International Journal of Instruction
e-ISSN: 1308-1470 ● www.e-iji.net

July 2023 ● Vol.16, No.3
p-ISSN: 1694-609X
pp. 191-204

Article submission code: 20220730112359

Received: 30/07/2022
Revision: 22/12/2022

Accepted: 13/01/2023
OnlineFirst: 02/04/2023

Development of Three Tier Open-Ended Instrument to Measure Chemistry Students' Critical Thinking Disposition Using Rasch Analysis

Andi Wahyudi
Alma Ata University, Indonesia, *andiwahyudi@almaata.ac.id*

Rino Richardo
Alma Ata University, Indonesia, *rinorichardo@almaata.ac.id*

Ingo Eilks
Bremen University, Germany, *ingo.eilks@uni-bremen.de*

Christoph Kulgemeyer
University of Bremen, Germany, *kulgemeyer@physik.uni-bremen.de*

Being able to assess students' critical thinking disposition is a prerequisite for measuring students' achievements before and after corresponding instruction. This study provides an instrument that teachers can use to identify students' critical thinking disposition regarding the acid-base chemistry concept. The critical thinking disposition in the acid-base-chemistry instrument (CTD-AB) has been utilized by the 3D+1I model (defining, designing, developing, and implementing). Participants were 31 students' vocational high school (16 females and 15 males). Data from the implementation was analyzed by Rasch analysis and Confirmatory Factor Analysis to assess the psychometrical quality of the instrument. Factor analysis confirmed the CTB-AB model. Rasch analysis shows that the CTD-AB is unidimensional and has an acceptable fit. CTD-AB has good item reliability and good internal consistency. The results of this study indicate that CTD-AB is a reliable and valid measurement tool that can be used to measure students' critical thinking disposition regarding the acid-base concept. Further research should reveal cross-validation of the CTD-AB in a bigger sample (N>100) and different educational settings.

Keywords: confirmatory factor analysis, critical thinking disposition, instrument development, rasch analysis, chemistry students

INTRODUCTION

Critical thinking (CT) is suggested as an essential skill in the Industrial Revolution (IR) 4.0 (Lee et al., 2018) and, probably, one of the most valuable human assets to face global competitiveness (Qiang et al., 2020). In education, the importance of critical

Citation: Wahyudi, A., Richardo, R., Eilks, I., & Kulgemeyer, C. (2023). Development of three tier open-ended instrument to measure chemistry students' critical thinking disposition using rasch analysis. *International Journal of Instruction*, *16*(3), 191-204. https://doi.org/10.29333/iji.2023.16311a

thinking because it will help students in forming habits to evaluate knowledge and information from credible sources (Robillos, 2022). Most researchers agree that CT is divided into two dimensions: critical thinking skills (CTS) and critical thinking disposition (CTD) (Chen et al., 2020). CTS covers individuals' abilities to analyze and comprehend problems and develop reasonable solutions for the identified problems (Sosu, 2013). CTD is a tendency to think critically (Ennis, 1996), or an operational manner in which individuals approach a task (Facione, 2000; Qiang et al., 2020). A person has the ability to think critically only if he possesses CTS, but also if the disposition is present in his life.

After the concept of CTS was established, researchers and educators began to conduct intense studies on CTD (Q. Chen et al., 2020). A person who has CTD is able to relate his life to life in general through the process of seeking information, being curious, and believes that reasoning is an investigative process (Nugroho et al., 2018). Many research explored the various learning innovation to analyze their impact on CTD (Seçer, 2022; Ulger, 2018; Zhou et al., 2012). The most popular instrument to measure CTD is the California Critical Thinking Disposition Inventory (CCTDI), including several features, such as truth-seeking, open-mindedness, analyticity, systematicity, and self-confidence, inquisitiveness, and maturity of judgment (Facione, 1991). CCTDI is a standard psychometric instrument to measure the disposition to engage with problems and make decisions using critical thinking (Facione, 2000). Although widely used, the instrument is not free of problems. Critique on the CCTDI came from cross-validation studies that show inconsistencies in pattern loading, construct overlap, and instability structure (Sosu, 2013). This critique has motivated other researchers to develop and create a different view on the CCTDI. Several methods were used to find alternatives to measure CTD by making instruments more specifically (Walsh et al., 2007)

Despite any critique, many authors started applying the CCTDI to measure the effect of learning strategies for improving students' CTD (Wang et al., 2019; Zhou, 2012). The CCTDI, however, showed limitations in measuring any improvement of students' CTD based on a changed learning strategy (Syahfitri et al., 2019). The progress of CTD may be due to other factors such as the influence of the personal living environment. Many alternative instruments were developed, for example, a shorter CCTDI version using a different CTD scale (Sosu, 2013), a need for cognition scale (Cacioppo et al., 1996), or the NEO five-factor inventory (Costa & McCrae, 1992). Another study focused CTD in biology education in the context of certain content (Syahfitri et al., 2019). To contribute filling existing gaps, the present study aims at developing an instrument to measure CTD in the context of acid-base-chemistry. The instrument can be used to see the effect of learning strategies on the improvement of students' CTD in acid-base-chemistry specifically.

