



THE DEVELOPMENT AND VALIDATION OF CRITICAL AND CREATIVE THINKING SKILLS TEST IN ENZYME FOR UNDERGRADUATE CHEMISTRY STUDENTS

Andi Wahyudi^{1✉}, S. Liliarsari², F.M.T. Supriyanti, Nahadi

¹Student of Doctoral Program of Science Education, Universitas Pendidikan Indonesia

²Department of Chemistry Education, Faculty of Mathematics and Natural Science Education, Universitas Pendidikan Indonesia

Article Info

Received January 2019

Accepted February 2019

Published February 2019

Keywords:

Creative thinking skills, critical thinking skills, test development and validation.

Abstract

The aim of this study was to develop and validate a test of students' critical and creative thinking skills on enzyme. This study was conducted using development and validation methods through six steps, i.e. (1) defining the construct and formulating objectives, (2) formatting items, (3) constructing items, (4) creating a scoring guide, (5) Judging items by experts, and (6) calculating validity, reliability, level of difficulty and discrimination by empirical study. This test was constructed two-tier test, which consists of the first tier as a multiple choice and the second tier as the reason from the first tier option. The participants were 61 undergraduate chemistry students who studied enzyme in a chemistry department. The results showed that item CVI value was 0.86 and item Cronbachs' Alpha value was 0.843 (very high category). Item difficulty index (P) was 0.501, and Item discrimination index (D) was 0.328 which showed that the level of difficulty item in the medium category and level of discrimination in the good category. Out of initial 14 items, 3 items were excluded after the analysis of CVI, item difficulty index and item discrimination index. The revised 11-item test was found to have sufficient validity, reliability, item difficulty index and item discrimination index to measurement of enzyme concept mastery, critical thinking skills, and creative thinking skills.

© 2019 Universitas Negeri Semarang
p-ISSN 2252-6617
e-ISSN 2252-6232

✉Corresponding author:

Andi Wahyudi

Student of Doctoral Program of Science Education

Universitas Pendidikan Indonesia

Email address: andywhayudy23@gmail.com

INTRODUCTION

Currently, we are faced fourth industrial revolution (4IR) that emphasizes internet-based digitalization. The Internet of Things (IoT) is a globalized name as a pioneer of the fourth industrial revolution (Lee *et al.*, 2018). It has an impact on the needs of skilled workers who must have higher order thinking skills (HOTS). Higher education as a provider of skilled workers must adjust its curriculum by social needs, no exception in the department of chemistry.

According to 21st-century skills, chemistry graduates should be having four basic skills (i.e. critical thinking, creativity, communication, and collaboration) (Goldberg, 2012). Critical and creative thinking skills were components of higher order thinking skills (HOTS). HOTS is an important skill to solve problems, make decisions, learn new things to prepare quality human resources. All experts agreed that critical and creative thinking skills are important dimensions in science education (Stephenson & Sadler-Mcknight, 2016; Yoon *et al.*, 2014; Zhou *et al.*, 2013). Critical and creative thinking in science education have a supplementary role about scientific process. Critical thinking skills are the ability to analyzing and evaluating information through observation, experience, communication or reading systematically (Ku *et al.*, 2014). This definition reflects a convergent view of critical thinking as logical reasoning and argumentation. Creative thinking skills are the ability to create original ideas that contribute to scientific information, present experimental results to understand science product, develop design for scientific activities and determine extraordinary plans (Kutlu & Gökdere, 2015). This ability can be grown by developing curiosity and imagination through learning activities. Creative thinking is innovation and a key factor for developing personal entrepreneurship and social competence. Creative person will produce various ideas from one problem and be able to produce something new.

Critical and creative thinking skills can be developed through continuous learning activities in each subject. One of the topics discussed in the biochemistry course is about enzymes. Enzyme is biochemistry core because it is important for

understanding other topics such as metabolic pathways, transcription, and translation (Bretz & Linenberger, 2012). The enzyme is also a complicated topic because one of the subtopics of an enzyme that is enzyme kinetics requires calculus knowledge apart from introductory biology and chemistry. Enzyme kinetics is commonly taught in introductory biochemistry course by delving directly into the mathematical equations that define Michaelis–Menten kinetics (Florjanczyk *et al.* 2018; Hinckley, 2012; House *et al.* 2016).

