RESEARCH ARTICLE

Concept inventory test on cellular respiration for senior high school students: An assessment of reliability and validity

John Daryl C. Bace[a,1,*], **Shaina D. Guiral**[a,2], **Nichelle Karla C. Galicia**[a,3], **Mark Jay Reyes**[a,4], **Sylvester Cortes**[a,5]

[a] Cebu Technological University, M.J. Cuenco Ave., Cor R Palma St., Cebu, Philippines

[1]basejohndarylcasul@gmail.com*; [2]shainaguiral@gmail.com; [3]nikagaliciac@gmail.com; [4]cybermaster.mjr@gmail.com; [5]sylvester.cortes@ctu.edu.ph

Abstract: Given the complexity of cellular respiration, there is a significant need for effective assessment tools that can accurately gauge students' understanding, identify misconceptions, and provide educators with actionable insights to improve instruction. Concept Inventory Test (CIT) has emerged as a valuable instrument in science education for this purpose. The aim of this study is to develop a Concept Inventory Test on cellular respiration and assess its reliability and validity. This study employed a mixed-method sequential exploratory approach for each assessment. The study was conducted in a private school in Cebu, targeting the senior high school students. The results of the study showed that the developed Concept Inventory Test's reliability and validity are made apparent by the process used to gather supporting data, as well as by the findings and observations. Furthermore, this Concept Inventory Test serves a significant instrument for classroom assessment, promotes additional research on students' critical comprehension, and shows how these diagnostic inventories link to the students' competencies. However, a greater number of respondents should be taken to account to observe the consistency of the Concept Inventory Test results.

Keywords: assessment; cellular respiration; concept inventory test; science education; senior high school

For correspondence:
basejohndarylcasul@gmail.com

Article history:
Received: 28 September 2024
Revised: 4 November 2024
Accepted: 18 November 2024
Published: 30 November 2024

10.22219/jpbi.v10i3.36635

p-ISSN: 2442-3750
e-ISSN: 2537-6204

How to cite:
Bace, J.D.C., Guiral, S.D., Galicia, N.K.C., Reyes, M.J., & Cortes, S. (2024). Concept inventory test on cellular respiration for senior high school students: an assessment of reliability and validity. *JPBI (Jurnal Pendidikan Biologi Indonesia), 10*(3), 1128-1138. https://doi.org/10.22219/jpbi.v10i3.36635

Introduction

Mastering complex biological topics is crucial not only for academic success but also for fostering a deeper understanding of broader concepts in science education (Bybee, 2014). Cellular respiration has been one of the most complex topics in biology. It is essential for understanding how organisms convert biochemical energy from nutrients into adenosine triphosphate (ATP), which powers various cellular processes (Voet, & Pratt, 2006). Despite its foundational role in biological education, cellular respiration is often recognized as one of the most challenging topics for students to comprehend. The intricacies of glycolysis, the Krebs cycle, and the electron transport chain, along with the abstract nature of energy transformation and molecular interactions, contribute to widespread student misconceptions and learning difficulties (Dauer & Long, 2015; DiBattista & Kurzawa, 2011). For senior high school students, grasping this topic is crucial not only for academic success but also for fostering a deeper understanding of broader biological concepts (Bybee, 2014).

Given the complexity of cellular respiration, there is a significant need for effective assessment tools that can accurately gauge students' understanding, identify misconceptions, and provide educators with actionable insights to improve instruction. Concept inventories have emerged as valuable instruments in science education for this purpose. A concept inventory is a standardized assessment designed to evaluate students' understanding of core concepts within a specific domain by focusing on common misconceptions and alternative conceptions that students may hold (Adams & Wieman, 2010). These tools have been widely used across various scientific disciplines, including physics, chemistry, and biology, to assess student learning, guide curriculum development, and evaluate the effectiveness of instructional interventions (Smith & Tanner, 2017).

However, the development of a concept inventory is not a trivial task. To be effective, a concept inventory must demonstrate strong psychometric properties, particularly in terms of reliability and validity.

Reliability refers to the consistency of the assessment results—whether the inventory consistently measures what it intends to measure across different populations and over time (McNeish, 2017). Validity, on the other hand, concerns the accuracy and appropriateness of the inferences drawn from the assessment results. Content validity refers to the extent to which the test items represent the entire domain of the concept being measured (Messick, 1989), —in this case, cellular respiration. Typically, if a CIT passed the reliability and validity assessment, this would build confidence that the test measures the intended concepts consistently, and that the CIT is meaningful reflected through the students' understanding of the subject matter.

