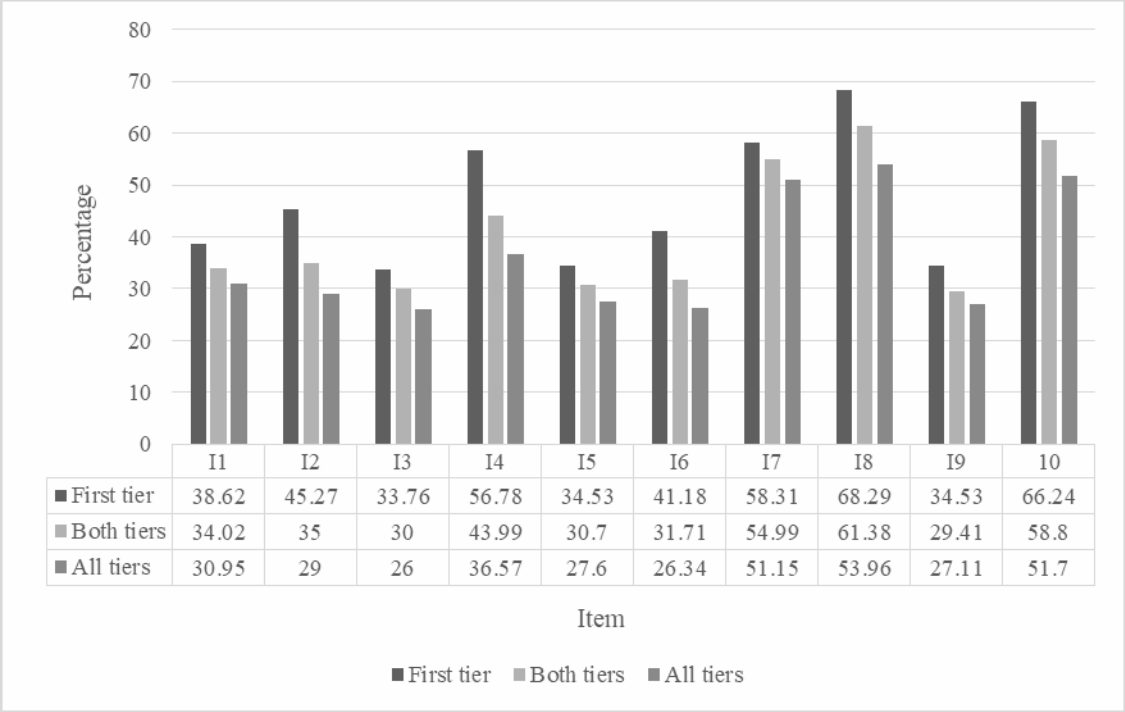


Figure 2. Percentages of correct responses regarding the number of tiers

Likewise, Figure 3 illustrates the percentages of misconceptions corresponding to the number of tiers. Across all items, the percentage of misconceptions decreases as the number of tiers increases.



	I1	I2	I3	I4	I5	I6	I7	I8	I9	10
First tier	38.62	45.27	33.76	56.78	34.53	41.18	58.31	68.29	34.53	66.24
Both tiers	34.02	35	30	43.99	30.7	31.71	54.99	61.38	29.41	58.8
All tiers	30.95	29	26	36.57	27.6	26.34	51.15	53.96	27.11	51.7

Figure 3. Percentages of misconceptions regarding the number of tiers

Some students might have provided incorrect answers in the first tier without necessarily holding misconceptions; these errors could be attributed to false negatives or a lack of knowledge. Both two-tier tests and three-tier tests can reveal the percentages of false negatives. However, unlike two-tier tests, three-tier tests can also detect cases of lack of knowledge, which is discernible through the third tier. As anticipated, the average percentages of misconceptions decreased with the increase in tiers (M-FT: 47.75%; M-BT: 41.05%; M-AT: 36.01%). The difference between the mean percentages of the one-tier and the two-tier tests was 6.6%, possibly because even some incorrect answers for the first tier were due to a false negative. As anticipated, the average percentage of misconceptions decreased with the increase in tiers (M-FT: 47.75%; M-BT: 41.05%; M-AT: 36.01%). Furthermore, the figure also reveals that the three-tier test is the most effective in accurately capturing the actual percentages of student misconceptions.