The effectiveness of a learning process can be measured from the learning outcome. It causes that the instrument is an important part to see the effectiveness of a learning process. Critical and creative thinking skills test in enzyme is an instrument to measure critical and creative thinking skills developed through biochemistry course on the topic of enzymes. The focus of this test not only the measure of students' abilities in critical thinking skills and creative thinking skills but also enzyme material and enzyme students' understanding.

Appropriate instruments are instruments that passed the steps of test development. In the development test, the instrument must be valid and reliable. Validity is level of test ability to gather information about the quality of the instrument intended to be assessed (Adams & Wieman, 2011). Reliability is the degree of consistency of the instrument with the same measurement of repeated attributes which will give identical conditions or similar conditions. Reliability can be measured by test-retest reliability, two semi-test reliability, Kuder Richardson-20 and Cronbachs' alphas reliability (Kara, 2015).

Previous studies have developed critical thinking skills test and creative thinking skills test, such as Watson-Glaser critical thinking skills test, California Critical Thinking Skills Test (CCTST), Torrance Test of Creative Thinking, and YanPiaw Critical and Creative Thinking Test. The tests in previous studies were standard tests. These will be bias when we want to see an increase in critical and creative thinking skills through the effects of learning methods. The increase caused may be from other factors, such as family, social environment and cultural

environment. The aim of this study was to develop a test that can measure the critical and creative thinking skills integrated with enzymes concept mastery.

METHODS

This study used methods of development and validation through six steps of developmental research by (Tiruneh, et al., (2017) (Fig.1). The first step was defining and selecting the indicator of critical thinking and creative thinking that should be targeted in this test. Based on 12 indicators of critical thinking skills by Ennis, the most suitable indicators were chosen for the topic of the enzyme, including: (1) inducing and judging induction, (2) deducing and judging deduction, (3) observing and judging observation reports, and (4) judging the credibility of a source. Indicators of creative thinking skills were adapted from Torrance and were selected 3 indicators of creative thinking skills including fluency, elaboration, and originality. Meanwhile, mastery of concept adapted by blooms' revised taxonomy which consist of C3 and C4 level. The second step was plotting item critical thinking, creative thinking and mastery concepts. In this section, every item that is suitable for development is examined along with the problem item. The third step was constructing item format with two-tier test, which consists of the first tier as a multiple choice and the second tier as the reason from the first tier option. The fourth step was creating a scoring guide which in each first tier item was given a score of 2 and the second tier was given a score of 3. The fifth step was expert validation from five lectures who are expert in assessment, HOTS, and biochemistry. Expert validation analysis used content validity ratio (CVR) by (Wilson et al. (2012). The sixth step was empirical study of 61 undergraduate chemistry students who studied enzyme at one of the universities in Bandung. This was used to see construct validity, reliability, level of difficulty and discrimination. The result of empirical study was analyzed by IBM SPSS series 20.