Research on CITs often overlooks the specific cognitive and pedagogical challenges that senior high school students face when learning complex topics like cellular respiration (Liu, Lin, & Tsai, 2008). Exploring how these factors influence the reliability and validity of concept inventory tests at this educational level could provide valuable insights for improving test design and interpretation. Also, while concept inventories exist for various biological topics, only few are specifically tailored to assess understanding of cellular respiration at the senior high school level (Freeman *et. al*., 2014). Most existing inventories either focus on broader biological concepts or are designed for higher education students, leaving a gap in the availability of tools suited for senior high school students.

Particularly, this study aims to develop and validate a Concept Inventory Test on Cellular Respiration specifically designed for senior high school students. The primary objectives of the research are to rigorously assess the reliability of the inventory and to establish its validity within the context of senior high school biology education. The inventory will be designed to target common misconceptions about cellular respiration, providing a tool that not only evaluates student understanding but also informs instructional practices. By providing a robust assessment tool, this study seeks to contribute to the enhancement of biology education at the senior high school level, helping students to overcome the challenges associated with learning about cellular respiration and, more broadly, to achieve a deeper understanding of biological processes.

Method

Research Design

This study utilized a mixed-method sequential exploratory approach that focused on the development, implementation, and assessment on the validity and reliability of a Concept Inventory Test (CIT) in Cellular Respiration. For the qualitative part of the study, content validity was evaluated which involved expert reviews and comments with the subject experts who provide feedback on whether the test items accurately reflect the key concepts of cellular respiration. On the other hand, quantitative methods involve the analysis for reliability and validity assessment. These methods provide objective, numerical data on the consistency of the test. The use of this research design allows a thorough and well-rounded assessment of the CIT. By integrating these methods, the research can address both the statistical robustness and the practical and conceptual appropriateness of the test.

Research Environment and Participants

The participants in this study are composed of senior high school students from a private institution known for its academic excellence internationally. These students were in grade 11 and 12 that differed in their strand, which were HUMSS, ICT, and STEM. The test was administered by their respective advisers to ensure accurate responses. Additionally, there were three experts who were chosen to ensure the validity of the developed concept inventory test. They are composed of experts who are currently teaching biological education in senior high school and college (Table 1).

Table 1. Profile of the respondents.

Male	Female	Total
119	81	200

Research Instruments

This study aimed to create and evaluate the reliability and validity of a concept inventory test on cellular respiration. There were three research instruments employed in conducting the data collection approach; various literature reviewers and science amplitude tests were used for the formulation of the test questions based on the DepEd's General Biology 1 and 2 Most Essential Learning Competencies (MELC); Table of Specifications (TOS) was used to show items under bloom's taxonomy cognitive approach; and the checklist by Morales, 2003 was also used for the content validation assessed by the validators.

Data Collection Procedure

The development and the overall collection of data was created in four primary processes according to the previous works of (Aligway *et al,* 2024; Moradas *et al*, 2024): (i) preparing the CIT design; (ii) developing the CIT; (iii) validating the CIT; and (iv) administering of CIT to the selected participants.

Step 1: Preparation of CIT

The formulation of the CIT began with a careful selection and compilation of key topics in the field of cellular respiration. The test questions were solely from various literature reviewers of the book, "Exploring life through science" second edition (Ramos & Ramos, 2022). and science amplitude tests that included DepEd's General Biology 1 and 2 Most Essential Learning Competencies as the foundation for the CIT.

Step 2: Development of CIT

Based on the topics from phase 1, a Table of Specifications (TOS) was developed using Bloom's revised taxonomy to guide the development of the CIT. The test questionnaire contains 20 question items with 60% Lower-Order Thinking Skills (LOTS), which highlights the ability to remember and understand key concepts, and 40% Higher-Order Thinking Skills (HOTS), which focuses on the ability to apply, analyze, evaluate, and create concepts according to their understanding. The percentages of each order thinking skills were marginally based on the studies of (Winanda & Anwar, 2022). The choice of students or respondents should have experience learning or taking the Cellular Respiration topic in general biology as a subject course, hence, the choice of senior high school students. Revision of test questions were also made after validation of the experts (Table 2).

Table 2. Table of Specifications (TOS) with their corresponding items.

