



Development and Validation of a Four-Tier Test for the Assessment of Secondary School Students' Conceptual Understanding of Amino Acids, Proteins, and Enzymes

Katarina B. Putica¹

Accepted: 6 October 2022 / Published online: 17 October 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

Previous studies noted the scantiness of diagnostic instruments for the assessment of students' understanding of fundamental biochemistry concepts. Consequently, within this study, a four-tier test for the examination of secondary school students' conceptual understanding of amino acids, proteins, and enzymes has been developed. Items in the test consist of the answer tier, the reason tier, and a confidence rating for each of these tiers. Following the three-phase development process, the final version of the test, comprising eight items, was administered to 123 students in the main study. The internal consistency of both cognitive scores and confidence ratings proved to be adequate (Cronbach's alpha values were 0.76 and 0.87, respectively) and the test–retest reliability was also satisfactory (Pearson's r value was 0.74 for the cognitive scores and 0.88 for the confidence ratings). If the correct responses to both the answer and the reason tier of items are considered, the mean score on the test was 3 out of 8 (38%). The students were more confident when producing the correct compared to the wrong answers but, overall, their confidence was not high (3.65 out of 6). Furthermore, the confidence in the incorrect answers that were provided (3.52 out of 6) indicates the presence of alternative conceptions. Such results show that the content regarding amino acids, proteins, and enzymes is conceptually challenging for the students. Therefore, the test can help secondary school teachers to uncover the students' conceptual difficulties related to this content and develop effective strategies for their remediation and prevention.

Keywords Four-tier test · Conceptual understanding · Amino acids · Proteins · Enzymes · Secondary school students

✉ Katarina B. Putica
puticakatarina@gmail.com

¹ Innovative Centre of the Faculty of Chemistry, University of Belgrade, Studentski trg 12-16, 11000 Belgrade, Republic of Serbia

Introduction

Thus far, our insight into students' conceptual understanding of biochemistry has been limited (Villafañe et al., 2011). Furthermore, it was noted that educators lack efficient and reliable diagnostic instruments to explore students' understanding of fundamental concepts in this field (Bretz & Linenberger, 2012; Villafañe et al., 2021).

According to constructivism, each student actively forms its own concepts by building new information upon the foundation of prior knowledge and experiences (Powell & Kalina, 2009). However, some of the notions that students bring to the classroom, which are deeply rooted in their ways of thinking, can be scientifically incorrect (Treagust, 1988). Such conceptions that differ from those commonly held in the scientific circles are labeled as alternative conceptions (ACs) (Wandersee et al., 1994). Besides being highly resistant to change, ACs can easily be transferred between courses, causing a long-term detrimental effect on the learning process, which spans across several interrelated disciplines (Yong & Kee, 2017).

Concepts regarding amino acids, proteins, and enzymes are fundamental for biochemistry and of great importance for the entire field of life sciences (Tansey et al., 2013). Ordinarily, students are introduced to these concepts at the secondary school level. Consequently, this represents a crucial time for the detection and remediation of the corresponding ACs, thus preventing their transfer to the university level and a prolonged negative impact on learning in fields ranging from biotechnology and agriculture to various health sciences.

To enable the assessment of secondary school students' conceptual understanding of amino acids, proteins, and enzymes, within this study, the four-tier test entitled 4AAPE has been developed. Prior studies only explored conceptual challenges related to the given content area among university students. Furthermore, none of them used four-tier tests, which overcome the limitations of all the previously developed diagnostic instruments for the examination of conceptual understanding (Caleon & Subramaniam, 2010a). The present study provides a detailed overview of the three-phase process of the development and validation of 4AAPE, followed by the quantitative data regarding its reliability, difficulty, and discrimination power, as well as several parameters derived from the confidence ratings, all of which are used to assess secondary school students' understanding of the abovementioned concepts.

Diagnostic Instruments for the Assessment of Conceptual Understanding

Conceptual understanding can be explored through the use of several diagnostic instruments. At the same time, it is important to acknowledge that most of these instruments possess limitations. For example, interviews and open-ended questions (OEQs) are commonly used to detect students' conceptual difficulties related to various science topics (Treagust, 1986). However, students often fail to provide detailed answers to OEQs (Gurel et al., 2015), while interviews are time-consuming and unsuitable for screening of large research samples (Chandrasegaran et al., 2007). Concept maps are invaluable for detecting gaps in the understanding of interrelated concepts (Ross & Munby, 1991), but teaching students how to compose them requires time, and the answers that they provide in this form are often incomplete (Kinchin, 2000).

Tests consisting of multiple-choice questions (MCQs) are markedly time-efficient, but their great weakness lies in the relatively high probability of producing the correct responses