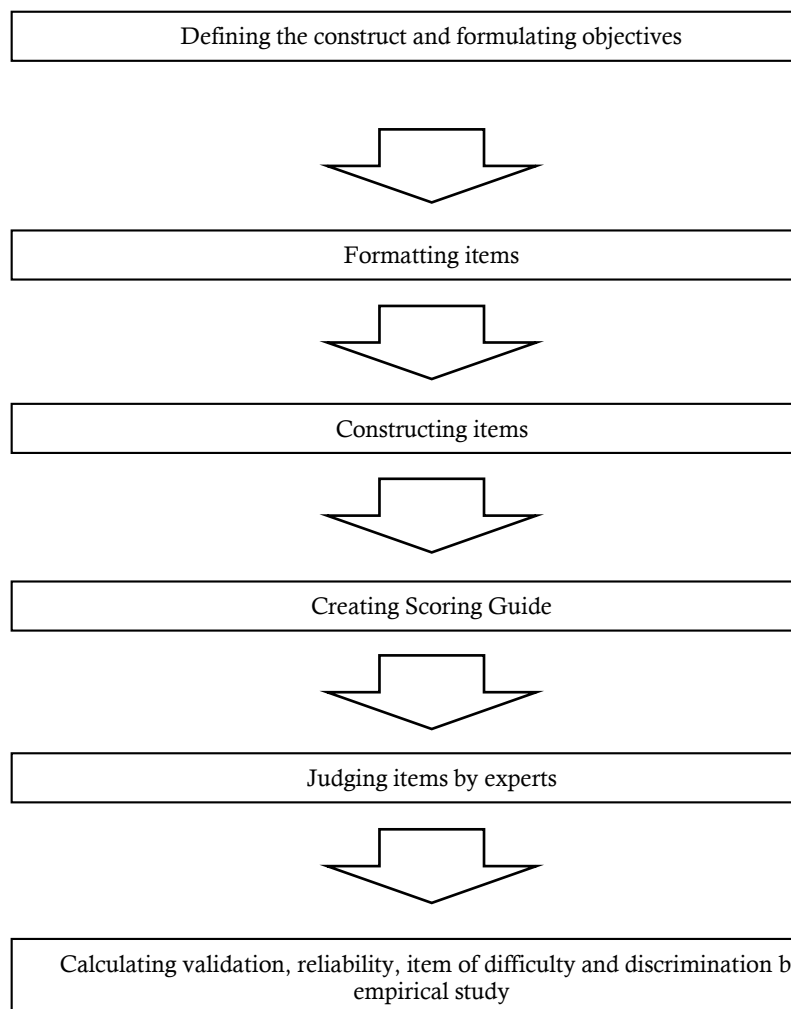


Figure 1. Schema of Instrument development

RESULTS AND DISCUSSION

Content validity

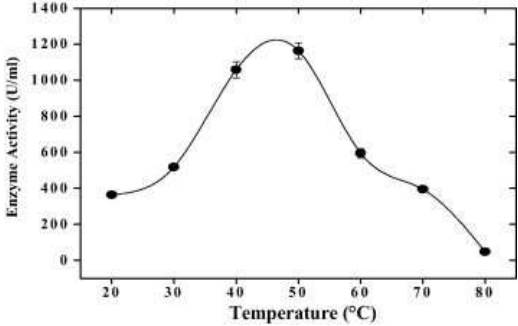
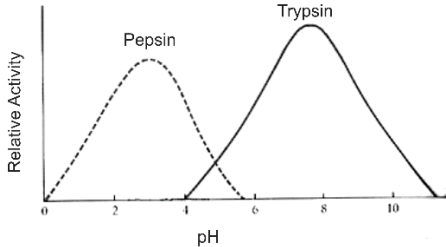
Content validity is the most crucial step in developing an instrument because at this step analysis of the suitability between items with the characteristics that are wanted to be measured (Xie, 2018; Yasar & Gundogan, 2014). This content validation is evaluated by qualified experts in the assessment domain, HOTS, and biochemistry. Method of content validity analysis used CVR which was developed by Wilson et al., (2012). Total of 14 items in the form of two-tier tests were analyzed by 5 experts. They make a score of 1 if the item is relevant to the standard and gives a score of 0 if the item is not relevant to the standard. The results of content validity analysis on 14 items can be seen in Table 1.

Table 1 showed the results of content validity analysis using CVR. According to Wilson, the critical value of CVR for the five panelists is 0.877 at the significance level (α) of 0.05. From the 14 items were analyzed, 10 items are in the appropriate interpretation (CVR = 1), 3 items must be revised (CVR = 0.6), and 1 item must be eliminated (CVR = 0.2). Content validity index (CVI) for each form of this test is 0.86. Items 4, 7, and 11 were revised in advance to be consulted with the panelists who considered that the items were not relevant. The items were repaired and judged by the five experts until the panelists agree to the items. The sample revised items can be seen in Table 2.

Table. 1. Analysis of Content Validity

Items	Relevant	Not relevant	CVR	Interpretation
1	5	0	1.00	Appropriate
2	5	0	1.00	Appropriate
3	5	0	1.00	Appropriate
4	4	1	0.60	Need Revision
5	5	0	1.00	Appropriate
6	5	0	1.00	Appropriate
7	4	1	0.60	Need Revision
8	5	0	1.00	Appropriate
9	5	0	1.00	Appropriate
10	3	2	0.20	Eliminated
11	4	1	0.60	Need Revision
12	5	0	1.00	Appropriate
13	5	0	1.00	Appropriate
14	5	0	1.00	Appropriate

Table 2. Description of Sample Items Revision

Item Before Revision	Item After Revision
<p>Item Number 11 The following graph shows the effect of temperature on enzyme activity (%)</p>  <p>What is optimum temperature for graph above....</p> <ol style="list-style-type: none"> 25°C 35°C 40°C 50°C 60°C 	<p>Item Number 11 The following graph shows the effect of temperature on catalase activity</p>  <p>Which statement is TRUE for the graph above....</p> <ol style="list-style-type: none"> Pepsin and trypsin are active in acidic conditions Pepsin and trypsin are active in alkaline conditions Pepsin is active in acidic conditions and trypsin is active in an alkaline condition Pepsin is active in alkaline conditions and trypsin is active in an acidic condition The difference between pepsin and trypsin is caused by inhibitors