TOPICS / LEARNING COMPETENCIES	OBJECTIVES						TOTAL
	LOTS		HOTS				
	Remember	Understand	Apply	Analyze	Evaluate	Create	
Cellular Respiration							
*Explain the major features and sequence the chemical events of cellular respiration	Q1, Q6	Q11					3
*Distinguish major features of glycolysis, Krebs cycle, electron transport system	Q2, Q3, Q5	Q9					4
*Describe reactions that produce and consume ATP	Q4	Q8, Q12		Q15	Q17		5
*Describe the role of oxygen in respiration and describe the pathways of electron flow in the absence of oxygen		Q7	Q14	Q16		Q19	4
*Explain the advantages and disadvantages of fermentation and aerobic respiration		Q10	Q13		Q18	Q20	4
Total Number of Items:	**6**	**6**	**2**	**2**	**2**	**2**	**20**
Percentage:	**30%**	**30%**	**10%**	**10%**	**10%**	**10%**	**100%**

Step 3: Validation of the CIT

Third phase of the methodology involves validation of CIT to determine the quality of the test. Test items are rigorously evaluated and refined as part of the validation process for concept inventory tests. Three experts with an extensive experience and expertise in analyzing test questions in Cellular Respiration were to guarantee content validity, confirming that the items effectively cover the desired structures and misconceptions. All of them are experts on the field of biological education which still currently practices teaching for senior high school and college students. The 20-item multiple-choice conceptual test (v1) was revised based on the comments of the evaluators. The evaluators had the option to accept, modify, or reject the items. Statistical studies, including item difficulty and discrimination indices, are used to measure the performance of each question (Madsen, McKagan, & Sayre, 2017). The second version of the conceptual test (v2) was subject for content validation, which uses 20 item-Content Validation Checklist, and Aiken's validity test.

Step 4: Administration of CIT

The fourth step is administering the CIT. The CIT was then performed and completed by two hundred (200) selected participants that had previously studied the topic matter, Cellular Respiration under General Science subject course. Respondents were represented from a private school in Cebu. There were 119 males and 81 females, ranging in age from 17 to 18 years old. The selection process was carefully executed to ensure that the participants are competent and suitable to perform the concept inventory tests. The researchers have asked permission to the school to administer the exams through face-to-face setup. After an hour, the tests were collected and were subject for data analysis.

Data Analysis

After the previous methods, all raw data collected then underwent data analysis. The content validation was conducted using the aid of Table of Specifications, expert judgments, and Content Validity Index scores calculation. After the administration of the CIT, the collected data were subject for assessment of reliability, evaluating item difficulty, item discrimination, discrimination level, item classification, and Cronbach's alpha. Using appropriate statistical tests, the data were cross-sectioned to carry out any significant variations in scores among the students.

Results and Discussion

The objective of the study is to develop, implement, and assess the validity and reliability of a multiple-choice concept inventory test on Cellular Respiration. Twenty (20) items in total comprised the topics under cellular respiration in this study. There were four choices for each question; one is a correct response, while the other three serves as distractors.

Table 3 illustrates the percentage distribution of items based on Bloom's Taxonomy of Cognitive Domain. About 40% of the test questions, consisting of eight (8) items, were created using Higher-Order Thinking Skills. Furthermore, about 60% of the Lower-Order Thinking Skills test consists of twelve (12) items. The percentage distribution of each item is displayed in the third column. It shows that remembering and understanding items have equally distributed 30%. The four categories in HOTS items - applying, analyzing, evaluating, and creating have equally share 10% (Table 3).

Table 3. Percentage Distribution of Items Based on Bloom's Taxonomy of Cognitive Domain (v.1)

Bloom's Taxonomy of Cognitive Domain	No. of Items	Percentage out of 100	Percentage of LOTS and HOTS
		LOTS	
Remember	6	30%	
Understand	6	30%	60%
		HOTS	
Apply	2	10%	
Analyze	2	10%	
Evaluate	2	10%	
Create	2	10%	40%
Total	***20***	***100***	

Content Validation

The CIT was assessed and verified by the three (3) biology experts who are professional educators. This procedure provides a test of good quality. The test will be evaluated in both qualitative and quantitative terms by the three professional evaluators. Every item on the checklist is rated by the evaluators, who further provide feedback on each one. Table 4 presents a comparison of the overall mean of the content validity of version 1 (v.1) and version 2 (v.2). In version 1, it has the overall mean of 3.75 out of 5. This shows that the content validity is between Undecided to Agree (Appendix A). Furthermore, it also displays the constructive comments of the three evaluators who recommended revisions to certain items. Columns 2, 3, and 4 of version 1 (v.1) state the individual remarks from the evaluators for test development.

The revised test (v.2) which consists of the same 20 number of items was restructured according to the comments for improvement. The same three experts rated it with an overall mean of 4.72 out of 5 as shown in Table 3. The overall mean demonstrates that the content validity of version 2 is between Agree to Strongly Agree (Appendix A). This indicates that the remarks have been applied into context which denotes that there is an improvement from the first validation. This demonstrates a good quality in construction and validity of the test items.

