ORIGINAL ARTICLE

Development of a Protein Concept Inventory: A Proposal for Item Scoring and Responding

Güntay Taşçi

Erzincan Binali Yıldırım University, Erzincan, Turkey

***Abstract**: The present study has aimed to develop and validate a protein concept inventory (PCI) consisting of 25 multiple-choice (MC) questions to assess students' understanding of protein, which is a fundamental concept across different biology disciplines. The development process of the PCI involved a literature review to identify protein-related content, validation interviews to iteratively validate and refine the created items (n = 26), and data collection from a large sample (n = 291) for statistical analysis. An expert interview was held with two different field experts regarding the content validity of the draft PCI tool, the suitability of the options, and the clarity of the items. Free choice format (multiple marking) was used to answer the developed MC items. In scoring these items, positive points were given to correct options, and negative points were given to incorrect options. Evidence regarding the psychometric properties of the PCI trial form was collected through factor analysis, group differentiation, internal consistency, and item analysis using quantitative data. The evidence collected demonstrates that the validity and reliability of the PCI as a measurement tool have been confirmed. PCI's scoring approach and the use of response patterns created by multiple markings in teaching are discussed.*

Science Insights Education Frontiers 2024; 23(2):3755-3777
DOI: 10.15354/sief.24.or617

How to Cite: Taşçi, G. (2024). Development of a protein concept inventory: A proposal for item scoring and responding. Science Insights Education Frontiers, 23(2):3755-3777.

***Keywords**: Protein, Conceptual Understanding, Free Choice Format*

***About the Author**: Güntay Taşçi, Educational Faculty, Erzincan Binali Yıldırım University, Erzincan, Turkey*. *E-mail: gtasci@erzincan.edu.tr*

***Correspondence to**: Güntay Taşçi at Erzincan Binali Yıldırım University of Turkey.*

***Conflict of Interests**: None*

***AI Declaration**: The author affirms that artificial intelligence did not contribute to the process of preparing the work.*

Introduction

CONCEPT inventories (CI) are widely used as a robust approach for evaluating conceptual understanding through the identification of misconceptions (D'Avanzo, 2008; Kalas et al., 2013; Klymkowsky & Garvin-Doxas, 2008; Libarkin, 2008). Starting with the Force Concept Inventory (Hestenes et al., 1992), CIs have been developed for various concepts (Newman et al., 2016). In relation to this, more studies have focused on developing CIs for teaching biology (Bybee, 2012; D'Avanzo, 2008; Garvin-Doxas et al., 2007). While these CIs may include a wide range of topics, including general biology, genetics, and microbiology, some concepts, such as homeostasis (McFarland et al., 2017) and natural selection (Anderson et al., 2002), have been selected. The widespread use of the CI approach has resulted in an essential body of literature on developing these tools. Researchers investigating the methods used for CI development have reported that these development processes include many variations (Lindell et al., 2007; Jorion et al., 2015). In this context, Adams and Wieman (2011) proposed a robust development framework with four fundamental phases for CI development. This framework included recommendations to identify the content and construction of the core concept, developing test specifications, field testing, validation interviews, scoring rules, and answering procedures.

Even though MC items are widely used in CIs due to their advantages, CIs are different from ordinary MC tests (Kalas et al., 2013; Smith & Knight, 2012). Stressed characteristics of CIs are measuring conceptual understanding, diagnosing understanding levels in learners (D'Avanzo, 2008; Garvin-Doxas et al., 2007; Libarkin, 2008), and containing research-based misconceptions in the options (distracters). This approach has been named distracter-driven MC items (Herrmann-Abell & DeBoer, 2011; Shin et al., 2019). Additionally, answering and scoring are other features of MC items discussed in the literature. Generally, the examinee can mark one of the options in an MC item, including the keyed item and other distracters. Its scoring is mainly determined by giving one (1) point for the correct answer and zero for the incorrect answer. This system is called "number right" (Kurz, 1999, p1). However, various other suggestions on scoring and responding to MC items are proposed in the literature (Frary, 1980; Frary, 1989; Hsu et al., 1984). According to Hsu et al. (1984), in the free-choice (multiple marking or multiple select) format instructions, the examinee can mark multiple options if they are correct. This method assumes that the examinee has preconceptions regarding the number of possible correct answers (Frary, 1989). Hsu et al. (1984) reported that in scoring items in the free-choice format, four points are given for the correct answer and minus points for each incorrect marking. According to Frary (1980), two different scoring methods exist for free-choice format items. The first was suggested by Coombs (1953), and the other was by Arnold and Arnold (1970). Thus, different response patterns score numerically. In studies on the development of biology concept inventories, MC items were mainly preferred, while multiple true-false items, open-ended questions, and stratified diagnostic items were less preferred (Queloz et al., 2017). In these studies, single-option marking has been generally used. The chance of the keyed answer, emphasized in the literature, is an essential limitation in this case. In their CI development study, Newman et al. (2016) used the multiple-select directive for MC items. Researchers reported that multiple-select effectively reduces random guessing.

Several studies directly related to protein structure, folding, chemical bonds, and amino acids in its structure have reported misconceptions in the field of biology (Fisher, 1985; Guzman & Bartlett, 2012; Kasapoglu, 2011; Harle & Towns, 2013; Linenberger & Bretz, 2014; Selvi & Yakışan, 2004; Robic, 2010; Villafañe et al., 2011). Robic (2010) summarized common misconceptions concerning protein structure with ten items. Harle and Towns (2013) indicated misconceptions of the formation of primary and secondary protein structures and the roles of chemical bond interactions by the students. Villafañe et al. (2011) have reported misconceptions about bond energy and protein alpha-helix structure. Misconceptions about protein synthesis, genes, DNA, chromosomes, mutation, and protein are also encountered in genetic and molecular genetics studies (Smith & Knight, 2012; Marbach-Ad, 2001; Gericke & Wahlberg, 2013; Smith & Williams, 2007; Guzman & Bartlett, 2012; Kasapoglu, 2011). Following this, White and Bolker (2008) outlined the close relationships between the genetic, biochemistry, and molecular genetic disciplines. Furthermore, several studies have reported protein-related misconceptions regarding diet, digestion, energy gain, and growth (Mak et al., 1999; Yilmaz et al., 2017).

Protein is crucial because it has a widespread teaching process from high school to higher education. It is also a critical core concept for fundamental biology disciplines like biochemistry and molecular genetics (White & Bolker, 2008). This literature shows that protein is an important core concept for different subject areas in biology. Accordingly, a conceptual understanding of the protein has been considered necessary. Therefore, the necessity of developing a PCI formed the basis of this study. The misconceptions were used as distracters, in line with other studies to measure this situation. In the study, the free choice format, which is in the measurement literature and rarely preferred in CI development studies, was chosen as the answering guide. Along with this response guide, unlike other studies, a scoring rule was used to create item scores, giving negative points for incorrect options and positive points for correct options. Thus, in the study, it was tried to determine qualitatively and quantitatively whether the response patterns obtained with multiple answers reflected the cognitive structure of the participants. In this context, the research's operational definition of conceptual understanding was accepted as the absence of misconceptions regarding the intended concepts in the cognitive structure of the learners. In the study, the response patterns were obtained with multiple responses. It was tried to determine whether the test reflects the cognitive structure and to examine the scores, validity, and reliability of these cognitive patterns.