In item 11, the panellist commented that this item did not show the cognitive level of analysis (C4). This item develops critical thinking skills on observing indicators and judging observation reports. Panelists also considered that the item should show activity of obtaining information from a source through the use of senses. This can be achieved with items arranged by presenting experimental data. The

item was changed by observing the graph of the relative activity of enzymes at different pH conditions. Students are asked to choose the correct conclusions from the graph with the reasons. The reason for the influence of pH on enzyme activity is due to the structure of enzyme ions which can be in the form of positive ions, negative ions or double charged ions (zwitter ions). Therefore, changes in environmental pH

will affect the effectiveness of the active part of the enzyme in forming enzyme-substrate complex (Daniel & Danson, 2013). Enzymes like proteins are made up of many amino acids through peptide bonds. Amino acids are composed of carboxyl groups (-COOH) and amine groups (-NH₂) which are bound to the same C atom (also called alpha carbon) (Chikezie, 2014; Demir *et al.*, 2012). The carboxyl group provides acidic properties and the amine group provides basic properties. In the form of a solution, amino acids are amphoteric (they tend to become acidic in alkaline solutions and become alkaline in acidic solutions).

Construct Validity and Reliability

Construct validity and reliability were obtained from the results of empirical research. These items have been judged and corrected based on expert advice. From the 14 items prepared, 1 item was eliminated because it had a low CVR. 13 items were analyzed by calculating the Pearson Product Moment (ppm) correlation to determine the construct validity (Nahadi *et al.* 2018). It saw a correlation between each item with a total score. Meanwhile, reliability was analyzed by calculating the correlation between different items on the same test (internal consistency) (Bajpai & Bajpai, 2014; Hermosilla & Alvarado, 2016). The results of the validity and reliability analysis through SPSS can be seen in Table 3.

Based on Table 3, there are 4 items in the high category, 6 items in the enough category, 1 item in the low category and 2 items in the very low category. Items that have a very low category in item number 4 and 7 were eliminated because they are not significant at the significance level (α) 0.05 (2 tailed). In item 13, although it had a low validity but can still be used because it is still significant at the significance level (α) of 0.05 (2 tailed) (Yu *et al.*, 2017). Analysis of reliability values using internal consistency has obtained a value of 0.843 with a very high category which states that this instrument has excellent consistency value in measuring critical and creative thinking skills.

Table 3. Recapitulation Validity and Reliability of Items

Item	Validity	Reliability
------	----------	-------------

	Value	Category	Value	Category
1	0.639*	Enough		
2	0.563*	Enough		
3	0.553*	Enough		
4	0.120	Very Low		
5	0.661*	High		
6	0.620*	High		
7	0.104	Very Low	0.843	Very High
8	0.430*	Enough		
9	0.598*	Enough		
10	0.568*	Enough		
11	0.796*	High		
12	0.779*	High		
13	0.334*	Low		

Note: *P<0.05 (sig. 2 tailed).

Level of difficulty and discrimination

Level of difficulty is the learners' ability to answer an item correctly and the ranges of 0.0 to 1.0 (Foertsch, 2014; Shin *et al.*, 2015). If the value approaches 0 then it can be said that the item is very easy, whereas if the item is close to 1 then the item is very difficult. Level of discrimination is the ability of an item to distinguish between high and low ability learners (Sünbül & Yormaz, 2018). The value of the level of discrimination ranges from -1 to 1. Score 1 shows that these items have an excellent level of discrimination. The results of the analysis in 13 items are shown in Table 4.

In Table 4, it is shown that the level of difficulty items in the range of 0.119 to 0.919. 5 items i.e. item 3, 8, 10, 11, and 12 were in the easy category, 4 questions in item 1, 2, 5, and 13 were in the medium category, and 4 questions in numbers 4, 6, 7, and 9 were in the difficult category. The average difficulty index is 0.501, it indicated that items were in the moderate difficulty level. Other studies that developed an instrument showed that the average level of difficulty in the moderate category (moderate difficulty) (Boopathiraj & Chellamani, 2013; Koçdar, 2016; Shin *et al.*, 2015). However, the overall test must consist of items that are easy, medium and difficult category. Medium category must have a larger percentage than easy and difficult category (Odukoya *et al.*, 2018; Sabri, 2013).