Table 4. Content Validity of Version 1 (v.1) and Version 2 (v.2)

	v.1			v.2		
	Expert 1	Expert 2	Expert 3	Expert 1	Expert 2	Expert 3
Mean	3.7	3.85	3.75	4.65	4.9	4.6
Comments	1. " There are certain items that require revision. Some of the items should be formulated under the create category."	1. "There are items subject for revision. The questions should be rephrased to provide a clear understanding of the subject matter."	1. "Improve the structure of the questions."	1. " The items have been organized into categories for recalling, understanding, applying, analyzing, evaluating, and creating."	1. " The questions were presented clearly."	No comments
Over-all Mean			3.75 out of 5		4.72 out of 5	

Moreover, the content validity coefficient has been determined for each checklist item to further assure that each item qualifies as a content-valid test. This was done based on Aiken's content validity coefficient. The average content validity coefficient in version 1 is 0.69. However, in version 2 displays an average content validity coefficient of 0.93, which is closer to 1 as shown in Table 5. In version 2, every item on the checklist earned a rating that was closer to 1, indicating a high level of content validity.

Table 5. Content Validity Coefficient

Checklist Items	Aiken's V (Content Validity Coefficient)	
	v.1	v.2
1	0.58	0.92
2	0.50	0.92
3	0.67	0.92
4	0.50	0.92
5	0.50	0.92
6	0.75	1
7	0.75	0.92
8	0.75	0.92
9	0.75	0.92
10	0.67	0.92
11	0.75	0.92
12	0.75	1
13	0.58	0.92
14	0.67	0.92
15	0.75	0.92
16	0.75	0.92
17	0.67	0.92
18	0.75	1
19	0.75	0.92
20	0.92	0.92
Average	**0.69**	**0.93**

Item Analysis for Validation

Table 6 shows an effective item analysis that involves grouping items according to their difficulty and discrimination to ensure that tests accurately reflect student learning. According to (Odukoya, Adekeye, Igbinoba, & Afolabi, 2017), item analysis is essential for ensuring that assessments appropriately represent students' comprehension and learning outcomes. Furthermore, item analysis assists educators in making informed judgments about which items to retain or adjust, thereby improving the reliability and validity of tests used in educational contexts.

Table 6. Content Validity Coefficient

Item No.	Item Difficulty Index	Item Difficulty Remarks	Item Discrimination Index	Item Discrimination Remarks	Decision
1	0.96	Very Easy	0.15	Moderately Discriminating	Reject
2	0.8	Very Easy	0.50	Discriminating	Reject
3	0.71	Easy	0.63	Very Discriminating	Needs Revision
4	0.61	Easy	0.39	Discriminating	Needs Revision
5	0.73	Easy	0.46	Discriminating	Needs Revision
6	0.67	Easy	0.59	Discriminating	Needs Revision
7	0.65	Easy	0.70	Very Discriminating	Needs Revision
8	0.45	Moderately Difficult	0.80	Very Discriminating	Accept
9	0.77	Easy	0.67	Very Discriminating	Needs Revision
10	0.65	Easy	0.57	Discriminating	Needs Revision
11	0.67	Easy	0.69	Very Discriminating	Needs Revision
12	0.66	Easy	0.46	Discriminating	Needs Revision
13	0.35	Difficult	0.33	Discriminating	Accept
14	0.68	Easy	0.57	Discriminating	Needs Revision
15	0.58	Moderately Difficult	0.61	Very Discriminating	Accept
16	0.79	Easy	0.46	Discriminating	Needs Revision
17	0.77	Easy	0.59	Discriminating	Needs Revision
18	0.62	Easy	0.65	Very Discriminating	Needs Revision
19	0.57	Moderately Difficult	0.54	Discriminating	Accept
20	0.35	Difficult	0.39	Discriminating	Accept

Item Difficulty

The test items' level of difficulty was based on Ebel (1972), as shown in Table 6. Each item's difficulty index was determined, and it was divided into categories such as "very easy", "easy", "moderately difficult", "difficult", and "very difficult". The results in item difficulty shows a considerable tilt towards easier items, as seen in Table 7, with 65% of the test items classified as Easy. The presence of 10% Very Easy (VE) items adds to this pattern, demonstrating that a substantial portion of the test is accessible to a diverse range of individuals. In contrast, just 15% of the questions were categorized as Moderately Difficult (MD), and only 5% as Difficult (D). Notably, there were no Very Difficult (VD) items, raising concerns about the assessment's comprehensiveness in addressing the highest levels of participant ability. The average difficulty index of 0.61 is consistent with the observed distribution, indicating that the test is primarily easy.