by chance (Milenković et al., 2016). For this reason, Treagust (1986, 1988) proposed the use of two-tier tests (2TTs). Items in these tests consist of the answer tier (AT) and the reason tier (RT) in the MCQ format. Within the RT, the justification for the response to the AT is provided. A two-tier item is answered correctly only if responses to both tiers (BTs) are correct, which lowers the probability of guessing the correct answer (Milenković et al., 2016). However, 2TTs cannot distinguish incorrect responses due to ACs from those caused by the lack of knowledge, or estimate the strength that ACs are harbored with (Caleon & Subramaniam, 2010a), which led to the development of three- and four-tier tests. Three-tier test items consist of the AT, RT, and a confidence rating for the responses to these two tiers (Caleon & Subramaniam, 2010b). Thus, it is not possible to ascertain whether students possess different levels of confidence for their answers to the AT and RT, which might be the case given that these tiers measure different levels of knowledge and that, despite their relatedness, students sometimes perceive them as independent MCQs. Therefore, within four-tier tests, a separate confidence rating is added for each tier (Caleon & Subramaniam, 2010a). To distinguish between wrong answers caused by the lack of understanding and a lack of knowledge on these tests, Caleon and Subramaniam (2010a) suggested the following approach. Firstly, significant ACs represent all distracters or erroneous answer-reason combinations selected by at least 10% of the sample above the percentage of students that could have made this choice by chance. Next, confidence ratings of significant ACs are used as a measure of their strength. Since significant ACs often represent individual distracters from the AT and RT, separate confidence ratings for the two tiers provide the opportunity to accurately ascertain the strength of such ACs, which wouldn't have been possible with a three-tier test. Furthermore, overall confidence rating for an erroneous answer-reason combination on a three-tier item may not represent a completely accurate approximation of its strength, if students have different levels of confidence for their answers to the AT and RT. The strength of significant ACs of this type is more accurately determined if the confidence with which they are expressed is calculated from the individual confidence ratings for the answer and reason in question (Yan & Subramaniam, 2018). Subsequently, spurious ACs which are caused by the lack of knowledge represent significant ACs expressed with low confidence (mean confidence rating below 3.50 on a six-point confidence scale), while genuine ACs, caused by the lack of understanding, represent significant ACs expressed with high confidence (mean confidence rating above 3.50 on a six-point confidence scale). Genuine ACs can be further classified as moderate (mean confidence rating ranging from 3.50 to 4.00) and strong (mean confidence rating above 4.00). Finally, correct answers expressed with high confidence (mean confidence rating above 3.50) indicate good conceptual understanding, while correct answers accompanied by low confidence (mean confidence rating below 3.50) are indicative of a lack of knowledge. So far, four-tier tests have been used for the assessment of understanding of chemistry concepts among university (Habiddin & Page, 2019; Sreenivasulu & Subramaniam, 2013, 2014) and secondary school students (Yan & Subramaniam, 2018), but they have never been applied for this purpose in the field of biochemistry, at either of the two educational levels.

Uncovering Conceptual Challenges Related to Amino Acids, Proteins, and Enzymes

Up to date, only a small number of studies explored students' conceptual difficulties related to amino acids, proteins, and enzymes (Table 1).

As can be seen in Table 1, the previous studies only explored conceptual difficulties regarding amino acids, proteins, and enzymes among university students, so there is a complete lack of literature concerning the challenges with the understanding of the given content area among secondary school students. When it comes to amino acids, university students experienced the greatest difficulties with the understanding of acid–base properties of these compounds and interactions between their non-polar side chains. Regarding proteins, problems were uncovered in relation to the understanding of all four levels of protein structure while, in terms of enzymes, the students struggled with the understanding of enzyme kinetics, enzyme–substrate interactions, and different types of enzyme inhibition.

An insight into students' understanding of key biochemistry concepts can only be gained through appropriate diagnostic assessment. Therefore, the lack of literature on ACs related to particular biochemistry concepts which are elaborated within a certain course or educational level primarily originates from the lack of diagnostic instruments that specifically examine the understanding of these concepts, among the students of the given cohort (Villafañe et al., 2011). While a limited number of diagnostic tools for the assessment of university students' understanding of proteins are already in use, up to date, no instrument for the exploration of secondary school students' understanding of this topic has been developed. Development of such instrument is all the more important in light of the anecdotal evidence derived from interactions with secondary school students and teachers, as well as students who are about to embark in introductory biochemistry course at the university, which warns of the existence of numerous ACs related to the abovementioned content. Interviews with secondary school students and teachers conducted within this research, discussed in greater detail in the following section, confirmed the existence of such ACs regarding the xanthoproteic reaction, acid–base properties of amino-acids, stability of alpha helices, principles of sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE), effects of temperature on enzyme activity, competitive and noncompetitive inhibition, enzyme kinetics, and protein precipitation with ammonium sulfate. Table 1 further shows that prior studies used interviews, OEQs, or tests consisting of MCQs, and all of these instruments possess limitations that can be overcome through the implementation of four-tier tests. Furthermore, four-tier tests have previously never been used for the exploration of conceptual understanding of biochemistry at any educational level. Therefore, the principal objective of this study was to:

- Develop and validate a four-tier test for the assessment of secondary school students' conceptual understanding of amino acids, proteins, and enzymes.

The development and implementation of this instrument would enrich the literature on both the conceptual challenges related to amino acids, proteins, and enzymes and the use of four-tier tests for the assessment of understanding in the field of biochemistry at the secondary school level. Furthermore, secondary school teachers would be provided with an accurate and reliable diagnostic instrument for a time-efficient exploration of their students' understanding of the given content. The feedback from the test would enable teachers to develop effective strategies for remediation of the detected conceptual difficulties and preclusion of their future reoccurrence. Rectification of the uncovered ACs would also prevent their transfer to the university level and a prolonged negative impact on learning related to the entire field of life sciences.

Table 1 Conceptual challenges regarding amino acids, proteins and enzymes

Study	Educational level	Diagnostic instrument	Content area of difficulty	Detected ACs
Halmo et al. (2018)	University	Interview	Interactions between non-polar amino acid side chains	Non-polar amino acid side chains cannot interact with each other, because they have no charge Phenylalanine side chains can interact with each other through hydrogen bonding
Watters and Watters (2006)	University	Interview	Acid–base properties of amino acids	If pH value is higher than its pK_a value, the side chain of the given amino acid will be protonated
Harle and Towns (2013)	University	Interview	Primary protein structure	A bond between two amino acids represents an ester/glycosidic linkage
Harle and Towns (2013)	University	Interview	Secondary protein structure	Hydrogen bonds protrude out of the alpha helix
Villafañe et al. (2011)	University	Test consisting of MCQs	Tertiary and quaternary protein structure	The interior of an alpha helix is occupied by the side chains of amino acids that form the helix
Halim et al. (2018)	University	OEQs	Enzyme–substrate interactions	The quaternary protein structure consists of various secondary level structures (the tertiary structure is entirely overlooked)
Bretz and Linenberger (2012)	University	Test consisting of MCQs	Enzyme–substrate interactions	Substrate contains the active site Enzyme and substrate only interact at the active site Enzyme–substrate binding happens exclusively at transition state
Rodriguez et al. (2019)	University	Interview	Enzyme kinetics	The x-axis of the Michaelis–Menten graph represents time The y-axis of the Michaelis–Menten graph represents the velocity (V) instead of the initial velocity (V_0) of an enzymatic reaction
Rodriguez and Towns (2021)	University	OEQs	Enzyme inhibition	An enzymatic reaction within which the rate of formation of the product is constant over time represents a first-order reaction
Bretz and Linenberger (2012)	University	Test consisting of MCQs	Enzyme inhibition	Inhibitors are able to attach to the substrate Competitive inhibition is the only type of inhibition
Halim et al. (2018)	University	OEQs	Enzyme inhibition	Enzymes themselves function as inhibitors Non-competitive inhibition cannot be reversed Irreversible inhibitors exclusively target the allosteric site of the enzyme Reversible inhibitors exclusively target the active site of the enzyme Irreversible inhibitors bind non-covalently to the enzyme