Method

The study was conducted using the methodology reported in CI development studies in the literature. These studies include stages in which various qualitative and quantitative data collection stages are carried out. **Table 1** shows an overview of the CI's development process.

The study's qualitative data were obtained from a systematic literature review and validation interview with participants. Data analysis was provided with qualitative content analysis (Mayring, 2002) using MS Excel and MAXqda trial programs. SPSS, JASP, and MS Excel were used to analyze quantitative data for statistical evidence regarding validity and reliability.

Participants

The study included 317 participants. The validation interviews were conducted with 26 volunteers from the group. The participants were from different educational levels, including high school students (f = 10), university students (f = 13), biology teachers (f = 2), and one instructor.

The remaining 291 participants included 142 high school students (f = 120 females and f = 22 males), 96 university students (f = 80 females and f = 16 males), and 53 instructors. The high school level participants came from different high schools and attended various grade levels. Participants at the university level came from the departments of biology (f = 5), biology education (f = 14), science education (f = 50), and others (f = 27; health sciences, classroom teachers, and chemistry teachers). The teacher participants consisted of biology teachers (f = 31), science teachers (f = 13), and other branches (f = 3), a total of 47 teachers (17 males and 30 females). The average service period of these teachers was 14 years; however, working periods varied between 1 and 35 years. Six faculty members (three females and three males) worked at the university level. The teachers and faculty members were classified as the expert group in both applications of the study. However, this classification does not express any expertise in protein or biochemistry. It is solely based on the level of education gained.

Table 1. Overview of Applications in the Process of PCI Development and Criteria of Validity and Reliability.

Criteria	Applications	Content
Content Validity	1. Literature and fundamental resources reviewing	Identifying misconceptions and determining the conceptual framework for protein
Content validity	2. Forming of CI 2.1. Design of MC-questions 2.1.1. Item writing	
	2.1.2. The validation interviews (participants)	Understanding items by participants, defining misconceptions participants, and wording participants about items The accordance of response with argumentation
Construct validity	2.2. Draft PCI and test specification 2.2.1. Design response instruction and item-scoring rules	
	2.2.2. The validation interviews (experts)	The experts' opinions on the suitability content of PCI determined The accordance of item content with the aimed misconception
Construct validity	3. Statistical analysis 3.1. Validity 3.1.1. Factor analysis 3.1.2. Group comparisons	A large sampling application
	3.2. Reliability 3.3. Item analysis	

Literature and Fundamental Resources Reviewing

The first stage was a literature review to identify common misconceptions. To conduct a thorough literature analysis on misconceptions related to protein, various databases including "Google Scholar," "ERIC," and "SCOPUS" were searched using keywords such as "Protein," "Misconceptions in Biology," and "Misconceptions." Those related to protein and genetic concepts were selected first among the articles accessed. Then, articles that directly or also included protein-related misconceptions address protein-related misconceptions were selected. As a result of this, thirty-four articles from different countries were reached, and misconceptions about protein and protein-related concepts were reviewed. Simultaneously with this stage, essential sources (Nelson & Cox, 2016; Sadava et al., 2011; Simon et al., 2016) were examined. In this way, the conceptual framework of PCI and the list of misconceptions defined in various studies have been reached. Thanks to this list, multiple choice items began to be written to form the PCI.

Forming of Protein Concept Inventory

The researcher wrote items in a conventional MC format using the list of misconceptions area. The MC questions are about these misconceptions, and their options contain the misconceptions. The researcher initially wrote 28 MC questions. Then, validation interviews were conducted to examine these MC questions according to participants' thinking and wording. The participants' thoughts and explanations about protein were investigated with these items. To do this, participants solved the questions by thinking aloud about the questions and making explanations about their answers. The data of the participants from different education levels who participated in this process enabled testing the misconceptions determined with the support of the literature, revealing the students' understandings, and improving the items. These validity interviews with the participants were conducted interactively with question development in individual and group interviews. Based on this process, minor improvements were made to eight questions written at the beginning, significant improvements were made to 20 questions (total changing item roots or options), and seven new questions were written. As a result, 25 MC items were obtained through revisions carried out in three stages.

After creating the draft PCI, we conducted comprehensive interviews with two experts. First, the expert with a doctorate in biochemistry concluded that the items were related to the measured misconception, the keyed options were appropriate, and the conceptual framework for protein was sufficient. Then, the appropriateness of the form was reviewed by the researcher in terms of assessment and measurement according to the related literature (Haladyna & Downing, 1989; Haladyna & Steven, 1989; Nolen et al., 1992). Afterward, it was discussed with a field expert with a doctorate in educational sciences. The expert suggested some improvements for the items, such as making the expressions of item roots similar, avoiding two negative expressions, condensing lengthy option statements, having similar option lengths, and creating meaningful propositions with the question statement. The experts also expressed positive opinions about the test's presentation and response instructions. Twenty-five candidate items formed the PCI for a large sample application.

Design Response Instruction and Item-Scoring Rules

In scoring the participants' response patterns, scoring rules and their meanings were determined under the literature. The scorings of the response (cognitive) patterns are shown in **Table 2**.

Furthermore, these scores were compared with the number-right item scoring methods. Per the response instruction, each distracter (misconception) marked was given minus one (-1) point, and the correct option was given four (4) points; the sum of them formed the item scores. By scoring in this way, the highest score that can be obtained from an item is four (4) points,

Table 2. Comparison of Rules for Item Scoring and Levels of Understanding.

			Item Scores		
Levels of Understanding	Response combinations	Indices	Cognitive scores	pattern	Number-right scores
Scientific Conceptions (SC)	The only TO	I_1	4		1
Partial Understanding (PU)	A WO with TO	I_2	3		0
	Two WO with TO	I_3	2		0
	Three WO with TO	I_4	1		0
	All of the options	I_5	0		0
Lack of Knowledge (LK)	Do not know	I_0	0		0
	Do not know with TO or WO	I_{01}	0		0
Not Understanding (M)	Only a WO	M_1	-1		0
	Two WOs	M_2	-2		0
	Three WOs	M_3	-3		0
	Four WOs	M_4	-4		0

TO: true option; WO: wrong option

and it indicates scientific conception (SC) level (I1), namely, complete comprehension. Other positive scores (I_2–I_4) are called the level of partial understanding (PU). These scores suggest a cognitive structure in which the respondent has combined scientific conception with different misconceptions. Marking all options (I_5) is scored as zero (0) points in both scoring methods. This pattern is classified as PU because the respondents in test items are also given the option "I do not know." Furthermore, if the participant only marked the distracters, the item scores became a minus point (M_1–M_4), indicating a complete misconception or lack of understanding (M).