Table 7. Item Difficulty of Version 2 (v.2)

Item Difficulty Interpretation	Item Difficult Range	Number of Items	Percentage
Very Easy (VE)	0-81 and above	2	10%
Easy (E)	0.61- 0.81	13	65%
Moderately Difficult (MD)	0.41-0.06	3	155
Difficult (D)	0.21-0.4	2	10%
Very Difficult (VD)	0.00 - -.2	0	0%
Total		**20**	**100%**
Test Difficulty		**0.61**	**61.76**

Item Discrimination

Using the higher and lower groups of test-takers is one efficient way to obtain item discrimination based on student responses. The researchers determine which group, the upper (27% of scores) and lower (27% of scores) are the highest and lowest, respectively, out of all the responses. Then, the percentage of right answers were compared for each evaluation item to evaluate how well these two groups performed. The summary of the percentage of all item discrimination indexes were shown in Table 4.

The item discrimination analysis in Table 8 demonstrates a favorable distribution of item effectiveness, with a majority of items (60%) being classified as discriminating (D) and a significant proportion (35%) as very discriminating (VD) and identified 5% as moderately discriminating. In addition to this, the item discrimination analysis revealed no questionable or non-discriminating items.

Table 8. Item Discrimination of Version 2 (v.2)

Item Discrimination Interpretation	Range	Number of Items	Percentage
Questionable (Q)	(-) 1.00 – (-) 0.60	0	0%
Not Discriminating (ND)	(-) 0.59 – (-) 0.09	0	0%
Moderately Discriminating (MD)	0.10 – 0.20	1	5%
Discriminating (D)	0.10 – 0.60	12	60%
Very Discriminating (VD)	0.60 – 1.00	7	35%
Total		**20**	**100%**
Discriminating Level		**0.54**	

Additionally, using the discrimination index and difficulty index determine whether Items are classified as accepted, modified, or rejected. Table 9 shows the items classified based on the item analysis of v2. The 20-item multiple choice test's item analysis results showed that 50% are acceptable, 40% require revision, and 10% ought to be rejected.

Table 9. Item Classified Based on Item Analysis of Version 2 (v.2)

	Number of Items	Percentage
Accept	10	50%
Revise	8	40%
Reject	2	10%
Total	**20**	**100%**

Reliability

A Cronbach's alpha of .377 indicates a low level of internal consistency among the items being measured, suggesting that the items do not correlate well with one another (Willems *et al.*, 2023). This low alpha indicates that the items may not be reliable for assessing the intended construct, and researchers may need to revise or replace items to enhance internal consistency (Taber, 2017). Furthermore, while a low alpha implies reliability issues, it does not address the validity of the instrument, implying that consistent responses do not guarantee that the items adequately measure the intended concept (Table 10).

Table 10. Reliability Statistics Results

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.377	.365	18

Distractor Analysis

Table 11 shown below is the sample used for the Distractor analysis. This table provides a clear sample of the data collected for this study. In analyzing the performance of students on item, a majority of students in the upper group selected letter D as their answer. Conversely, a significant portion of students in the lower group opted for letter C, which was identified as a distractor. According to (DiBattista & Kurzawa, 2011), well-constructed distractors can enhance the diagnostic power of assessments by revealing the cognitive processes involved in answering questions incorrectly, particularly for lower-performing students.

In item Figure 1, both the upper and lower groups predominantly selected answer B as the correct response. However, a significant number of students in the lower group also chose answers C and D, indicating potential confusion regarding the distractors. This pattern suggests that while the item was correctly identified by the majority, the presence of misleading options may have contributed to the incorrect selections among lower-performing students. Research indicates that well-constructed items should have a larger proportion of right responses from higher-performing students, although lower-performing students typically chose distractors (Bhat & Prasad, 2020).

Table 11. Distractor Analysis Sample.