Methodology

Development of 4AAPE

The three-phase process of the development of 4AAPE was conducted in Serbia, where amino acids, proteins, and enzymes are elaborated in the fourth (final) year of secondary school. Serbian chemistry curriculum for this educational level is primarily focused on biochemistry and biotechnology. Teaching topic *Proteins* represents one of the key topics encompassed by the curriculum (along with the topics devoted to carbohydrates, lipids, nucleic acids, principles of metabolism, and principles of biotechnology) with five to six weeks, including two lesson periods per week, allocated for its elaboration. Although no previous research examined secondary school students' understanding of this topic, anecdotal evidence, gathered through interactions with students and teachers, suggests that the former encounter considerable difficulties while learning about proteins. Author's and her colleagues' experience with students who recently embarked on introductory biochemistry course at the university indicates that a thorough examination of the knowledge brought from secondary school is warranted at the very beginning of the course, to prevent the various ACs related to acid–base properties of amino acids, all four levels of protein structure, and protein precipitation and denaturation, as well as enzyme kinetics, from interfering with further learning. Additional insight into students' conceptual difficulties regarding proteins is provided later in this section, when student and teacher interviews are discussed. Development of 4AAPE followed the procedure proposed by Treagust (1986, 1988) for 2TTs, with certain minor modifications.

Phase 1 of work, *defining the content area of the study*, included the following steps:

1. Identification of the key concepts related to amino acids, proteins, and enzymes, which are elaborated at the secondary school level. Upon a thorough examination of the chemistry curriculum and chemistry textbook for the fourth year of secondary school, the following concepts were selected and incorporated into the corresponding concept map (Fig. 1): chemical reactions and acid–base properties of amino acids, protein structure, and protein separation by SDS PAGE, enzyme kinetics, enzyme precipitation with ammonium sulfate, and factors affecting enzyme activity.
2. Generating propositional knowledge statements pertinent to the concepts under investigation and relating them directly to the concept map, thus defining the content area of the study and confirming its internal consistency.
3. Content validation of the concept map and propositional knowledge statements by one university and one secondary school chemistry educator. The validators confirmed that the selected content is scientifically correct and relevant to the knowledge about amino acids, proteins, and enzymes that should be acquired at the secondary school level.

Phase 2, *identifying students' conceptual challenges*, presupposed the following:

1. A literature review that, as already explained, only produced an overview of the university students' conceptual challenges regarding the content area of the study.
2. Interviews with two highly experienced secondary school teachers and six fourth-year secondary school students, in which the participants expressed their views on common difficulties related to the concepts under investigation. Except for the reactions of amino acids, the key difficulties identified by the two sides were relatively similar. Nevertheless, student interviews uncovered the existence of several additional ACs (Table 2).

3. Preparation of the preliminary version of the test, with items consisting of the AT in the MCQ format and the RT in the form of an OEQ, which were constructed to address the conceptual difficulties detected within the first three steps of phase 2. The preliminary version of the test was subsequently administered to 46 fourth-year secondary school students.

Phase 3, *development of the diagnostic instrument*, included the following:

1. Preparation of the pre-pilot version of the test, with the AT and RT in the MCQ format. The distracters for the RT of items in the pre-pilot version were crafted from the most frequent incorrect answers to the RT of the corresponding items in the preliminary version, as well as the information gathered through the interviews and the literature review.
2. Validation of the pre-pilot test by one university and one secondary school chemistry educator. The validators checked whether the items were written clearly, without any ambiguities and grammatical errors, whether there is only one correct answer to the AT and RT of each item, whether answer keys provided for the two tiers are correct and whether the proposed distracters are satisfactory and represent potential ACs. The validators suggested minor changes in the phrasing of three distracters in the RT, which were taken into account in the final version of the instrument. Aside from this, the validators' feedback was very positive, concluding that the instrument is appropriate for the assessment of secondary school students' conceptual understanding of amino acids, proteins, and enzymes.
3. Preparation of the pilot test, by adding six-point confidence scales ranging from *Just guessing* (1) to *Absolutely confident* (6), for the AT and RT of each item. Such scales were also used in other previously developed four-tier chemistry tests (Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018) and their implementation is justified for several reasons. Firstly, previous research established that rating scales should optimally have five or six response categories. Five or six categories are also most frequently used by subjects when expressing their confidence on the continuous scale (McKelvie, 1978). However, scales with five (or less) categories generally have lower discrimination power, as well as lower validity and reliability compared to the six-point scales (Preston & Colman, 2000), which therefore represent the most appropriate tool for the examination of students' confidence in their responses to the AT and RT of items in the test. Following the addition of confidence scales, the test which comprised ten four-tier items was administered to a new sample of 62 fourth-year secondary school students for pilot testing.
4. Prior to taking the pilot test, the students were informed about its novel format and explained how to use the confidence scales, following which they were given one chemistry lesson period lasting 30 min to complete it (ordinarily, lesson periods last 45 min, but due to the COVID-19 pandemic, their duration was shortened). Subsequently, the students confirmed that no clarifications in the phrasing of any of the items were needed and that they had no trouble with the use of the confidence ratings. However, the students warned that 30 min was barely enough time to complete the test and suggested omission of one of the items to overcome this problem. Twelve days after the pilot test, interviews were conducted with five students. Firstly, the students were requested to respond to the test items when the answer options for the AT and RT were covered. Afterwards, the answer options were revealed and the students were asked to select their responses in a think aloud manner, thus exposing the reasoning behind their choice of answers.