Statistical Analysis

Validity

Factor Analysis

The hypothesis on the structure of the items, which has been reached based on qualitative results, was tested using factor analysis. Exploratory factor analysis (EFA) was used to test the structure to be measured based on the relationship matrix between the participants' responses. The goodness of fit indices related to this structure was examined with confirmatory factor analysis (CFA). For this, the minimum residual (unweighted least square) method and Promax rotation, one of the oblique rotation techniques, were used for EFA. Both eigenvalue (> 1) and parallel analysis were performed to determine the number of dimensions. Finally, the PCI's construct was confirmed with CFA using JASP software (JASP Team, 2022). This stage was performed resampling using the bootstrap: 5000 and 95% confidence interval (95% CI) criteria.

Groups Comparisons

In high school education, students learn about proteins in various units throughout different educational periods. This means that their understanding of proteins is expected to increase over time. An analysis of variance (ANOVA) was conducted to test the statistical significance of the differences in PCI scores among high school participants from different educational periods. Owing to this analysis another piece of evidence for construct validity was provided based on the significance of the differentiation between groups in test scores.

Reliability and Item Analysis

For reliability, which is also related to validity, indices for internal consistency were calculated considering the construct validity findings. Firstly, Cronbach's alpha as an internal consistency criterion was calculated for reliability estimates compatible with the PCI factor structure. Additionally, item factor loads and error variance values produced by EFA and CFA were used in reliability estimation. Based on this, Guttman's lambda 6 and McDonald's Omega were calculated (Revelle & Zinbarg, 2009; Yurdugul, 2005). In addition, an item analysis was performed to evaluate the psychometric properties of the items. Item difficulty index, item discrimination (in the lower and upper 27% segments), and item-total correlation were calculated using understanding scores for item statistics. In these calculations, positive scores ranging from 0 to 8 were used by adding four to the cognitive pattern (understanding) scores (-4/4). The difficulty indices were calculated by the ratio of the item average to the maximum score that the item could get (Tan, 2016; p243). However, cognitive pattern scores (-4/4) were used in other statistical procedures.

Results

Results of Literature Review

Table 3. Descriptive Statistics of Articles in the Literature Review.

Sample of Studies	f	Methods	f	Type of article	f	Topics	f
Middle School	7	Experimental	3	Literature Review	2	Atom	1
Highschool	4	Document Analysis	5	Review	2	The plants	2
University	16	Mixed	3	Thesis	1	General biology concepts	5
Mixed	2	CI Development	2	Article	29	Circulation Systems	1
Biology Teacher In high school	1	Qualitative	5			Enzyme	5
		Survey	12			Digestive system	1
						Genetics and Genetics Concepts	10
						Protein structure and synthesis	8
						Cell Physiology	1

Previous studies on the misconceptions about protein or related concepts were selected to describe the misconceptions area regarding protein. **Table 3** presents the descriptive information about these articles.

In this way, 34 articles from different countries were reviewed in the literature analysis. The articles reviewed were published between the years 1985 and 2018. The samples of investigated research articles cover 6574 people from different education levels, and the book review studies include 19 national biology books from different countries. Especially review articles on misconceptions regarding the concept of protein (Robic, 2010) and general biology concepts (Vogel, 2000) were included. A part of the examined articles was on genetics and genetics concepts (f = 10), general biology concepts (f = 5), and enzymes (f = 5). In addition to this, articles directly related to protein structure and protein synthesis were found in eight studies. The cases of misconception reached by systematic literature analysis and their distribution according to subject areas are presented in **Table 4**.

Table 4 outlines the conceptual framework and misunderstandings related to proteins. These misunderstandings have been verified by analyzing various general biology and biochemistry resources. Based on these findings, the areas of study that form the conceptual framework of PCI, as well as the framework of misconceptions identified in the literature, have been obtained. This conceptual framework has been utilized to create test specifications for PCI questions. When the content of the test was formed, two main criteria were considered: protein topics and misconceptions. Accordingly, **Table 5** shows the distribution of the questions selected for PCI from the developed items according to topics and misconceptions about protein.

Table 4. The Subject Framework and Misconceptions about Protein.

Subject Domain	Misconceptions Area	Source
Amino acid	Misconceptions about the structure, synthesis, types of amino acids, and relationships among amino acids and genetic concepts	Vogel (2000); Fisher, (1985)
Protein structure	Misconceptions about the structural levels of the protein, its folding mechanism, the bonds in its structure, its formation in the cell, its 3-dimensional formations, denaturation, and its classification and specificity.	Villafañe et al., (2011); Vogel (2000); Smith & Knight (2012); Robic (2010); Lewis et al., (2000); Harle & Towns (2013)
Functions of the proteins	Misconceptions about the roles of protein, enzyme, and hormone concepts in vital events	Tekkaya et al.,(2000); Yip (1998a); Couch et al., (2015); Linenberger & Bretz (2014); Hershey (2004); Dikmenli et al., (2009)
Protein synthesis	Misconceptions about "The Central Dogma" information flow, Transcription, Translation, start and stop mechanisms, code, gene, codon, mRNA, anticodon, tRNA, and polysome.	Cho et al. (1985); Smith & Williams (2007); Guzman & Bartlett (2012); Kasapoglu (2011)
Genetic concepts and mutation	The misconceptions regarding gene, DNA, genotype to be protein, the role of the genetic codes in protein synthesis, and the effect of mutation.	Smith & Knight (2012); Cho et al. (1985); Gericke & Wahlberg (2013) Temelli (2006); Marbach-Ad (2001); Lewis et al., (2000)
Nutrient – Digestion – Cell - Metabolism	Misconceptions regarding the protein content of foods, the digestive mechanism, the state of their use in energy production, and that protein synthesis occurs only in eukaryotic cells.	Lewis et al. (2000); Yip, (1998b); Yilmaz et al.(2017); Mak et al. (1999) Wynn et al. (2017); Herrmann-Abell et al. (2016)

Results of Validation Interviews

With the validity-interviews analysis, 681 codes regarding different topics misconceptions. According to these findings, misconceptions about protein structure have the highest frequency (34.5%), followed by protein synthesis (21.29%) and genetic concepts (12.78%). Accordingly, it can be said that the misconceptions determined in the literature can be detected in the participants with the prepared items. Although the small number of participants in the validation interview and the absence of a quantitative sampling did not allow interpretations regarding the prevalence of the misconceptions, it indicates that the common misconceptions defined in the literature can be measured in these participants. However, it is critical to determine whether the options containing the misconceptions in the response patterns provided by the participants with multiple markings reflect their cognitive status. Thus, the consistency between the participants' response patterns and explanations has been investigated thanks to their marked options and explanations. **Table 6** presents some of these findings.