Table 5 presents an analysis of students' conceptual knowledge based on their responses across all three tiers. The mean score for scientific knowledge is generally the highest, demonstrating the effectiveness of teaching Evolution in high schools within Thua Thien Hue province. Nevertheless, for three specific items (item 7, item 8, and item 10), the percentage of misconceptions surpasses that of scientific knowledge. This observation underscores these concepts' considerable challenge as students grapple with confusion and incomplete understanding. The elevated prevalence of misconceptions in these instances highlights the complexity of the concepts and the ongoing struggles students face in fully grasping them. This result reinforces the notion that misconceptions present a formidable challenge for educators. As Lucariello (2013) stated, ordinary forms of instruction, such as lectures, labs, discovery learning, or simply reading texts, are sometimes unsuccessful at overcoming student misconceptions.

Table 5. Analysis of Students' Conceptual Knowledge

Item	Scientific knowledge	Lucky guess	Misconception	False-positive	False-negative	Lack of knowledge
1	54	2.3	31	4.6	3.8	4.6
2	43	4.9	29	3.3	5.9	14
3	53	3.1	26	4.6	0.5	13
4	31.2	4.09	36.57	5.37	7.16	15.6
5	48.1	4.6	27.6	6.65	3.07	9.97
6	46.8	3.84	26.34	3.84	5.12	14.07
7	30.18	2.56	51.15	6.39	0.26	9.46
8	24.81	4.09	53.96	1.02	3.32	12.79
9	57.03	5.63	27.11	1.79	1.79	6.65
10	26.3	2.81	51.7	3.07	5.63	10.5
Mean	41.44	3.79	36.04	4.06	3.66	11.06

According to the relevant literature on evaluating three-tier tests, content validity is affirmed by calculating the percentage of false negatives and false positives. Hestenes and Halloun (1995) recommended that the probability of false negatives should be less than 10%. Ensuring low probabilities of false negatives and positives contributes to higher validity in multiple-choice tests (Hestenes & Halloun, 1995). In this study, the obtained percentage of all items' false positives and false negatives are both below 7% and 8%, respectively, aligning with values associated with normal distribution.

Cataloglu (2002) proposed a correlation between high scores and confidence as evidence of construct validity. According to him, respondents with high scores were expected to be more confident than those with lower scores. A robust positive correlation indicates the proper functioning of test items. In this study, an investigation was conducted into the correlation between scores from both tiers and the certainty scores. A statistically significant positive correlation of 0.372 was identified at $p < 0.01$, further substantiating the construct validity of the OEDTK.

Furthermore, incorporating alternative blank space into the first two tiers aids in revealing other students' explanations regarding concepts. For instance, in the question illustrated in Figure 1 (question 10), a few students selected both options A and B for the first tier. In the second tier, the rationale behind students choosing both options is that "natural selection and the processes of labor and cognition are both influential factors in human evolution". Here, the students have mistaken the factors influencing the development of adaptive traits for the factors influencing the evolution of the human species. Despite the students' inaccuracies in their answers, these responses still possess value for teachers in identifying appropriate teaching strategies to refine students' comprehension in the future.

CONCLUSION

The present study focuses on developing an online three-tier test to diagnose students' misconceptions about evolution. Based on theoretical research, the procedure to build the OEDTK consisted of five main phases: (1) making a list of possible evolutionary misconceptions, (2) constructing items, (3) embedding the three-tier multiple-choice items into an online tool, (4) conducting a pilot study, (5) revising and completing the OEDTK.

The results indicate that the OEDTK has moderate difficulty, reliability, and value in investigating high school students' conceptual understanding as a data collection method for assessing evolution misconceptions. The OEDTK enables researchers to more accurately categorize students' perceptions, whether as a lack of knowledge, a lucky guess/lack of

confidence, or a misconception. The findings of this study are consistent with previous research, which demonstrated that three-tier diagnostic tests are appropriate instruments to identify misconceptions with high reliability.

With new improvements in constructing items with alternative blank answers in the first two tiers, the OEDTK combines the advantages of both a simple objective multiple-choice test and an open-ended test. This enables the online tool to be used by teachers to swiftly determine students' pre-existing knowledge, even within a large group of students. Simultaneously, students also have the opportunity to express their alternative evolutionary perceptions along with their corresponding reasons.