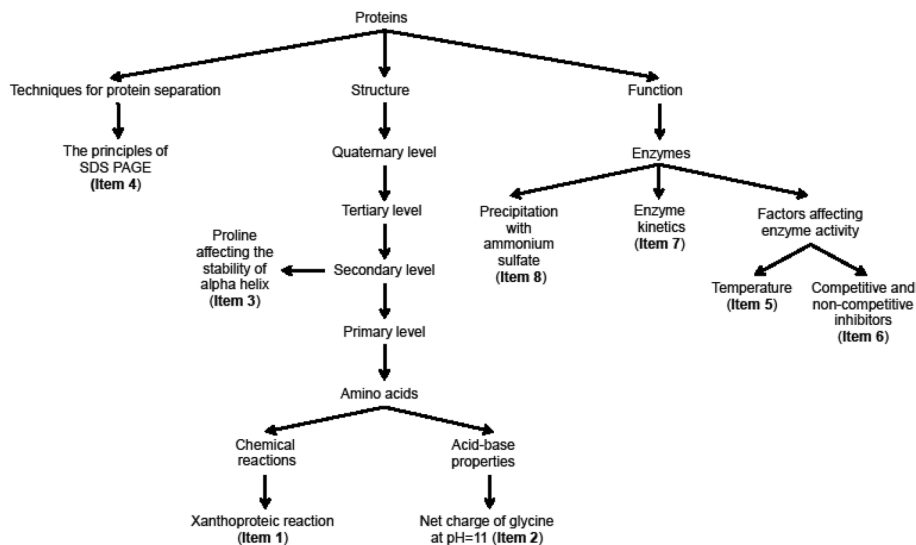


Fig. 1 The concept boundaries of the study

Occasionally, the students even attempted to explain the way in which other answer options could have been selected, noting that during certain phases of learning about proteins, they also reasoned in such a manner. Overall, the interviews confirmed good understanding of the phrasing and requirements of the questions in the two tiers of each item and satisfactory ability of the RT answer options to detect common ways of secondary school students' reasoning about concepts of interest. All interviewed students and around 68% of all students who took the pilot test had different confidence ratings for the AT and RT of at least six items, thus justifying the use of the four-tier format in the main study. Further analysis of the collected data showed that all ten items had satisfactory discrimination index (DI) values of above 0.2 (Mitra et al., 2009) for BTs. The facility indices (FIs) for BTs of eight items were within the acceptable range between 0.25 and 0.75 (Sreenivasulu & Subramaniam, 2013), but one item proved to be too difficult ($FI(BTs)=0.18$), while another one was too easy ($FI(BTs)=0.83$). Consequently, both items were omitted from the test and, thus, the final version of 4AAPE, comprising eight items, was prepared. The full contents of this test are presented in Online Resource 1, while a sample item is provided in Fig. 2.

Participants

The final version of 4AAPE was administered to 123 fourth-year students (aged 18–19 years) from three secondary schools in Serbia, who recently completed the five-week long elaboration of amino acids, proteins, and enzymes. None of the students or their teachers took part in the previous phases of this study. The implementation of 4AAPE was approved by the director of each participating school. All students voluntarily accepted to take the test, after they were explained that its purpose was to uncover the difficulties with the understanding of the abovementioned content, so that effective remediating strategies could be developed to overcome them. The students were assured that only the summarized results of the study would be published, that their

Table 2 Secondary school students’ common difficulties with the concepts encompassed by this study

Concept	Difficulties detected by the teachers	Additional ACs revealed by the students
Reactions of amino acids	No difficulties were detected	Side chains of amino acids that take part in the xanthoproteic reaction contain sulfur
Acid–base properties of amino acids	It is unclear how pK_a values of ionizable groups within an amino acid molecule influence their ionization state and the net charge of the entire molecule at the given pH	All amino acids exist in their zwitterion form at $pH = 7$ pI value represents the pH value at which ionizable groups in an amino acid molecule have no charge
Stability of alpha helices	Hydrogen bonds are weak and easily breakable, so it is unclear how they can be crucial for the stability of alpha helices It is unclear between which atoms are hydrogen bonds formed in alpha helices	Alpha helices are stabilized by disulfide bonds
SDS PAGE	Polypeptide chains are separated according to their charge Protein subunits are not separated in the gel	The largest polypeptides are the most negatively charged and travel the furthest in the gel
Effect of temperature on enzyme activity	Thermic inactivation of enzymes only occurs at temperatures over 100 °C	Shape of the graph that depicts the effect of temperature on the rate of enzymatic reaction is hyperbolic
Competitive and noncompetitive inhibition	Both types of inhibition are irreversible	Noncompetitive inhibitors cannot bind to the enzyme–substrate complex
Enzyme kinetics	The Michaelis–Menten graph depicts the change of velocity or concentration of the reaction product with time	Plateau region of the graph is reached when all the substrate in the reaction system is consumed
Protein precipitation with ammonium sulfate	Precipitation also causes protein denaturation	Both precipitation and subsequent denaturation are irreversible

performance on the test would only be known to the researcher, and that it would, thus, have no impact on their chemistry grade. The students were explained about the test's novel format and assigned 30 min to complete it.

Treatment of Data

The treatment of data followed the procedure devised by Caleon and Subramaniam (2010a). For each item in 4AAPE, the students' answers to the AT and RT were assigned the score of 1 if correct, or 0 when otherwise. Furthermore, the BTs score was assigned the value of 1 if both the AT and RT were answered correctly, or 0 in other instances. Based on these scores, the FI and DI values were calculated for the AT, RT, and BTs of all items in 4AAPE.