Table 6 shows the options containing the misconceptions selected by participants for the MC items. The explanations for these answers also include these misconceptions. Item 18 asked which of the following does not

Table 5. Distribution of Protein Concept Inventory Items by Topic and Subtopics.

Topics	Misconception Area	f	Item No
Genetic Concepts	Protein synthesis	1	3
	Protein diversity	1	12
	Mutation	1	17
	Genotype-phenotype protein relationship	1	18
	Protein-containing	1	23
Metabolism	Protein metabolism	1	6
Protein Functionality	Structure-function relationship	1	5
	Protein activity	1	11
	Specific binding	1	16
Protein Synthesis	Cell types	1	4
	Protein synthesis process	2	10-14
	Cell	1	15
Functions of the protein	Cellular functions	2	9 - 19
Structure of Protein	Amino acid	2	1 - 7
	Peptide bond	1	2
	Chemical bonds	1	8
	Types of protein	1	13
	Features of protein structure	1	20
	Formation of the original structure	1	21
	Distinctive feature	1	22
	Denaturation	1	24
	Structural levels	1	25

Table 6. Marked Options and Related Explanation.

Marked options (code of participants/question number)	The reason (code of participants/question number)
"Amino acids synthesize in the ribosome" (13/Question 2) "Amino acids are formed by the translation" (11/Question 2) "Amino acid is produced by tRNA. Amino acid synthesized in the ribosome" (5/Question 2)	"Ribosomes synthesize protein; therefore, they are the source of amino acids." (13/Question 2) "Amino acids are formed as a result of transcription. Encoded as reading (by translation)" (11/Question 2) "(Answer is) 2 (tRNA) and 3 (translation) because there is a production" (5/Question 2; 12). "Because the ribosome is a protein-specific organelle and synthesizes amino acids." (21/Question 2)
"Neuron, fungal cell, digestive tract, plant cell" (3/Question 18) "Neuron, digestive tract" (1/Question 18) "Neuron" (4/Question 18)	"Protein is produced in the ribosome. The ribosome is found in prokaryotes and eukaryotes in all living things. Bacteria produce protein. The neuron, fungal, plant cells, and digestive tract contain proteins, but these do not produce it." (3/Question 18) "For a protein-producing, it must have a ribosome organelle. It is absent in the neuron and digestive tract." (1/Question 18) "Neurons cannot reproduce, they cannot renew themselves (therefore) they cannot produce a protein (my answer) Neuron (nerve cell)" (4/Question 18)

realize protein synthesis. This question measures the misconception that some organisms or cell types cannot realize protein synthesis. According to their explanation in quotations, some participants (1, 3, and 4) believe that because neurons cannot divide or do not have ribosomes, they cannot produce protein. Similarly, the second question, which measures the misconceptions that amino acids are not synthesized in metabolic pathways in the cell, reveals the participants' misconception of the protein synthesis mechanism. These findings lead to the result that the response patterns formed by the developed questions are compatible with the cognitive structures of the participants.

Findings Regarding the Validity and Reliability

EFA Findings

The Kaiser–Meyer–Olkin value (0.939) in the EFA results is greater than 0.80, and Bartlett's test (Chi-square value = 2301.348. $p < 0.001$) is statistically significant. These findings show that the dataset is suitable for factor analysis in terms of sample size and correlation between items. Based on these findings, the scree plot was created for both a parallel analysis and eigenvalues.

Figure 1 shows that only one data point in the graph has an eigenvalue greater than the simulation data (triangle). The values indicate that the one-dimensional model developed by parallel analysis is acceptable. The variance explained by one dimension is 0.309, and the sum of the squares of the factor loads is 7.736. Based on this, the model has been accepted as one-dimensional.

Modification indices were obtained using CFA for the single-factor structure obtained by EFA. The goodness-of-fit criteria (X^2/df: 1.386; comparative fit index: 0.949; Tucker–Lewis Index: 0.944 root-mean-square error of approximation: 0.036; goodness-of-fit index: 0.907) were found to have a good level of fit in the CFA assessment of the one-dimensional structure obtained with the EFA result model (Hu & Bentler, 1999; Kline, 2005). **Table 7** shows the factor loads of the items calculated with CFA and EFA.

The CFA results (bootstrap 5000 and 95% CI) show that the predicted values for the items are significant within the confidence intervals. The factor loads calculated by CFA ranged from 0.346 to 0.655, and the EFA results were close. These results show that PCI has a one-dimensional factor structure.

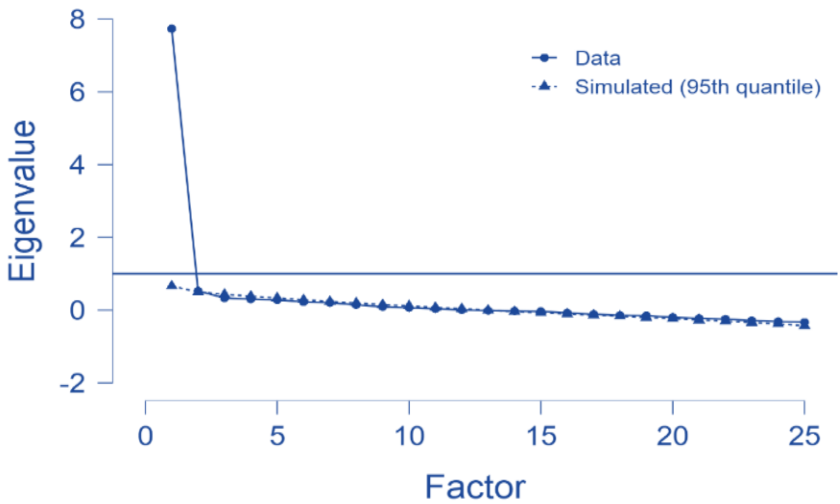

Figure 1. Scree Plot of the Eigenvalues and Parallel Analysis.

Table 7. Results for EFA and CFA.