Based on the analysis results, it is evident that despite students having learned about evolution, many still hold misconceptions. Therefore, to help students overcome these alternative conceptions, teachers must find effective strategies to achieve conceptual change. The OEDTK provides teachers with information about students' prior knowledge and alternative conceptions, enabling the design of appropriate lessons to rectify these misconceptions and enhance students' understanding of evolution.

Moreover, Biology educational research in Vietnam still lacks online three-tier tests to address students' misconceptions. Thus, there is a need for more studies to develop online three-tier tests to identify misconceptions in other Biology topics to support teachers in teaching Biology and contribute to enhancing the quality of education in high schools.

Acknowledgment: Pham Thi Phuong Anh was funded by the Master, PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2022.TS004. The authors would like to thank teachers and students in Thua Thien Hue province, Vietnam, who voluntarily participated in this study.

REFERENCES

1. Allen, M. (2014). Misconceptions in Primary Science. McGraw-Hill Education (UK).
2. Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. International Journal of Science Education, 34(11), 1667-1686. doi: 10.1080/09500693.2012.680618.
3. Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. International Journal of Science Education, 32(7), 939–961. doi: 10.1080/09500690902890130.
4. Cataloglu, E. (2002). Development and validation of an achievement test in introductory quantum mechanics: The quantum mechanics visualization instrument (QMVI). The Pennsylvania State University.
5. Champagne, A. B., & Klopfer, L. E. (1983). Naive knowledge and science learning. Retrieved from https://files.eric.ed.gov/fulltext/ED225852.pdf.
6. Cunningham, D. L., & Wescott, D. J. (2009). Still more "fancy" and "myth" than "fact" in students' conceptions of evolution. Evolution: Education and Outreach, 2(3), 505-517.
7. Doran, R. L. (1972). Misconceptions of selected science concepts held by elementary school students. Journal of Research in Science Teaching, 9, 127-137 Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory. Physics Teacher, 3, 502-506.
8. Eryilmaz, A. (2010). Development and Application of Three-Tier Heat and Temperature Test: Sample of Bachelor and Graduate Students. Eurasian Journal of Educational Research (EJER), (40), 53-76. Retrieved from: https://ejer.com.tr/wp-content/uploads/2021/01/ejer_2010_issue_40.pdf#page=54.

9. Hasyim, W., Suwono, H., & Susilo, H. (2018). Three-tier test to identify students' misconception of human reproduction system. Jurnal Pendidikan Sains, 6(2), 48-54. Retrieved from: http://journal.um.ac.id/index.php/jps/article/view/11684.
10. Helm, H. (1980). Misconceptions in physics amongst South African students. Physics Education, 15, 92-97. doi: 10.1088/0031-9120/15/2/308.
11. Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory. Physics Teacher, 33, 502-506. Retrieved from: https://www.halloun.net/wp-content/uploads/2016/10/INTERFCI_TPT95.pdf.
12. Kanli, U. (2015). Using a Two-Tier Test to Analyse Students' and Teachers' Alternative Concepts in Astronomy. Science Education International, 26(2), 148-165. Retrieved from: https://files.eric.ed.gov/fulltext/EJ1064041.pdf.
13. Krasner, S. D., Nordlinger, E., Geertz, C., Skowronek, S., Tilly, C., Grew, R., & Trimberger, E. K. (1984). Approaches to the State: Alternative Conceptions and Historical Dynamics. Comparative Politics, 16(2), 223. doi: 10.2307/421608.
14. Lucariello, J., & Naff, D. (2013). How do I get my students over their alternative conceptions (misconceptions) for learning? Retrieved from: https://www.apa.org/education-career/k12/misconceptions.
15. Ministry of Education and Training (MOET) (2018). The general education program in Biology (Issued with Circular No. 32/2018/TT-BGDDT dated December 26, 2018, of the Minister of Education and Training). Hanoi.
16. Milenkovic, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Development of a three-tier test as a valid diagnostic tool for identification of misconceptions related to carbohydrates. Journal of Chemical Education, 93(9), 1514–1520.
17. Nehm, R. H., & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. BioScience, 57(3), 263-272.
18. Nguyen, T. D. (Chief Editor), Pham, V. L. (Editor), Dang, H. L., & Mai, S. T. (2012). Biology 12 [Sinh hoc 12]. Ha Noi, Vietnam Education Publishing House.
19. Morrison, J. A., & Lederman, N. G. (2003). Science teachers' diagnosis and understanding of students' preconceptions. Science Education, 87(6), 849-867. doi:10.1002/sce.10092.
20. Novak, J. D. (1966). The role of concepts in science teaching. In Analyses of Concept Learning (pp. 239-254). Academic Press. doi: 10.1016/B978-1-4832-3127-3.50020-1.
21. Pesman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. The Journal of Educational Research, 103(3), 208-222.
22. Putri, L. O. L., Rahman, T., & Priyandoko, D. (2017). Analyzing concepts mastery and misconceptions about evolution of biology major students. In Journal of Physics: conference series (Vol. 812, No. 1, p. 012083). IOP Publishing.
23. Ola Adeniyi, E. (1985). Misconceptions of selected ecological concepts held by some Nigerian students. Journal of Biological Education, 19(4), 311 316. doi:10.1080/00219266.1985.9654758.
24. Saat, R. M., Fadzil, H. M., Aziz, N. A. A., Haron, K., Rashid, K. A., & Shamsuar, N. R. (2016). Development of an online three-tier diagnostic test to assess pre-university students' understanding of cellular respiration. Journal of Baltic Science Education, 15(4), 532. Retrieved from: https://media.proquest.com/media/hms/PFT/1/07bdD?_s=Wp7zgoIx6qbPGxVin8kZ7CgcrSs%3D.