Next, using the students' confidence ratings, the following parameters were determined for the AT, RT, and BTs of each test item:

- The mean confidence (CF) represents the sum of all confidence ratings for the given tier(s), divided by the total number of students.
- The mean confidence of students when answering correctly (CFC) represents the sum of the confidence ratings for all the correct answers, divided by the total number of students who produced them.
- The mean confidence of students when answering wrongly (CFW) represents the sum of the confidence ratings for all the incorrect answers, divided by the total number of students who produced them.
- The confidence discrimination quotient (CDQ) shows whether the students can differentiate between what they know and what they do not know, and it is calculated as $(CFC - CFW)/\text{standard deviation of all confidence ratings for the given tier(s)}$.
- The confidence bias (CB) shows whether the students' confidence matches the accuracy of their responses, and it is calculated as $[(CF - 1)/5]$ – proportion of students who answered BTs of the given item correctly.

To examine the test–retest reliability, 4AAPE was re-administered to a sub-group of 55 students after three weeks, following which the corresponding Pearson's r values were calculated. The reliability of the test was further examined through the calculation of Cronbach's alpha and split-half reliability coefficients for both the cognitive scores and confidence ratings. Additionally, all the cognitive scores and confidence ratings were independently determined by the author and one recently retired secondary school chemistry teacher and the agreement between them was 100%.

Results and Discussion

The reliability statistics of 4AAPE are presented in Table 3. As can be seen, the Pearson's r values regarding both the cognitive scores and confidence ratings for the AT, RT, and BTs were above 0.70 and of statistical significance, indicating good test–retest reliability. Furthermore, Cronbach's alpha and split-half coefficient values, in terms of both the cognitive scores and confidence ratings, were higher than 0.70, proving 4AAPE's adequate internal consistency. The calculated Cronbach's alpha values for the AT, RT, and BTs, in respect to

The optimum temperature of enzyme X is 20 °C. When the temperature of the system within which the reaction catalyzed by enzyme X unfolds is raised from 15 °C to 20 °C, the rate of this enzymatic reaction increases three times. What will happen with the rate of the enzymatic reaction if the temperature of the system is raised to 37.2 °C?

- a) The rate of the enzymatic reaction will increase, compared to the rate at the temperature of 20 °C
- b) The rate of the enzymatic reaction will drop, compared to the rate at the temperature of 20 °C
- c) The rate of the enzymatic reaction will stay the same as it was at the temperature of 20 °C

The confidence rating for your answer:

1	2	3	4	5	6
Just guessing	Very unconfident	Unconfident	Confident	Very Confident	Absolutely confident

The reason why you chose this answer is:

- a) At the optimum temperature, the catalytic activity of an enzyme reaches its maximum, following which it remains unchanged if the temperature is further increased
- b) In accordance with the collision theory, the rate of all reactions increases as the temperature is increased
- c) Due to the thermic denaturation of enzyme X, its catalytic activity is decreased

The confidence rating for the reason you provided:

1	2	3	4	5	6
Just guessing	Very unconfident	Unconfident	Confident	Very confident	Absolutely confident

Fig. 2 Item 5 from 4AAPE

cognitive scores, were more satisfactory compared to those reported in the previous studies on four-tier chemistry tests (Habiddin & Page, 2019; Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018). At the same time, these values were lower than the corresponding values for the confidence ratings, thus complying with the trend that was already documented in the abovementioned prior studies. Overall, the results presented in Table 3 confirm that 4AAPE represents a reliable instrument for the assessment of secondary school students’ conceptual understanding of amino acids, proteins, and enzymes.

The values of the FIs and DIs for the AT, RT, and BTs of each item in 4AAPE are presented in Table 4. As can be observed, all FIs were in the acceptable range between 0.25 and 0.75, while all DIs were above the lowest acceptable value of 0.2. The mean DI values for the AT, RT, and BTs were around 0.70, indicating a good discriminatory power of the test. There was no significant difference between the students’ scores on the AT and RT ($t(122)=1.64, p=0.104$) and the mean FI values of 0.55 and 0.52, respectively, imply that the difficulty of these tiers was moderate. On the other hand, the

mean FI(BTs) value of 0.38, along with the mean score 3 out of 8, indicates that, overall, 4AAPE was difficult for the students. This conclusion is in alignment with the findings of the previous studies on four-tier chemistry tests (Habiddin & Page, 2019; Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018) and it is not unexpected, given that 4AAPE represents a diagnostic rather than an achievement test (Sreenivasulu & Subramaniam, 2014). Such findings further emphasize that a more accurate insight into the students' conceptual understanding can be gained if the responses to the AT are considered along with the justifications for their selection, provided in the RT.

Table 5 presents the values of the confidence parameters for each item in 4AAPE. The mean CF(BTs) value of 3.65 out of 6, which is similar to those reported in the previous studies on four-tier chemistry tests (Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018), shows that the students' overall confidence in their answers on 4AAPE was only 60.83%. Furthermore, it is important to acknowledge that the students' confidence was not adequately calibrated, given that the mean CB(BTs) value of +0.16 implies that the students were inclined to overestimate their performance.

Although their scores on the two tiers were similar, the students were significantly more confident in their answers to the AT, in comparison to the RT ($t(122)=4.32$, $p<0.0001$). This is an indication that, overall, the students did find answering to the RT to be more challenging compared to the AT, which is a tendency that was also observed in the previous studies on 2TTs (Treagust, 1986, 1988) and four-tier chemistry tests (Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018).

The CFC(BTs) values for seven out of eight items in 4AAPE were higher than the CFW(BTs) values, so the corresponding CDQ(BTs) values were above zero and the mean CDQ(BTs) value was +0.28. Such results are consistent with the finding that students tend to be more confident when they are right, than when they are wrong (Lundeberg et al., 2000). However, the CDQ(BTs) value for item 6 was just below zero indicating that, in this instance, the students were slightly more confident when producing the incorrect as opposed to the correct answer. Furthermore, the abovementioned mean CDQ (BTs) value shows that the students' ability to discern between what they do and don't know was relatively modest, which was also noted in the prior studies of this type (Sreenivasulu & Subramaniam, 2013, 2014; Yan & Subramaniam, 2018).