Item No.	Upper (n=54)				Lower (n=54)				Remarks
	A	B	C	D	A	B	C	D	
1	*54	0	0	0	*46	1	3	**4	**D** poses a misconception
2	2	*52	0	0	**12	*25	5	**12	**A** poses a misconception
3	0	0	*53	1	**21	5	*19	9	**A** poses a misconception
4	2	1	7	*44	8	9	**14	*23	**C** poses a misconception
5	0	*52	0	2	**15	*27	6	6	**A** poses a misconception
6	0	1	*51	2	8	2	*19	**25	**D** poses a misconception
7	*51	0	0	3	*13	7	17	**17	**D** poses a misconception
8	1	*50	3	0	14	*7	12	**21	**D** poses a misconception
9	*54	0	0	0	*18	7	**17	12	**C** poses a misconception
10	3	*51	0	0	11	*20	**15	8	**C** poses a misconception
11	1	*53	0	0	7	*16	6	**25	**D** poses a misconception
12	0	*46	8	0	11	*21	**11	**11	**C** poses a misconception
13	15	4	*28	7	10	**23	*10	11	**B** poses a misconception
14	2	*52	0	0	**12	*21	8	**13	**A** poses a misconception
15	0	4	*49	1	**14	8	*16	**16	**D** poses a misconception
16	0	0	*54	0	**13	1	*29	11	**A** poses a misconception
17	*53	0	0	1	*21	5	**19	9	**C** poses a misconception
18	0	2	*52	0	**19	10	*17	**18	**A** and **D** poses a misconception
19	1	*47	0	6	7	*18	**14	**15	**D** poses a misconception
20	0	5	**19	*30	10	13	**22	*9	**C** poses a misconception

* correct answer
**portable source of misconception

19. Supposed you design a simple experiment to measure the rate of cellular respiration in yeast cells. Which of the following could be used as an indicator of respiration rate?
I. Amount of glucose present
II. Amount of CO2 produced
III. Amount of oxygen consumed
a. I & III
b. II & III
c. I & III
d. d. I, II, III

Figure 1. Sample Item

In analyzing the results of item Figure 2 from the distractor analysis, we observe significant discrepancies between the upper and lower groups of students, which reveal underlying misconceptions related to the distractors presented (Shin, Guo, & Gierl, 2019). The upper group, consisting of 52 students, predominantly selected the correct answer (C), while only 17 students from the lower group did so. Notably, 19 students from the lower group chose distractor A, and 18 opted for distractor D. This pattern suggests that distractors A and D may embody common misconceptions that mislead students who lack a solid understanding of the material (Wind *et al*, 2019). The fact that both distractors A and D garnered significant attention from the lower group implies that they were not only plausible but potentially misleading, leading students to incorrect conclusions (Rezigalla *et al*., 2024). One reason students may choose A is that they confuse the effects of increased respiration with those of impaired respiration. Impaired respiration can lead to decreased ATP production, but increased respiration should enhance ATP synthesis (Raimondi, Ciccarese, & Ciminale, 2020). The selection of D could stem from the mistaken belief that increased respiration is linked to lactic acid accumulation, when in fact, lactic acid builds up under anaerobic conditions (Zhao, Jiang, Zhang, & Yu, 2019).

18. Given that the electron transport chain produces reactive oxygen species (ROS), what might be a potential downside of increased cellular respiration?
a. Decreased ATP production.
b. Increased glucose consumption.
c. Damage to cellular structures due to ROS.
d. Accumulation of lactic acid.

Figure 2. Sample Item

The distractor analysis for item Figure 3, reveals that option C may have been constructed based on common misunderstandings related to the content, which aligns with findings that effective distractors often reflect frequent student misconceptions, thereby highlighting the need for careful item design to address these errors (Testa, Toscano, & Rosato, 2018). On the contrary, item #16 shown, distractors like option A can lead to misconceptions, as students may select it based on familiarity or incorrect reasoning, which can detract from their overall performance and understanding of the material (Caldwell & Pate, 2013).

17. Why is it more efficient for human cells to perform aerobic respiration rather than anaerobic respiration?

a. It produces more ATP per glucose molecule.
b. It produces lactic acid.
c. It does not produce carbon dioxide.
d. It is faster.

Figure 3. Sample Item

16. An organism is found to produce ATP through both aerobic and anaerobic pathways. What can you infer about this organism's environment?

a. It is always oxygen-rich.
b. It is always oxygen-poor.
c. It can experience varying oxygen levels.
a. d. It does not need oxygen.

Figure 4. Sample Item

Interestingly, for item Figure 4, the presence of a significant number of students in the lower group opting for option D highlights the need for careful item construction, as distractors that resonate with student misconceptions can reduce the discriminative power of the assessment, making it less effective in distinguishing between varying levels of student understanding (Testa, Toscano, & Rosato, 2018). The choice of letter D as a distractor in this scenario likely stems from a misunderstanding of the metabolic pathways involved in cellular respiration, particularly among lower-performing students who may confuse fermentation with other processes. Students may incorrectly associate the lack of ATP production and increased NADH with fermentation, which typically occurs when oxygen is absent and is characterized by a different metabolic response than what is observed in the presence of oxygen during cellular respiration (Nadanaciva *et al*., 2012). Therefore, misconceptions enhance the effectiveness of distractors by reflecting common mistakes that students make, raising the plausibility of incorrect choices and challenging their comprehension of the content.