25. Sanger, M. J., & Greenbowe, T. J. (1997). Common student misconceptions in electrochemistry: Galvanic, electrolytic, and concentration cells. Journal of Research in Science Teaching, 34(4), 377–398. doi:10.1002/(sici)1098-2736(199704)34:4<377::aid-tea7>3.0.co;2-o.
26. Smith III, J. P., diSessa, A. A., & Roschelle, J. (1994). Misconceptions Reconceived: A Constructivist Analysis of Knowledge in Transition. Journal of the Learning Sciences, 3(2), 115–163. doi:10.1207/s15327809jls0302_1.
27. Soeharto, S., Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. Jurnal Pendidikan IPA Indonesia, 8(2), 247-266. doi: 10.15294/jpii.v8i2.18649.
28. Trinh, V. B., Vo., D. B, Tran., K. C. (2017). Using "hands-on" learning to detect misconceptions and build correct concepts for students in teaching physics in high school [Su dung phuong phap "ban tay nan bot" de phat hien quan niem sai lech va xay dung quan niem dung cho hoc sinh trong day hoc Vat li ơ trung hoc pho thong]. Journal of Education, June special issue, 179-183.
29. Uzuntiryaki, E., & Geban, Ö. (2005). Effect of conceptual change approach accompanied with concept mapping on understanding of solution concepts. Instructional Science, 33(4), 311–339. doi:10.1007/s11251-005-2812-z.
30. Vo, D. B. (2019). Detecting misconceptions and building correct concepts for students in teaching the part "Mechanics" (Physics 10) [Phat hien quan niem sai lech va xay dung quan niem dung cho hoc sinh trong day hoc phan "Co hoc" (Vat li 10)]. Journal of Education, April special issue, 207-209; 233.
31. Yates, T. B., & Marek, E. A. (2015). A study identifying biological evolution-related misconceptions held by prebiology high school students. Creative Education, 6(08), 811.

About the authors

– M.Ed. Pham Thi Phuong Anh is an educator and researcher in Biology teaching methods at The University of Education, Hue University. She is also a predoctoral fellow in Theory and Methods of Teaching in Biology in the Biology Department at the University of Education, Hue University. Her main research fields include constructivism as a theory for teaching and learning, critical thinking skills, and measures to develop and evaluate students' competencies. She has published fifteen articles in prestigious national and international scientific conferences/ journals, conducted seven research projects, and written three books on this research area.

– Assoc. Prof. Dr. Phan Duc Duy is an educator and researcher in Biology teaching methods at the University of Education, Hue University. His research fields are case exercises in teaching, developing biology teaching skills, and developing critical thinking skills for students. He has published fifty articles in prestigious national and international scientific conferences/ journals, conducted ten research projects, and written six books o this research area.

– Dr. Nguyen Thi Dieu Phuong is an educator and researcher in Biology teaching methods at the University of Education, Hue University. Her main research fields are teaching to develop learners' competencies, organizing experiential activities in teaching biology, testing and evaluating students' competencies, and integrated teaching. She has published over forty articles in prestigious national and international scientific conferences/ journals, participated in thirteen research projects, and written five books on this research area.