METHOD

Research Design

Determining construct validity is an important part of instrument development (Farzad et al., 2020). One approach is to collect arguments for an appropriate construct validity based on confirmatory factor analysis (CFA). CFA depends on classical test theory (CTT) and aims to investigate the relationship between each item and an underlying construct (latent variable) (Marsh et al., 2020). In addition to investigating the relationship between items and latent variables, CFA is also suggested to evaluate the model. Model evaluation is analyzed with fit satisfaction to test model suitability (Vaingankar et al., 2020)

Another important aspect of the development of a suitable instrument is testing the hypothesis that the items cover the same construct. This part can be achieved using Rasch Measurement Theory (RMT) (Behmke & Atwood, 2013). This approach is suggested to deal with a limitation of CFA as one approach in CTT. CTT is unable to compare two tests in different groups (Lu & Bi, 2016). When two different types of tests are given to two different groups, the test items cannot be compared. Probabilistic test theory (PTT) can be used to overcome this limitation and to develop accurate measurement instruments (Deng & Wang, 2017). PTT, namely the so-called Rasch Model, is an analytical tool that can test the reliability of research instruments and test the suitability of persons and items simultaneously (Wren & Barbera, 2014).

PTT has several advantages. It fulfills the five principles of a good measurement model: it can provide a linear scale with the same intervals; predict missing data; provide a precise estimate; detect model inaccuracies; and produce replicable measurements (Hale et al., 2016). Also, PTT can be used to discuss the relationship between the person and items by providing a clear comparison on the same scale.

Many studies developed instruments combining CFA and Rasch analysis. This combination was done, e.g., in the instrument development of the Persian version of Patient-Rated Wrist and Hand Evaluation (Farzad et al., 2020), Bangla fear of Covid-19 scale (Sakib et al., 2020), or reading assessment (Randall & Engelhard, 2010).

This study uses a combination of CFA and Rasch analysis. The research used in this study focused on the development through the 3D+1I model (defining, designing, developing, and implementing) (Figure 1). A literature study was carried out for the defining step through the assembling of the CTD indicators and acid-base content. The designing step was carried out by making a three-tier open-ended test. In this step, the question items were compiled based on the CTD indicators. The developing step related the instruments by experts' judgment to calculate the content validity ratio (CVR). The implementation step relates to the empirical study of the instrument through the analysis of implementation results using CFA and Rasch analysis.

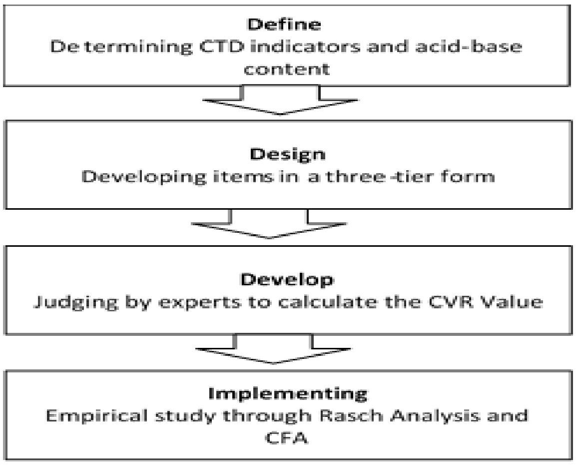

Figure 1
Steps of the research design

Sample and location

Participants in this study were 31 students of a vocational high school in Cimahi, one of districts in West Java, Indonesia. Map of Cimahi can be shown in Figure 2. The samples were 16 females (Teteh) and 15 males (Akang). Akang and Teteh are gender identifiers given to young people in Sundanese. The Sundanese are a tribe in West Java (Indonesia), where this study is located. The average age was 16.8 years. The study was carried out in Bahasa, the dominant language in Indonesia. All legal regulations for data collection and studies with human beings were taken into account.