Table 4. Recapitulation Level of Difficulty and Discrimination

Item	Difficulty Index (P)		Discrimination Index (D)	
	Value	Category	Value	Category
1	0.381	Medium	0.299	Mediocre
2	0.445	Medium	0.497	Excellent
3	0.787	Easy	0.511	Excellent
4	0.119	Difficult	0.204	Mediocre
5	0.394	Medium	0.483	Excellent
6	0.239	Difficult	0.314	Good
7	0.068	Difficult	0.073	Poor
8	0.816	Easy	0.48	Excellent
9	0.255	Difficult	0.181	Poor
10	0.919	Easy	0.333	Good
11	0.839	Easy	0.070	Poor
12	0.703	Easy	0.286	Mediocre
13	0.548	Medium	0.532	Excellent
Average	0.501	Medium	0.328	Good

The level of discrimination from 13 items found that 3 items (7, 9, 11) in the poor category, 3 items (1,4,12) in the mediocre category, 2 items (6,10) in the good category, and 5 items (2,3,5,8,13) in the very good category. The average of discriminant level was 0.328 with the good category. This value states that the items can discriminate between the high group and low groups learners. The research conducted by Nahadi et al. (2018) on the development of virtual chemistry test based multiple representations found that the discrimination index was in the range of 0.44 to 0.67. Yu et al. (2017) calculated the discrimination index on the development of mechanical critical thinking scale and obtained a discriminant index value of 0.45.

Item 4 has a CVR value of 0.60, construct validity value of 0.120 (very low), difficulty index of 0.119 (difficult) and discrimination index of 0.204 (mediocre). Therefore, this item cannot be used and must be eliminated. Item 7 must also be eliminated because it has a CVR of 0.60, construct validity of 0.104 (very low), difficulty index of 0.068 (difficult) and

discrimination index of 0.073 (poor). Overall, there are 11 items left that have been successfully developed. These eliminated items do not reduce the number of indicators developed because they are still represented by other items.

CONCLUSION

In this study, enzyme-critical and creative thinking skills test was developed through six steps, i.e. (1) defining the construct and formulating objectives, (2) formatting items, (3) constructing items, (4) creating scoring guide, (5) judging items by experts, and (6) calculating validity, reliability, level of difficulty and discrimination. This test has a CVI of 0.86, construct validity showed a range of 0.104 to 0.796, reliability of 0.843, an average of difficulty index (P) of 0.501 and an average of discrimination index (D) of 0.328. From the 14 questions was developed, 1 item was eliminated because the item has a low CVR. Then, 13 items were analyzed through empirical study of 61 undergraduate chemistry students. Item 4 and item 7 were eliminated because the items have low values of construct validity, difficulty index, and discrimination index.

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.
- Bajpai, R., & Bajpai, S. (2014). Goodness of Measurement: Reliability and Validity. *International Journal of Medical Science and Public Health*, 3(2), 112-115.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of Test Items on Difficulty Level and Discrimination Index in the Test for Research in Education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189–193.