The mean CFC(BTs) value of 3.87, being above the mid-point of the confidence scale, indicates that it is generally unlikely that the students' correct answers were produced through guessing. However, it also shows that the students were reluctant to assign the highest confidence ratings to their correct responses. Furthermore, the fact that the mean CFW(BTs) value was also just above the 3.50 mark implies a high likelihood of the presence of ACs. Overall, all these findings confirm that the content regarding amino acids, proteins, and enzymes is conceptually challenging for secondary school students.

Conclusion

This study focused on the development of the four-tier test entitled 4AAPE, which enables the assessment of secondary school students' conceptual understanding of amino acids, proteins, and enzymes. In the light of the scantiness of diagnostic instruments for the exploration of understanding of fundamental biochemistry concepts (Bretz & Linenberger, 2012; Villafañe et al., 2021), the present study builds upon the prior research which established that four-tier tests overcome the limitations of all the previously developed

Table 3 Reliability statistics for 4AAPE

Reliability measure	AT		RT		BTs	
	Score	Confidence	Score	Confidence	Score	Confidence
Pearson’s <i>r</i>	0.74**	0.82**	0.72**	0.80**	0.74**	0.88**
Cronbach’s alpha	0.74	0.88	0.75	0.83	0.76	0.87
Split-half coefficient	0.76	0.89	0.79	0.84	0.75	0.88

** $p < 0.01$

instruments for the assessment of conceptual understanding (Caleon & Subramaniam, 2010a) and represent efficient tools for identification of conceptual difficulties in the field of chemistry (Sreenivasulu & Subramaniam, 2014).

Following the three-phase development process, the final version of 4AAPE, consisting of eight items, was prepared. Cronbach’s alpha and split-half coefficient values, regarding both the cognitive scores and confidence ratings on the test were higher than 0.70, proving 4AAPE’s adequate internal consistency. Furthermore, Pearson’s *r* values for both the cognitive scores and confidence ratings were above 0.70, confirming satisfactory test–retest reliability. The mean DI(BTs) value of 0.73 shows that 4AAPE has good discrimination power. The AT and RT were of moderate difficulty, but the mean FI(BTs) value of 0.38 indicates that, overall, 4AAPE was difficult for the students. The mean CF(BTs) value of 3.65 out of 6 implies that the students’ overall confidence in their answers on the test was not high, while the mean CDQ(BTs) value of +0.28 shows that their ability to distinguish between what they do and do not know was modest. This value further implies that the students were more confident when producing the correct compared to the wrong answers, but the mean CFC(BTs) value of 3.87 shows that they were nevertheless hesitant to assign high confidence ratings to these correct responses. Furthermore, the mean CFW(BTs) value of 3.52 is indicative of the presence of ACs in the students’ knowledge base. Overall, such findings confirm that the content about amino acids, proteins, and enzymes is riddled with conceptual challenges for secondary school students.

Table 4 The difficulty level and discrimination power of 4AAPE

Item	FI values			DI values		
	AT	RT	BTs	AT	RT	BTs
1	0.64	0.63	0.47	0.58	0.70	0.76
2	0.56	0.47	0.37	0.61	0.42	0.52
3	0.55	0.54	0.41	0.73	0.85	0.76
4	0.49	0.46	0.31	0.82	0.82	0.76
5	0.46	0.44	0.31	0.79	0.85	0.70
6	0.54	0.52	0.35	0.58	0.67	0.67
7	0.61	0.59	0.44	0.88	0.76	0.82
8	0.53	0.47	0.33	0.82	0.82	0.82
Mean	0.55 ^a	0.52 ^b	0.38 ^c	0.73	0.74	0.73
SD	0.06	0.07	0.06	0.12	0.14	0.10

^aMean score was 4.39, *SD* = 2.36

^bMean score was 4.12, *SD* = 2.40

^cMean score was 3, *SD* = 2.35

Table 5 The confidence parameters for each item in 4AAPE

Item	AT				RT				BTs						
	CF	CFC	CFW	CDQ	CB	CF	CFC	CFW	CDQ	CB	CF	CFC	CFW	CDQ	CB
1	3.59	3.75	3.30	0.37	0.05	3.48	3.69	3.13	0.47	0.03	3.53	3.86	3.24	0.53	0.04
2	3.77	3.91	3.59	0.21	0.18	3.63	4.00	3.31	0.47	0.16	3.70	3.94	3.56	0.26	0.17
3	3.58	3.63	3.51	0.10	0.11	3.52	3.59	3.44	0.13	0.09	3.55	3.58	3.53	0.04	0.10
4	3.88	4.08	3.68	0.32	0.27	3.80	4.02	3.62	0.34	0.25	3.84	4.22	3.67	0.47	0.26
5	3.75	3.95	3.58	0.28	0.24	3.76	4.04	3.55	0.40	0.24	3.76	4.03	3.63	0.33	0.24
6	3.73	3.57	3.93	-0.26	0.20	3.69	3.83	3.54	0.22	0.19	3.71	3.70	3.72	-0.01	0.19
7	3.54	3.68	3.33	0.24	0.07	3.37	3.38	3.36	0.01	0.03	3.46	3.62	3.33	0.21	0.05
8	3.72	3.92	3.47	0.36	0.21	3.53	3.83	3.26	0.47	0.18	3.63	3.98	3.45	0.44	0.20
Mean	3.70	3.81	3.55	0.20	0.17	3.60	3.80	3.40	0.31	0.15	3.65	3.87	3.52	0.28	0.16
SD	0.12	0.18	0.20	0.21	0.08	0.15	0.23	0.17	0.18	0.09	0.13	0.22	0.17	0.20	0.08