Figure 2
Maps of Cimahi, West Java –Indonesia
Source: https://www.google.com/maps/place/Kabupaten+Cimahi

The sample size is rather small and leads to limitations (W. H. Chen et al., 2014). For example, a small sample (N≤50) in our study might lead to incorrectly ordered item parameters. However, W. H. Chen et al., (2014) argue that a small sample (N>30) might still be appropriate for exploratory purposes if the results are treated with caution or confirmed using an appropriate alternative to item evaluation, such as a classical theory test or nonparametric statistics. In this case, we used Rasch Model plus Confirmatory Factory Analysis (CFA) to develop our instrument.

Data Analysis

The data analysis techniques in this study were CVR, Rasch analysis, and CFA. The CVR method was used to identify the quality of items and questions. CVR calculates the validity ratio based on the results of expert ratings (Wilson et al., 2012). The experts' rate how well items fit the model (CTD and acid-base). CFA was used to analyze construct validity. CFA investigates the relationship between latent and observed variables (Duarte et al., 2020). A measure for this relationship value is the factor loading. To accept a suitable construct validity, CFA should satisfy three criteria: factor loading > 0.5, composite variable > 0.7 and average variance extracted (AVE) > 0.7 (Fornell & Larcker, 1981). The Rasch instrument includes item and person measurement, level of difficulty and discrimination, and outlier or misfit of item questions (Hale et al., 2016; Lu et al., 2020). Twenty-six items were analyzed using scoring with WINSTEP 3.73. The scoring can be seen in Table 1. The scoring of the CTD-AB instrument used six categories and was modified from three-tier diagnostic tools (Milenković et al., 2016). Tier 1 and Tier 3 have a score of 1, while tier 2 has a score of 2. Tier 2 has a greater score than the other tiers because tier 2 describes and explains the item question.

Table 1
Scoring guided to the CTD-AB instrument

Item Score	Tier-score			Description
	Tier 1 (score1)	Tier 2 (score 1)	Tier 3 (score 2)	
None	-	-	-	Students cannot answer even one-tier.
0	x	x	x	Students' answers are not correct in all tiers.
1	√	x	x	Students answer the first or third tiers correctly, but not the second.
	-	x	√	
2	√	x	√	Students answer correctly in the first- and third-tiers, but get a score of 0 in the second tier; or students answer correctly in the second tier, but not the first and third.
	x	√	x	
3	√	√	x	Students answer correctly either the first or third-tier combined with a correct answer in the second tier, but not the remaining tier.
	x	√	√	
4	√	√	√	Students answer all tiers correctly.

FINDINGS AND DISCUSSION

Defining step

In the defining step, acid-base content and every indicator of CTD were identified. The acid-base content in this study included the Arrhenius-theory, acid-base properties, degree of ionization, pH, neutralizing reaction, the Brönsted-Lowry-theory, the Lewis-theory, hydration, hydrolysis, buffer, and titration. The CTD included truth-seeking, open-mindedness, analytical, systematicity and inquisitiveness. These indicators were the primary guide in developing the instrument items related to the acid-base content.

Design step

The design step was carried out by compiling question items in the form of a three-tier test. The first tier is a multiple-choice question, the second tier asks to justify the answer given in the first tier, and the third tier asks the trust in the answer provided in the second tier. Examples from the instrument are given in Figures 3.