- Bretz, S. L., & Linenberger, K. J. (2012). Development of the enzyme-substrate interactions concept inventory. *Biochemistry and Molecular Biology Education*, 40(4), 229–233.
- Chikezie, P. C. (2014). Kinetic Studies of Polyphenol Oxidase from White Yam (*Dioscorea rotundata* Poir) Tuber. *Journal of Nutrition & Food Sciences*, 05(02), 355–363.
- Daniel, R. M., & Danson, M. J. (2013). Temperature and the catalytic activity of enzymes: A fresh understanding. *FEBS Letters*, 587(17), 2738–2743.
- Demir, H., Çilem, Ç., & Çağlar, F. (2012). Purification and characterization of polyphenol oxidase enzyme from Iğdır apricot (*Prunus armeniaca* L.). *Bitlis Eren University Journal of Science and Technology*, 2(1), 22–26.
- Florjanczyk, U., Ng, D. P., Andreopoulos, S., & Jenkinson, J. (2018). Developing a three-dimensional animation for deeper molecular understanding of michaelis-menten enzyme kinetics. *Biochemistry and Molecular Biology Education*, 46(5), 561–565.
- Foertsch, M. (2014). Five steps to effective test development. *Revista Hospital Universitário Pedro Ernesto*, 13(4), 79–83.
- Hinckley, G. (2012). A method for teaching enzyme kinetics to nonscience majors. *Journal of Chemical Education*, 89(9), 1213–1214.
- House, C., Meades, G., & Linenberger, K. J. (2016). Approaching a Conceptual Understanding of Enzyme Kinetics and Inhibition: Development of an Active Learning Inquiry Activity for Prehealth and Nonscience Majors. *Journal of Chemical Education*.
- Kara, F. (2015). Development of Achievement Test: Validity and Reliability Study for Achievement Test on Matter Changing, 6(24), 21–27.
- Koçdar, S., & Author, M. (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context, 15(4), 16–24.
- Ku, K. Y. L., Ho, I. T., Hau, K. T., & Lai, E. C. M. (2014). Integrating direct and inquiry-based instruction in the teaching of critical thinking: an intervention study. *Instructional Science*, 42(2), 251–269.
- Kutlu, N., & Gökdere, M. (2015). The effect of purdue model based science teaching on creative thinking. *International Journal of Education and Research*, 3(3), 589–599.
- Lee, M., Yun, J., Pyka, A., Won, D., Kodama, F., Schiuma, G., ... Zhao, X. (2018). How to Respond to the Fourth Industrial Revolution, or the Second Information Technology Revolution? Dynamic New Combinations between Technology, Market, and Society through Open Innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, 4(3), 21.
- Levin-goldberg, J. (2012). Teaching Generation TechX with the 4Cs: Using Technology to Integrate 21st Century Skills, 1, 59–66.
- Nahadi, N., Firman, H., & Kurniadi, H. (2018). Development and Validation of Chemistry Virtual Test Based Multiple Representations. *Journal of Education and Learning (EduLearn)*, 12(1), 44.
- Odukoya, J. A., Adekeye, O., Igbino, A. O., & Afolabi, A. (2018). Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. *Quality and Quantity*, 52(3), 983–997.

- Sabri, S. (2013). Item Analysis of Student Comprehensive Test for Research in Teaching Beginner String Ensemble Using Model Based Teaching Among Music Students in Public Universities. *International Journal of Education and Research*, 1(12), 1–14.
- Shin, S., Jung, D., & Kim, S. (2015). Validation of a clinical critical thinking skills test in nursing. *Journal of Education Evaluation for Health Professions*, 12(1), 1–7.
- Stephenson, N. S., & Sadler-Mcknight, N. P. (2016). Developing critical thinking skills using the Science Writing Heuristic in the chemistry laboratory. *Chem. Educ. Res. Pract*, 17, 72–79. h
- Sünbül, Ö., & Yormaz, S. (2018). Effects of Test Level Discrimination and Difficulty on Answer-Copying Indices. *International Journal of Evaluation and Research in Education (IJERE)*, 7(1), 32. 8
- Tiruneh, D. T., De Cock, M., Weldelessie, A. G., Elen, J., & Janssen, R. (2017). Measuring Critical Thinking in Physics: Development and Validation of a Critical Thinking Test in Electricity and Magnetism. *International Journal of Science and Mathematics Education*, 15(4), 663–682.
- Trizano-hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach ' s Alpha Reliability in Realistic Conditions : Congeneric and Asymmetrical Measurements, 7(May), 1–8.
- Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197–210.
- Xie, R. (2018). Investigating the Content Validity of the Junior Middle School Entrance English Testing. *Journal of Education and Development*, 2(1), 45.
- Yasar, S., & Gundogan, A. (2014). Determining Validity and Reliability of Data Gathering Instruments Used By Program Evaluation Studies in Turkey. *Procedia - Social and Behavioral Sciences*, 131, 504–509.
- Yoon, H., Woo, A. J., Treagust, D., & Chandrasegaran, A. L. (2014). The Efficacy of Problem-based Learning in an Analytical Laboratory Course for Pre-service Chemistry Teachers. *International Journal of Science Education*, 36(1), 79–102.
- Yu, K. C., Lin, K. Y., & Chang, S. F. (2017). The development and validation of a mechanical critical thinking scale for high school students. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(5), 1361–1376
- Zhou, Q., Huang, Q., & Tian, H. (2013). Developing Students ' Critical Thinking Skills by Task-Based Learning in Chemistry Experiment Teaching. *Creative Education*, 4(12), 40–45.