Conclusion

The content validity and reliability test provide quality and effectiveness of the concept inventory test that would help to assess the students' understanding in the concept of cellular respiration. The test is designed to measure students' misconceptions, which can assist educators to assess their knowledge on the subject matter. The findings of the distractor analysis provided the educators with input on which specific concepts to revisit or reassess with the students in addition to highlighting the prevalent misconceptions represented by the specified number of students. The developed concept inventory test's validity and reliability are made apparent by the process used to gather supporting data, as well as by the findings and observations. Furthermore, the concept inventory test serves a significant instrument for classroom assessment, promotes additional research on students' critical comprehension, and shows how these diagnostic inventories link to the students' competencies. However, a greater number of respondents should be taken to account to observe the consistency of the concept inventory test results.

Recommendation

The researchers recommend future researchers to replicate the study in diverse educational settings, including other regions, school types, and cultural contexts. This would help determine the generalizability of the CIT and identify any necessary adaptations for different student populations. Also, the researchers recommend conducting longitudinal studies to assess the stability of the CIT over time. This would involve administering the test to the same group of students at different points in their academic journey to see how well the test maintains its reliability and validity across different stages of learning.

Acknowledgment

The authors would like to extend their warmest gratitude to the following people relevant to the conduction of the study; DOST STRAND-N for continuous financial support and guidance; students who willingly participated through the examination process; the advisers who did not hesitate to help disseminate the CIT to the students and help collect the data; the experts who tirelessly shared their useful comments of the developed CIT; to the relevant people of the authors that supported and inspired them to conduct the study; Sir Sly for tirelessly giving comments and useful ideas that the authors can cherish forever; and to the God almighty who guided the authors along the way. All these aforementioned have contributed equally to the success of the study. Truly, without the other, this would not be possible and successful.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author Contributions

J.D.C.B.: Introduction and methods; supervised the crafting of concept inventory test and contributed to the written manuscript; performed data analysis and manuscript revisions; supervised the content of the manuscript, **S.D.G.:** Performed data collection and data analysis; organized the collected data; contributed questions in the concept inventory test; performed the manuscript revisions; and contributed to research components, **N.K.C.G.:** Contributed in the methods, contributed questions to the concept inventory test; performed data analysis, created tables and figures; performed manuscript revisions; and contributed to research components, **M.J.R.:** Contributed to the concept inventory test; calculated the Cronbach's alpha, **S.T.C.:** Revised and finalized the manuscript.

References

Adams, W. K., & Wieman, C. E. (2010). Development and validation of instruments to measure learning of expert-like thinking. *International Journal of Science Education*, *33*(9), 1289-1312. https://doi.org/10.1080/09500693.2010.512369

Aligway, G. B., Delos Angeles, J. C., Collano, A. V., Barroca, E. P., Aves, A. D., Catubay, J. F., Edjec, J.T., Butaya, M.D., & Cortes, S. T. (2024). Validity and reliability of a concept inventory test in human physiology. *Jurnal Pendidikan Biologi Indonesia, 10*(1), 265-274. https://doi.org/10.22219/jpbi.v10i1.29558

Bhat, S. K & Prasad, K. H. L. (2020). Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology. *Indian Journal of Opthalmology*, *69*(2), 343-346. http://doi.org/ 10.4103/ijo.IJO_1610_20

Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of Science Teacher Education, 25*(2), 211-221. https://doi.org/10.1007/s10972-014-9381-4

Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *American Journal of Pharmaceutical Education, 77*(4), 71. http://doi.org/10.5688/ajpe77471

Dauer, J. T., & Long, T. M. (2015). Long-term conceptual retrieval by college biology majors followinhg model-based instruction. *Journal of Research in Science Teaching, 52*(8), 118-1206. https://doi.org/10.1002/tea.21258

DiBattista, D., & Kurzawa, L. (2011). Examination of the Quality of Multiple-choice Items on Classroom Tests. *The Canadian Journal for the Scholarship of Teaching and Learning, 2*(2), 25. http://dx.doi.org/10.5206/cjsotl-rcacea.2011.2.4

Ebel, R. L., & Frisbie, D. A. (1972). Essentials of Educational Measurement (5th Edition ed.). *Prentice-Hall International, Inc.*

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wnderoth, M. P. (2014). Active learning increases student perfromance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111

Liu, T.C., Lin, Y.C., & Tsai, C.C. (2009). Identifying senior high school students' misconceptions about statistical correlation, and their possible causes: an exploratory study using concept mapping with interviews. *International Journal of Science and Mathematics Education, 7,* 791–820. https://doi.org/10.1007/s10763-008-9142-y