Consequently, 4AAPE can be used by secondary school teachers to examine their students' conceptual understanding of the abovementioned content. The feedback from the test would also enable teachers to evaluate the effectiveness of their instruction on amino acids, proteins, and enzymes, and provide them with useful guidance about the ways in which it can be improved to ensure the overcoming of the detected difficulties and preclusion of their reoccurrence. Confidence ratings for the responses to the AT and RT of items in 4AAPE enable teachers to distinguish between spurious ACs which arise from the lack of knowledge, from genuine ACs caused by the lack of understanding. Thus, they are able to prioritize the implementation of remediating strategies, as rectification of genuine ACs which are held with high confidence and are likely to be strongly embedded into students' cognitive structures may require considerable time and effort on their behalf. Dealing with spurious ACs is expected to be less demanding and these ACs can be more easily tackled by students on their own (Sreenivasulu & Subramaniam, 2014). Previous research on four-tier chemistry tests mostly uncovered genuine ACs of moderate strength, which should be rectified through *precise* instruction that specifically addresses each of these ACs (Caleon & Subramaniam, 2010a). Furthermore, by identifying the content regarding which genuine ACs tend to develop, when planning their elaboration of proteins with future generations of students, teachers would be able to prioritize their instruction and put more emphasis on this content, thus ensuring that the previously detected ACs do not reoccur. 4AAPE can also be used by university educators to ascertain what ACs related to amino acids, proteins, and enzymes novice students bring to their courses from secondary school. By promptly acting to remediate them, the educators would be able to prevent their negative impact on the learning of more advanced related concepts.

Authors of previous research on four-tier chemistry tests greatly relied on educators' experience to identify common conceptual difficulties, but opted not to conduct student interviews to obtain this information, while preparing the preliminary version of the test. This research, however, indicates that conducting student interviews in this phase of work is of great importance. Previously, considerable discrepancies were detected between science teachers' views of common conceptual difficulties and the actual ACs that the students harbor, as teachers were only able to identify 42.7% of their students' ACs (Sadler & Sonnert, 2016). Within the present research, the greatest discrepancies of this type referred to the reactions of amino acids, as teachers noted that none of them caused problems to the students, while student interviews revealed considerable difficulties with the xanthoproteic reaction. Although the views of the two sides on common difficulties with the other concepts under investigation were relatively similar, student interviews still uncovered several additional ACs related to them. Therefore, student interviews can be used to verify teachers' observations and obtain further information about the ACs regarding the concepts of interest, thus prompting the development of additional items that previously weren't even considered and contributing to the overall improvement of the test's scope and quality.

An important limitation of 4AAPE lies in the fact that, with only eight items, it does not represent a comprehensive diagnostic test about amino acids, proteins, and enzymes. Although the key concepts regarding this content which are elaborated at the secondary school level are covered by the test, numerous other items could have been derived from them, thus providing a fuller insight into the students' conceptual understanding. Therefore, within future studies, more extensive tests of this type could be developed. Nevertheless, implementation of 4AAPE provides first evidence that four-tier tests represent efficient and reliable tools for exploration of secondary school students' conceptual understanding of biochemistry. Application of 4AAPE will also provide the much needed initial overview of secondary

school students' ACs about proteins that can be used as a stepping stone for the development of more refined and more comprehensive four-tier tests for further and deeper examination of understanding of this topic. Additionally, although the number of items in 4AAPE is not large, they all have a relatively broad scope. Thus, some items examine the understanding of a few additional concepts along with the central concept of interest (e.g., along with the principles of SDS PAGE, item 4 also examines the understanding of protein denaturation and quaternary protein structure). Other items explore the understanding of principles at the core of a certain concept which are also important for understanding of other related concepts, thus signaling the existence of further ACs associated with them (e.g., if within item 3 the importance of hydrogen bonds for the stability of alpha helices is not acknowledged, it is likely that this problem will also arise regarding beta pleated sheets and tertiary and quaternary protein structure). On the other hand, if the concept of interest is complex, such as enzyme kinetics in item 7, the corresponding item considers several different aspects of it and reveals which of them cause the most problems to students. The test also encompasses items of different type (e.g., problem-based items, items that require identification of a correct statement from a given set of statements or have answer options in the form of chemical formulae), thus probing the students' understanding in a nuanced manner. All of the abovementioned characteristics of 4AAPE are expected to facilitate the acquisition of a relatively good initial overview of secondary school students' ACs about proteins. It is, however, important to acknowledge the possibility that some of the confidence ratings in this study may not represent an adequate reflection of the students' assuredness in their answers. Furthermore, since 4AAPE is a diagnostic test, students' performance on it should not be used to assess their achievement in the abovementioned content area. Finally, the present study focused on the development of 4AAPE, so although its results indicate the presence of considerable conceptual challenges related to amino acids, proteins, and enzymes, no information is provided about the actual ACs that secondary school students harbor in regard to this content. Once a detailed overview of these ACs is provided, future studies may attempt to uncover the reasons for their occurrence and develop effective strategies for their rectification and prevention.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11165-022-10075-5>.

Funding This study was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Contract number: 451–03-9/2021–14/200288.

Declarations

Ethics Approval and Consent to Participate Informed consent was obtained from all students who took part in this study. Parental consent regarding the students' participation was not required, given that all students were of age (minimum 18 years old). The study was approved by the directors of all participating schools.

Competing Interests The author declares no competing interests.