	CFA						EFA	
				95% CI				
Items	Estimate	SE	z-value*	LL	UL	Std. Factor loads	Factor loads	Uniqueness
S1	1.000	0.000		1.000	1.000	0.404	0.403	0.838
S2	1.083	0.176	6.152	0.809	1.527	0.584	0.585	0.658
S3	1.331	0.222	6.004	0.981	1.915	0.552	0.554	0.693
S4	1.295	0.213	6.067	0.961	1.872	0.565	0.561	0.685
S5	1.144	0.202	5.667	0.815	1.680	0.488	0.489	0.761
S6	0.744	0.161	4.630	0.451	1.149	0.346	0.348	0.879
S7	1.281	0.205	6.236	0.933	1.875	0.604	0.603	0.637
S8	1.161	0.207	5.607	0.848	1.657	0.478	0.480	0.769
S9	1.339	0.223	6.013	0.967	1.955	0.554	0.553	0.694
S10	1.603	0.257	6.237	1.190	2.288	0.604	0.606	0.633
S11	1.119	0.201	5.561	0.793	1.619	0.471	0.473	0.776
S12	1.585	0.249	6.378	1.222	2.229	0.641	0.639	0.592
S13	1.371	0.230	5.960	1.017	1.965	0.543	0.543	0.705
S14	1.673	0.260	6.428	1.279	2.356	0.655	0.653	0.574
S15	1.286	0.210	6.133	0.942	1.851	0.580	0.578	0.666
S16	1.663	0.263	6.332	1.252	2.371	0.628	0.631	0.602
S17	1.477	0.230	6.409	1.115	2.084	0.649	0.649	0.579
S18	0.769	0.142	5.420	0.513	1.111	0.448	0.450	0.797
S19	1.195	0.202	5.925	0.860	1.718	0.536	0.537	0.711
S20	1.496	0.247	6.047	1.102	2.142	0.561	0.562	0.685
S21	1.597	0.259	6.170	1.192	2.268	0.588	0.586	0.656
S22	1.587	0.253	6.283	1.211	2.226	0.616	0.616	0.621
S23	1.409	0.225	6.276	1.047	2.006	0.614	0.613	0.624
S24	1.403	0.224	6.259	1.060	1.966	0.610	0.610	0.627
S25	0.960	0.177	5.408	0.633	1.454	0.447	0.445	0.802

p < 0.001

Findings Related To Group Comparisons

Table 8 shows the ANOVA test results regarding the differences in the mean scores of the high school participants based on their grade levels.

Table 8 shows a significant difference in grade levels ($F(3; 138) = 8.083$, $p = 0.000$). Multiple comparison results show the source of this difference. The difference between the ninth grade and other grades—10th grade ($MD = -10.00$; $SE = 2.8$), 11th grade ($MD = -11.85$; $SE = 3.741$), and 12th grade ($MD = -17.51$; $SE = 3.405$)—is statistically significant at 95% CI. The mean score of ninth-grade classes ($M = 0.19$; $SD = 10.38$) is significantly lower than those of other grade levels. Although no significant difference was found between the averages of understanding scores of other grade levels (10−12), it is observed that the mean increases as the grade level increases.

Item Analysis and Reliability

The item statistics and reliability were calculated using the data's item scores. **Table 9** shows the results of the item analysis.

It can be seen that the item difficulty values range from 0.22 to 0.63. These values show that the PCI includes items with acceptable difficulty levels ($0.20 < p < 0.80$; Crocker & Algina, 1986, p98; Fisseni, 1997, as cited in Bühner, 2006, p140). When the correlational item discrimination indexes (r_{itc}) are examined, it is found that while they are low (< 0.30) for item six, they are medium to high (> 0.50) for the remaining items. The discrimination index (D, 27%) changed between 0.25 and 0.67. These findings indicate that the item discriminations are acceptable (> 0.20; Crocker & Algina, 1986, p315). With these data, the internal consistency criterion for reliability estimate was determined and is shown in **Table 10**.

When the table is evaluated, the reliability estimates calculated with different internal consistency measures of the developed PCI show that the test has a significant internal consistency and structural reliability.

Discussion

Validity and Reliability

The study was carried out the standards used in developing the CI and the criteria for measurement and evaluation by considering the relevant literature (Adams & Wieman, 2011; Crocker & Algina, 1986, pp217-218; Bühner, 2006, p36). Accordingly, concept domain determination comes first among the applications used for content validity (Crocker & Algina, 1986, pp217-

Table 8. One-Way ANOVA Results for the Grade Levels of High School Level Participants.

High School Grades	Mean	SD	n	F(3, 138)	η^2
9	0.19	10.38	31		
10	10.20	14.17	46		
11	12.05	14.86	21	8.083*	0.149
12	17.70	18.90	44		
Total	10.61	16.34	142		

**p < 0.05.*

Table 9. Item Analysis Results in Understanding Scores.

Items	Mean	Std. Dev.	Item Total correlations (r_{itc})	Index of Discrimination (D)	Item Difficulties (p)
S1	4.96	3.58	0.40	0.44	0.62
S2	3.27	2.70	0.60	0.42	0.41
S3	3.31	3.43	0.52	0.44	0.41
S4	4.46	3.29	0.53	0.53	0.56
S5	2.92	3.33	0.50	0.46	0.37
S6	4.55	3.04	0.22	0.25	0.57
S7	3.19	3.03	0.58	0.43	0.40
S8	5.01	3.49	0.47	0.49	0.63
S9	4.59	3.47	0.54	0.55	0.57
S10	2.74	3.77	0.59	0.54	0.34
S11	2.73	3.40	0.45	0.47	0.34
S12	3.92	3.54	0.60	0.55	0.49
S13	4.21	3.61	0.53	0.53	0.53
S14	4.09	3.67	0.65	0.67	0.51
S15	3.84	3.18	0.57	0.48	0.48
S16	2.98	3.77	0.61	0.57	0.37
S17	3.36	3.26	0.62	0.47	0.42
S18	3.34	2.48	0.44	0.29	0.42
S19	3.08	3.20	0.52	0.43	0.38
S20	3.44	3.80	0.54	0.54	0.43
S21	3.99	3.86	0.57	0.58	0.50
S22	3.70	3.68	0.58	0.65	0.46
S23	3.74	3.29	0.61	0.50	0.47
S24	3.32	3.27	0.58	0.48	0.41
S25	1.76	3.05	0.45	0.29	0.22

Note: n = 291; Test Mean: 90.485; Test Variance: 1885.954; Test Std. Dev: 43.428

Table 10. Frequentist Scale Reliability Statistics.

Estimate	McDonald's ω	Cronbach's α	Guttman's λ_6	Average Inter-item Correlation
Point estimate	0.916	0.914	0.921	0.298
95% CI lower bound	0.902	0.899	0.912	0.256
95% CI upper bound	0.930	0.927	0.940	0.341

218; Bühner, 2006, p36; Lindell et al., 2007). Adams and Wieman (2011), for example, described the structure and purpose of the field to measure as a standard. The content of the PCI was determined with systematic literature reviews. This content includes the subject, subtopic, and related misconceptions in a comprehensive framework from high school to higher education. Furthermore, qualitative interviews obtained the proposed expert opinions on content validity (Crocker & Algina, 1986).

The study provided evidence for construct validity in ways compatible with the literature (Adams & Wieman, 2011; Briggs et al., 2017; Kalas et al., 2013; Smith & Knight, 2012). These qualitative findings showed that item roots and options were first appropriately understood according to their wording participants, and second, the candidate items of the inventory revealed the participants' common misconceptions. The findings also showed that respondents' responses reflected their understanding and cognitive structure (**Table 6**).