Madsen, A., McKagan, S. B., & Sayre, E. C. (2017). Best Practices for administering concept inventories. *The Physics Teacher, 55*(9), 530-536. https://doi.org/10.1119/1.5011826

McNeish, D. (2017). Thanks Coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412-433. https://doi.org/10.1037/met0000144

Messick, S. (1989). Validity. New York: *American Councsil on Education and Macmilan Publishing Company.*

Moradas, J. D., Socubas, F. B., Bacasmas, V. M. R., Arquilita, S. G. L., & Cortes, S. T.(2024).Development and validation of a concept inventory test inphotosynthesis for junior high schoolstudents. JPBI (Jurnal Pendidikan Biologi Indonesia), *10*(3),898-908.https://doi.org/10.22219/jpbi.v10i3.36634

Nadanaciva, S., Rana, P., Beeson, G. C., Chen, D., Ferrick, D. A., Beeson, C. C., & Will, Y. (2012). Assessment of drug-induced mitochondrial dysfunction via altered cellular respiration and acidification measured in a 96-well platform. *Journal of Bioenergetics and Biomembranes, 44*(4), 421-437. https://doi.org/10.1007/s10863-012-9446-z

Odukoya, J. A., Adekeye, O., Igbinoba, A. O., & Afolabi, A. (2017). Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. *Quality & Quantity, 52*, 983–997. https://doi.org/10.1007/s11135-017-0499-2

Raimondi, V., Ciccarese, F., & Ciminale, V. (2020). Oncogenic pathways and the electron transport chain: a dangeROS liaison. *British Journal of Cancer, 122*, 168-181. https://doi.org/10.1038/s41416-019-0651-y

Ramos, A. C., & Ramos. (2022). Exploring Life Through Science Series. Phoenix Publishing House.

Rezigalla, A. A., Eleragi, A. M., Elhussein, A. B., Alfaifi, J., ALGhamdi, M. A., Al Ameer, A. Y., Yahia, A. I. O., Mohammed, O.A., & Adam, M. I. E. (2024). Item analysis: the impact of distractor efficiency on the difficulty index and discrimination power of multiple-choice items. *BMC Medical Education, 24*. https://doi.org/10.1186/s12909-024-05433-y

Shin, J., Guo, Q., & Gierl, M. J. (2019). Multiple-choice item distractor development using Topic Modeling Approaches. *Frontiers Psychology, 10*, 825. http://doi.org/10.3389/fpsyg.2019.00825

Smith, J. I., & Tanner, K. (2017). The problem of revealing how students think: Concept inventories and beyond. *CBE- Life Sciences Education, 9*(1), 1-5. https://doi.org/10.1187/cbe.09-12-0094

Taber, K. S. (2017). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*, 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Testa, S., Toscano, A., & Rosato, R. (2018). Distractor efficiency in an item pool for a statistics classroom exam: Assessing its relation with item cognitive level classified according to bloom's taxonomy. *Frontiers in Psychology*, *9*. http://doi.org/ 10.3389/fpsyg.2018.01585

Voet, D., Voet, J. G., & Pratt, C. W. (2006). *Fundamentals of biochemistry: Life at the molecular level* (5th ed.). John Wiley & Sons.

Willems, I., Vera, V., Calders, P., Lapauw, B., & Craemer, M. D. (2023). Test–retest reliability and internal consistency of a newly developed questionnaire to assess explanatory variables of 24-h movement behaviors in adults. *International Journal of Environmental Research and Public Health*, *20*(5). https://doi.org/10.3390/ ijerph20054407

Winanda, G. Z., & Anwar, D. (2022). Analysis of the implementation of higher-order thinking skills in tasks and test of English subject at SMK Negeri 2 Padang. In *Proceedings of the 67th TEFLIN International Virtual Conference & the 9th ICOELT 2021* (pp. 106–109). Atlantis Press. https://doi.org/10.2991/assehr.k.220201.019.

Wind, S. A., Alemdar, M., Lingle, J. A., Moore, R., & Asilkalkan, A. (2019). Exploring student understanding of the engineering design process using distractor analysis. *International Journal of STEM Education*, *6*(1), 4. https://doi.org/10.1186/s40594-019-0163-6

Zhao, R. Z., Jiang, S., Zhang, L., & Yu, Z. B. (2019). Mitochondrial electron transport chain, ROS generation and uncoupling (Review). *International Journal of Molecular Medicine*, *44*(1), 3-15. https://doi.org/10.3892/ijmm.2019.4188