References

- Bretz, S. L., & Linenberger, K. J. (2012). Development of the enzyme-substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, 40(4), 229–233. <https://doi.org/10.1002/bmb.20622>
- Caleon, I., & Subramaniam, R. (2010). Do students know what they know and what they do not know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313–337. <https://doi.org/10.1007/s11165-009-9122-4>

- Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, 32(7), 939–961. <https://doi.org/10.1080/09500690902890130>
- Chandrasegaran, A. L., Treagust, D. F., & Mocerino, M. (2007). The development of a two-tier multiple-choice diagnostic instrument for evaluating secondary school students' ability to describe and explain chemical reactions using multiple levels of representation. *Chemistry Education Research and Practice*, 8(3), 293–307. <https://doi.org/10.1039/B7RP90006F>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Habiddin, H., & Page, E. M. (2019). Development and validation of a four-tier diagnostic instrument for chemical kinetics (FTDICK). *Indonesian Journal of Chemistry*, 19(3), 720–736. <https://doi.org/10.22146/ijc.39218>
- Halim, A. S., Finkenaedt-Quinn, S. A., Olsen, L. J., Gere, A. R., & Shultz, V. G. (2018). Identifying and remediate student misconceptions in introductory biology via writing-to-learn assignments and peer review. *CBE—Life Sciences Education*, 17(28), 1–12. <https://doi.org/10.1187/cbe.17-10-0212>
- Halmo, S. M., Sensibaugh, C. A., Bhatia, K. S., Howell, A., Ferryanto, E. P., Choe, B., Kehoe, K., Watson, M., & Lemons, P. P. (2018). Student difficulties during structure-function problem solving. *Biochemistry and Molecular Biology Education*, 46(5), 453–463. <https://doi.org/10.1002/bmb.21166>
- 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. <https://doi.org/10.1002/bmb.20719>
- Kinchin, I. M. (2000). Using concept maps to reveal understanding: A two-tier analysis. *School Science Review*, 81(296), 41–46.
- Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, C. (2000). Cultural influences on confidence: country and gender. *Journal of Educational Psychology*, 92(1), 152–159. <https://doi.org/10.1037/0022-0663.92.1.152>
- McKelvie, S. J. (1978). Graphic rating scales—How many categories? *British Journal of Psychology*, 69(2), 185–202. <https://doi.org/10.1111/j.2044-8295.1978.tb01647.x>
- Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Development of a three-tier test as a valid diagnostic tool for identification of misconceptions related to carbohydrates. *Journal of Chemical Education*, 93(9), 1514–1520. <https://doi.org/10.1021/acs.jchemed.6b00261>
- Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., & Judson, J. P. (2009). The levels of difficulty and discrimination indices in type A multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *International e-Journal of Science Medicine and Education*, 3(1), 2–7.
- Powell, K. C., & Kalina, C. J. (2009). Cognitive and social constructivism: Developing tools for an effective classroom. *Education*, 130(2), 241–250.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Rodriguez, J.-M.G., & Towns, M. H. (2021). Analysis of biochemistry students' graphical reasoning using misconceptions constructivism and fine-grained constructivism: Why assumptions about the nature and structure of knowledge matter for research and teaching. *Chemistry Education Research and Practice*, 22(4), 1020–1034. <https://doi.org/10.1039/D1RP00041A>
- Rodriguez, J.-M.G., Hux, N. P., Phillips, S. J., & Towns, M. H. (2019). Michaelis-Menten graphs, Lineweaver-Burk plots, and reaction schemes: Investigating introductory biochemistry students' conceptions of representations in enzyme kinetics. *Journal of Chemical Education*, 96(9), 1833–1845. <https://doi.org/10.1021/acs.jchemed.9b00396>
- Ross, B., & Munby, H. (1991). Concept mapping and misconceptions: A study of high school students' understandings of acids and bases. *International Journal of Science Education*, 13(1), 11–23. <https://doi.org/10.1080/0950069910130102>
- Sadler, P. M., & Sonner, G. (2016). Understanding Misconceptions: Teaching and learning in middle school physical science. *American Educator*, 40(1), 26–32.
- Sreenivasulu, B., & Subramaniam, R. (2013). University students' understanding of chemical thermodynamics. *International Journal of Science Education*, 35(4), 601–635. <https://doi.org/10.1080/09500693.2012.683460>
- Sreenivasulu, B., & Subramaniam, R. (2014). Exploring undergraduates' understanding of transition metals chemistry with the use of cognitive and confidence measures. *Research in Science Education*, 44(6), 801–828. <https://doi.org/10.1007/s11165-014-9400-7>

- Tansey, J. T., Teaster, B., Jr., Cox, M. M., Fox, K. M., Knight, J., Sears, D., & Bell, E. (2013). Foundational concepts and underlying theories for majors in biochemistry and molecular biology. *Biochemistry and Molecular Biology Education*, 41(5), 289–296. <https://doi.org/10.1002/bmb.20727>
- Treagust, D. F. (1986). Evaluating students' misconceptions by means of diagnostic multiple-choice items. *Research in Science Education*, 16(1), 199–207. <https://doi.org/10.1007/BF02356835>
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169. <https://doi.org/10.1080/0950069880100204>
- 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), 210–218. <https://doi.org/10.1039/C1RP90026A>
- Villafañe, S. M., Minderhout, V., Heyen, B. J., Lewis, J. E., Manley, A., Murray, T. A., Tienson-Tseng, H., & Loertscher, J. (2021). Design and implementation of a tool to assess students' understanding of metabolic pathways dynamics and regulation. *CBE-Life Sciences Education*, 20(3), 20–35. <https://doi.org/10.1187/cbe.20-04-0078>
- Wandersee, J. H., Mintzes, J. J., & Novak, J. D. (1994). Research on alternative conceptions in science. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 177–210). Macmillan.
- Watters, D. J., & Watters, J. J. (2006). Student understanding of pH: "I don't know what the log actually is, I only know where the button is on my calculator." *Biochemistry and Molecular Biology Education*, 34(4), 278–284. <https://doi.org/10.1002/bmb.2006.494034042628>
- Yan, Y. K., & Subramaniam, R. (2018). Using a multi-tier diagnostic test to explore the nature of students' alternative conceptions on reaction kinetics. *Chemistry Education Research and Practice*, 19(1), 213–226. <https://doi.org/10.1039/c7rp00143f>
- Yong, C. L., & Kee, C. Z. (2017). Utilizing concept cartoons to diagnose and remediate misconceptions related to photosynthesis among primary school students. In M. Karpudewan, A. N.Md. Zain, & A. L. Chandrasegaran (Eds.), *Overcoming students' misconceptions in science* (pp. 9–27). Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.