In addition, the literature suggested that factor analysis techniques and differentiation between-group techniques effectively provided statistical evidence for construct validity (Crocker & Algina, 1986, p. 231; Kummer et al., 2019; Ramlo, 2008). Accordingly, the one-dimensionality of the EFA results was supported by the theoretical expectation and verified by CFA. The factor loads were within the desired values (> 0.30; **Table 7**) in the literature (Floyd & Widaman, 1995; Kline, 1994). Another proof of construct validity is the statistical significance of the difference among educational levels regarding PCI scores. While education levels include a wider interval for teaching the concept of protein, there are minor teaching differences between high school classes, including more similar age groups. Therefore, a significant difference in understanding of protein was expected in favor of the groups that received more education. The result provides evidence that the PCI discriminates between groups with and without more education about protein, consistent with theoretical expectations.

Internal consistency indices were calculated using different techniques for the reliability evidence of the developed PCI. Shevlin et al. (2000, as cited in Bühner, 2006, p134) recommend performing CFA to ensure one-dimensionality and calculating Cronbach's α accordingly. In addition to Cronbach's α coefficient, McDonald's ω and Guttman's λ6 coefficients were determined. These reliability coefficients are indices more sensitive to test structure, namely, dimensionality. In this context, McDonald's ω has been reported as an important index for structure reliability (Revelle & Zinbarg, 2009; Yurdugül, 2005). The high level of internal consistency coefficients (> 0.90) calculated in this way (**Table 10**) indicates that the developed inventory has a high internal consistency. According to the item analysis results (**Table 9**), the items were by the reference ranges requested in the measurement literature (Crocker & Algina, 1986; Bühner, 2006) and the CI development studies (Kalas et al., 2013; Paustian et al., 2017; Jarrett et al., 2012). The findings summarized are strong evidence for important psychometric properties of the PCI for measurement instruments.

Features of the PCI

The item score is the sum of the patterns for each option with a misconception score of minus one and a scientific conceptualization score of four points. In this way, PCI can produce scores for each item between −4 and 4 and for total of −100 and 100 points. In this study's scoring approach, whether or not there is a misconception in the response patterns is an important criterion. This criterion is commonly used in the literature for rubrics regarding understanding measurement (Naah, 2015). This situation supports our operational definition and PCI's scoring system (**Table 2**). The scoring properly matches the study's operational definition of conceptual understanding and allows us to evaluate in a broader range than classical scoring (0–1). When examining understanding scores, as these scores increase from negative to positive, comprehension improves, decreasing misconceptions in the cognitive structure. This condition leads to a more sensitive quantitative measurement and evaluation of participant differences. Furthermore, the PCI has provided a limited representation of cognitive structure thanks to the multiple markings. In this way, the participants' response patterns qualitatively show misconceptions. Thus, processing the data collected with the PCI using data analytics approaches (MS Excel, Power Pivot, or Dashboards) can quickly provide qualitative and quantitative information about the students' levels of understanding individually or at the class level to teachers.

Conclusions

Qualitative research findings and, in line with this, large sample research findings show that PCI measures protein-related misconceptions. In addition, the statistical analysis provided evidence regarding the validity and reliability of PCI. Accordingly, the PCI developed under the basic principles of the literature measures the level of understanding of the protein concept (Adams & Wieman, 2011; Briggs et al., 2017; Garvin-Doxas et al., 2007; Kalas et al., 2013; Libarkin, 2008). This study contributes to studies on CI development with the distracter-driven items presented in the literature and quantitative and qualitative analyses of cognitive pattern scores based on free-choice (multiple marking). Obtaining both quantitative and qualitative results with MC items, which are preferred due to their rapid, objective, and quantitative scoring features in the literature, contribute significantly to the evaluation of conceptual understanding both individually and as a group. PCI is an effective tool for researchers and teachers to measure understanding. PCI can help identify areas that require further teaching by analyzing the conceptual patterns that lead to misconceptions. Teachers can also use PCI to assess their students' prior knowledge and adjust their teaching accordingly. Additionally, PCI will be helpful for formative assessment purposes.

Ethics Statement

The participants were presented with an informed consent form before participating. The participants were informed that their participation in the test was entirely voluntary and that they could leave the study at any time. During the data collection, the data were anonymized, and no personal and corporate information was collected or used in the research. This study was conducted with the approval of Erzincan Binali Yıldırım University Human Research Ethics Committee (Protocol Number: 05/15).

Acknowledgment

I thank the field experts for their participation in the expert interviews and their valuable opinions, as well as all of the study's volunteer participants.

References

Adams, W. K., & Wieman, C. E. (2011). Development and Validation of Instruments to Measure Learning of Expert-Like Thinking. *International Journal of Science Education*, 33(9):1289-1312. DOI: https://doi.org/10.1080/09500693.2010.512369

Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, 39(10):952-978. DOI: https://doi.org/10.1002/tea.10053

Briggs, A. G., Hughes, L. E., Brennan, R. E., Buchner, J., Horak, R. E. A., Amburn, D. S. K., McDonald, A. H., Primm, T. P., Smith, A. C., Stevens, A. M., Yung, S. B., & Paustian, T. D. (2017). Concept Inventory Development Reveals Common Student Misconceptions about Microbiology. *Journal of Microbiology & Biology Education*, 18(3). DOI: https://doi.org/10.1128/jmbe.v18i3.1319

Bühner, M. (2006). Einführung in die Test- und Fragebogenkonstruktion (2nd ed.). Pearson Studium.

Bybee, R. W. (2012). The Next Generation of Science Standards : Implications for Biology Education. *The American Biology Teacher*, 74(8):542-549. DOI: https://doi.org/10.1525/abt.2012.74.8.3

Cho, H. H, Kahle, J. B., & Nordland, F. H. (1985). An investigation of high school biology textbooks as sources of misconceptions and difficulties in genetics and some suggestions for teaching genetics. *Science Education*, 69(5):707-719. DOI: https://doi.org/10.1002/sce.3730690512

Couch, B. A., Wood, W. B., & Knight, J. K. (2015). The Molecular Biology Capstone Assessment : A Concept Assessment for Upper-Division Molecular Biology Students. *CBE-Life Sciences Education*, 14:1-11. DOI: https://doi.org/10.1187/cbe.14-04-0071

Crocker, L., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. Wadsworth Group- Thomson Learning.

D'Avanzo, C. (2008). Biology concept inventories: Overview, status, and next steps. *BioScience*, 58(11):1079-1085. DOI: https://doi.org/10.1641/b581111

Dikmenli, M., Çardak, O., & Öztaş, F. (2009). Conceptual problems in biology-related topics in primary science and technology textbooks in Turkey. *International Journal of Environmental and Science Education*, 4(4):429-440.

Fisher, K. M. (1985). A misconception in biology: Amino acids and translation. *Journal of Research in Science Teaching*, 22(1):53-62. DOI: https://doi.org/10.1002/tea.3660220105

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the aevelopment and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3):286-299. DOI: https://doi.org/10.1037/1040-3590.7.3.286

Frary, R. B. (1980). The effect of misinformation, partial information, and guessing on expected multiple-choice test item scores. *Applied Psychological Measurement*, 4(1):79-90. DOI: https://doi.org/10.1177/014662168000400109

Frary, R. B. (1989). Partial-Credit scoring methods for Multiple-Choice tests. *Applied Measurement in Education*, 2(1):79-96. DOI: https://doi.org/10.1207/s15324818ame0201

Garvin-Doxas, K., Klymkowsky, M., & Susan, E. (2007). Building, using, and maximizing the impact of concept inventories in the biological sciences: Report on a national science foundation-sponsored conference on the construction of concept inventories in the biological sciences. *CBE Life Sciences Education*, 6:297-306. DOI: https://doi.org/10.1187/cbe.07

Gericke, N., & Wahlberg, S. (2013). Clusters of concepts in molecular genetics: A study of Swedish upper secondary science students understanding. *Journal of Biological Education*, 47(2):73-83. DOI: https://doi.org/10.1080/00219266.2012.716785

Guzman, K., & Bartlett, J. (2012). Using simple manipulatives to improve student comprehension of a complex biological process: Protein synthesis. *Biochemistry and Molecular Biology Education*, 40(5):320-327. DOI: https://doi.org/10.1002/bmb.20638

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1):37-50. DOI: https://doi.org/10.1207/s15324818ame0201

Haladyna, T. M., & Steven, M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1):51-78. DOI: https://doi.org/10.1207/s15324818ame0201

Harle, M., & Towns, M. H. (2013). Students' understanding of primary and secondary protein structure: Drawing secondary protein structure reveals student understanding better than simple recognition of structures. *Biochemistry and Molecular Biology Education*, 41(6):369-376. DOI: https://doi.org/10.1002/bmb.20719

Herrmann-Abell, C. F., & DeBoer, G. E. (2011). Using distracter-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. *Chemistry Education Research and Practice*, 12(2):184-192.DOI: https://doi.org/10.1039/c1rp90023d

Herrmann-Abell, C. F., Koppal, M., & Roseman, J. E. (2016). Toward high school biology: Helping middle school students understand chemical reactions and conservation of mass in nonliving and living systems. *Cell Biology Education*, 15(4).DOI: https://doi.org/10.1187/cbe.16-03-0112

Hershey, D. R. (2004). Avoid misconceptions when teaching about plants. Available at: http://Www.Actionbioscience.Org/Education/Hershey.Html?Print. & https://files.eric.ed.gov/fulltext/ED501356.pdf

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3):141-158. DOI: https://doi.org/10.1119/1.2343497

Hsu, Tse-Chi, M., Moss, P. A., & Khampalikit, C. (1984). The merits of multiple-answer items as evaluated by using six scoring formulas. *The Journal of Experimental Education*, 52(3):152-158. Available at: http://www.jstor.org/stable/20151542

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 6(1):1-55. DOI: http://www.tandfonline.com/doi/abs/10.1080/10705519909540118

Jarrett, L., Ferry, B., & Takacs, G. (2012). Development and validation of a concept inventory for introductory-level climate change science. *International Journal of Innovation in Science and Mathematics Education*, 20(2). Available at: https://openjournals.library.sydney.edu.au/CAL/article/view/5814

JASP Team (2022). JASP (Version 0.16.4) [Computer software].

Jorion, N., & Gane, B. D., & DiBello, L. V., & Pellegrino, J. W. (2015, June): Developing and Validating a Concept Inventory [Paper presentation]. *ASEE Annual Conference & Exposition*, Seattle, Washington. DOI: https://doi.org/10.18260/p.23836

Kalas, P., O'Neill, A., Pollock, C., & Birol, G. (2013). Development of a meiosis concept inventory. *CBE Life Sciences Education*, 12(4):655-664. DOI: https://doi.org/10.1187/cbe.12-10-0174

Kasapoglu, E. (2011). The effect of teaching protein synthesis using concept maps on the academic achievement and attitudes of 12th grade high school students. Master thesis. Sel çuk University.

Kline, P. (1994). An easy guide to factor analysis. Routledge.

Kline, R. B. (2005). Principles and practice of structural equation modeling (Second). The Guilford Press.

Klymkowsky, M. W., & Garvin-Doxas, K. (2008). Recognizing student misconceptions through Ed's tools and the biology concept inventory. *PLoS Biology*, 6(1):0014-0017. DOI: https://doi.org/10.1371/journal.pbio.0060003

Kummer, T. A., Whipple, C. J., Bybee, S. M., Adams, B. J., & Jensen, J. L. (2019). Development of an evolutionary tree concept inventory. *Journal of Microbiology & Biology Education*, 20(2). DOI: https://doi.org/10.1128/jmbe.v20i2.1700

Kurz, T. B. (1999). A review of scoring algorithms for multiple-choice tests. *Annual Meeting of the Southwest Educational Research Association* (San Antonio, TX, January 21-23, 1999, 1:1-21. Available at: https://files.eric.ed.gov/fulltext/ED428076.pdf

Lewis, J., Leach, J., & Wood-Robinson, C. (2000). All in the genes? Young people's understanding of the nature of genes. *Journal of Biological Education*, 34(2):74-79. DOI: https://doi.org/10.1080/00219266.2000.9655689

Libarkin, J. (2008). Concept inventories in higher education science. STEM Education Workshop 2. Available at: https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_072624.pdf

Lindell, R. S., Peak, E., & Foster, T. M. (2007). Are they all created equal? A comparison of different concept inventory development methodologies. *AIP Conference Proceedings*, 883:14-17. https://doi.org/10.1063/1.2508680

Linenberger, K. J., & Bretz, S. L. (2014). Biochemistry students' ideas about shape and charge in enzyme-substrate interactions. *Biochemistry and Molecular Biology Education*, 42(3):203-212. DOI: https://doi.org/10.1002/bmb.20776

Mak, S. Y., Yip, D. Y., & Chung, C. M. (1999). Alternative conceptions in biology-related topics of integrated science teachers and implications for teacher education. *Journal of Science Education and Technology*,

8(2):161-170. DOI: http://www.jstor.org/stable/40188526
Marbach-Ad, G. (2001). Attempting to break the code in student comprehension of genetic concepts. *Journal of Biological Education*, 35(4):183-189. DOI: https://doi.org/10.1080/00219266.2001.9655775
Mayring, P. (2002). Einführung in die qualitative sozialforschung (5. edition). Beltz Verlag.
McFarland, J. L., Price, R. M., Wenderoth, M. P., Martinková, P., Cliff, W., Michael, J., Modell, H., & Wright, A. (2017). Development and validation of the homeostasis concept inventory. *CBE Life Sciences Education*, 16(2):1-13. DOI: https://doi.org/10.1187/cbe.16-10-0305
Naah, B. (2015). Enhancing preservice teachers' understanding of students' misconceptions in learning chemistry. *Journal of College Science Teaching*, 45(2):41-47. DOI: https://doi.org/10.2505/4/jcst15_045_02_41
Nelson, D. L., & Cox, M. M. (2016). Lehninger Principles of Biochemistry(Y. M. Elçin (ed.); 5th ed.). Palme Publshing.
Newman, D. L., Snyder, C. W., Fisk, J. N., & Wright, L. K. (2016). Development of the central dogma concept inventory (CDCI) assessment tool. *CBE Life Sciences Education*, 15(2). DOI: https://doi.org/10.1187/cbe.15-06-0124
Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1992). Uses and abuses of achievement test scores. *Educational Measurement: Issues and Practice*, 1992:9-15. DOI: https://doi.org/10.1111/j.1745-3992.1992.tb00234.x
Paustian, T. D., Briggs, A. G., Brennan, R. E., Boury, N., Buchner, J., Harris, S., Horak, R. E. A., Hughes, L. E., Katz-Amburn, D. S., Massimelli, M. J., McDonald, A. H., Primm, T. P., Smith, A. C., Stevens, A. M., & Yung, S. B. (2017). Development, validation, and application of the microbiology concept inventory. *Journal of Microbiology & Biology Education*, 18(3). DOI: https://doi.org/10.1128/jmbe.v18i3.1320
Queloz, A. C., Klymkowsky, M. W., Stern, E., Hafen, E., & Köhler, K. (2017). Diagnostic of students' misconceptions using the Biological Concepts Instrument (BCI): A method for conducting an educational needs assessment. *PLoS ONE*, 12(5):1-18. DOI: https://doi.org/10.1371/journal.pone.0176906
Ramlo, S. (2002). The Force and Motion Conceptual Evaluation. *Annual Meeting of the Mid-Western Educational Research Association*. Available at: https://files.eric.ed.gov/fulltext/ED471542.pdf
Ramlo, S. (2008). Validity and reliability of the force and motion conceptual evaluation. *American Journal of Physics*, 76(9):882-886. DOI: https://doi.org/10.1119/1.2952440
Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on sijtsma. *Psychometrika*, 74(1):145-154. DOI: https://doi.org/10.1007/s11336-008-9102-z
Robic, S. (2010). Mathematics, thermodynamics, and modeling to address ten common misconceptions about protein structure, folding, and stability. *Cell Biology Education*, 9:323-332. DOI: https://doi.org/10.1187/cbe.10
Sadava, D., Hillis, D. M., Heller, H. C., & Berenbaum, M. R. (2011). Life: The Science of Biology (E. Gündüz & İ. Türkan (eds.); 9th ed.). Palme Publishing.
Selvi, M., & Yakışan, M. (2004). Misconceptions about enzymes in university students. *Gazi Eğitim Fakültesi Dergisi*, 24(2):173-182. DOI: https://doi.org/10.17152/GEFD.02048
Shin, J., Guo, Q., & Gierl, M. J. (2019). Multiple-choice item distracter development using topic modeling approaches. *Frontiers in Psychology*, 10:1-14. DOI: https://doi.org/10.3389/fpsyg.2019.00825
Simon, E., Dickey, J., Reece, J., & Hogan, K. (2016). Campbell Essential Biology With Physiology (E. Gündüz & İ. Türkan (eds.); 5th ed.). Palme Publishing.
Smith, L. A., & Williams, J. M. (2007). "It's the X and Y thing": Cross-sectional and longitudinal changes in children's understanding of genes. *Research in Science Education*, 37(4):407-422. DOI: https://doi.org/10.1007/s11165-006-9033-6
Smith, M. K., & Knight, J. K. (2012). Using the genetics concept assessment to document persistent conceptual difficulties in undergraduate genetics courses. *Genetics*, 191(1):21-32. DOI: https://doi.org/10.1534/genetics.111.137810
Tan, Ş. (2016). Measurement and Evaluation in

Teaching (12th ed.). Pegem Akademi.
Tekkaya, C., Çapa, Y., & Yılmaz, Ö. (2000). Biology teacher candidates' misconceptions about biology. *HU Journal of Faculty of Education*, 18:140-147. Available at: https://dergipark.org.tr/tr/pub/hunefd/issue/7818/102721#article_cite
Temelli, A. (2006). Determination of misconceptions concerning genetic subjects of high school students'. *Kastamonu Education Journal*, 14(1):73-82. Available at: https://dergipark.org.tr/tr/download/article-file/819169
Villafañe, S. M., Loertscher, J., Minderhout, V., & Lewis, J. E. (2011). Uncovering students' incorrect ideas about foundational concepts for biochemistry. *Chemistry Education Research and Practice*, 12(2):1-6. DOI: https://doi.org/10.1038/nn.2744
Vogel, S. (2000). Mythology in introductory biology. Resonance, 2000:86-95.
White, B. T., & Bolker, E. D. (2008). Interactive computer simulations of genetics, biochemistry, and molecular biology. *Biochemistry and Molecular Biology Education*, 36(1):77-84. DOI: https://doi.org/10.1002/bmb.20152
Wynn, A. N., Pan, I. L., Rueschhoff, E. E., Herman, M. A. B., & Archer, E. K. (2017). Student Misconceptions about Plants - A First Step in Building a Teaching Resource. *Journal of Microbiology & Biology Education*, 18(1):0-13. DOI: https://doi.org/10.1128/jmbe.v18i1.1253
Yilmaz, M., Ertunç G., Hatun Diken, E., & Çimen, O. (2017). The analysis of biology topics in the 8th grade science textbook in terms of scientific content. *Erzincan University Journal Faculty of Education*, 19(3):17-35. DOI: https://doi.org/10.17556/erziefd.330600
Yip, D. Y. (1998a). Identification of misconceptions in novice biology teachers and remedial strategies for improving biology learning. *International Journal of Science Education*, 20(4):461-477. DOI: https://doi.org/10.1080/0950069980200406
Yip, D. Y. (1998b). Teachers' misconceptions of the circulatory system. *Journal of Biological Education*, 32(3):207-215. DOI: https://doi.org/10.1080/00219266.1998.9655622
Yurdugul, H. (2005). The comparison of reliability coefficients in parallel, tau-equivalent, and congeneric measurements. *Ankara University Journal of Faculty of Educational Sciences*, 39(1):15-37. DOI: https://doi.org/10.1501/egifak_0000000127

Received: February 25, 2024
Revised: March 4, 2024
Accepted: March 25, 2024