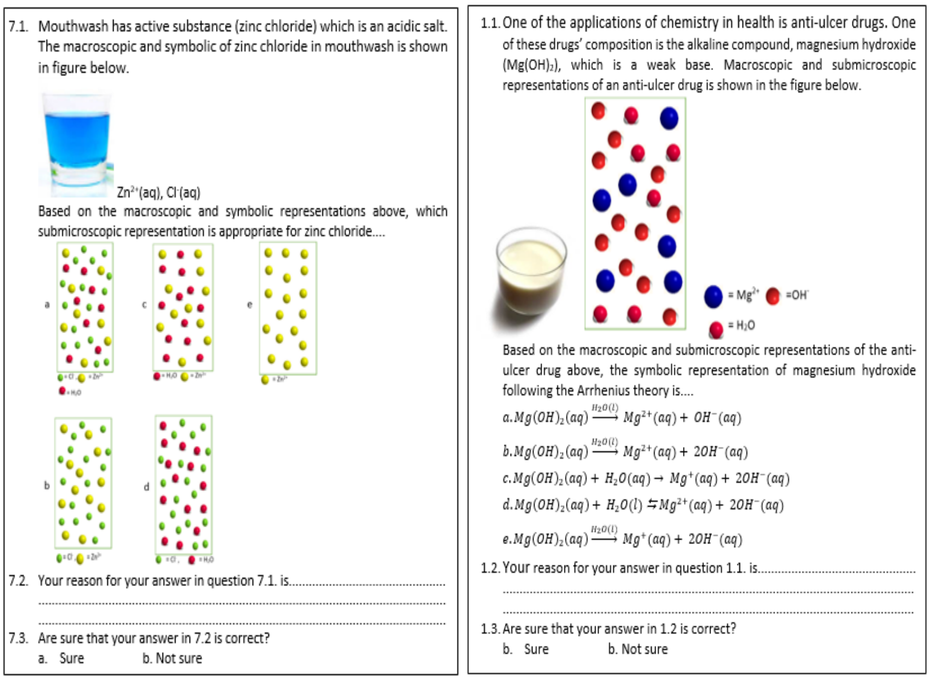

Figure 3
Question 1 (Q1) and Question 7 (Q7) of CTD-AB instrument

Item number 1 (Q1), examines students' truth-seeking in the Arrhenius-theory. Q1 is a question about anti-ulcer drugs where students are shown representations at the macroscopic and submicroscopic levels. In this item, students are asked to identify the correct symbolic representation of the behavior of $Mg(OH)_2$ in water. Students are then asked to justify their decision and how sure they are about it.

Item number 7 (Q7) examines students' inquisitiveness in the classification of acid and base components. Students are shown macroscopic and symbolic representations from zinc chloride ($ZnCl_2$) as an active substance in mouthwash. Students were asked to identify the most suitable submicroscopic representation of zinc chloride in water. A correct answer indicates that students have the curiosity to know things correctly, even if they are not immediately or obviously useful. Mapping CTD, acid-base content, and items can be seen in Table 2.

Table 2
Mapping items for CTD and acid-base content

CTD	Acid-Base content	items
Truth-seeking	Arrhenius theory	1, 14
	Lewis theory	3, 16
	Hydration	9, 22
	Buffer	11, 24
Open-Mindedness	Bronsted-Lowry theory	2, 15
	Hydrolysis	10, 23
Inquisitiveness	Acid-base classification	7, 17
	Degree of ionization	6, 19
Analyticity	Acid-base properties	5, 18
	pH	4, 20
Systematicity	Neutralizing reaction	8, 21
	Titration	12, 13, 25

CTD-AB consisted of 25 items and was made in the three-tiers form. In this instrument, there are five CTD indicators and 12 acid-base content. Each CTD indicator consists of two acid-base contents, except in truth-seeking. This is because the truth-seeking follows the characteristics of the more acid-base contents. For example, problem 1.1 requires finding the truth based on the given macroscopic and submicroscopic representations.

Development

Experts' judgment was used for looking suitability of the CTD indicators and the items. In the development step, 26 items were judged by five experts to calculate the content validity ratio (CVR). The CVR value is 0.73 calculated according to Wilson et al. (2012) in their description of the CVR concept from Lawshe (1975). The CVR shows that all item questions are in accordance with the content in question, except item 11. Item 11 has not been approved by one expert, so it received a CVR value of 0.60 and has been eliminated. The content validity index (CVI) is 0.98, which indicates that the items are in accordance with the content and CTD indicators.

Implementation

After Q11 has been eliminated, 25 items were administered to 31 students. The implementation was analyzed using the scoring criteria in Table 3. The Rasch-analysis revealed that the person measurement has a model-person reliability coefficient of 0.90 (which can be regarded as excellent). It is a little different from the real-person reliability coefficient of 0.89. This value indicates participants' consistency in answering the items questions. The person reliability indicates high potential that results would be replicated in a follow-up test (Hale et al., 2016).

Table 3
Result of person and item Rasch-analysis

Characteristics	Person	Item
Mean	22.9	29.5
SE of Mean	0.20	0.6
Real Reliability	0.89	0.93
Model Reliability	0.90	0.94
Real Separation	2.83	3.67
Model Separation	3.08	4.09

Rasch-analysis revealed an expected a posterior or plausible value (EAP/PV) of 0.93. The EAP/PV indicates predictive reliability in the items. The model of item reliability coefficient is 0.94 and little different from the real-item reliability coefficient is 0.93. The coefficient of the item reliability is higher than the person reliability which parallels other cases **(He et al., 2016)**. It indicates that the item has the ability to be answered consistently. **DeVellis (2003)** notes that scale reliabilities of 0.65-0.70 might be acceptable and of 0.70-0.85 are respectable for instrument to be used for research purposes (Rauch & Hartig, 2010). The values show the consistency of the items being developed. The instrument stability is excellent to provide results when applied in different research settings. This assumption is based on the item value of good reliability and small standard error value <.3.

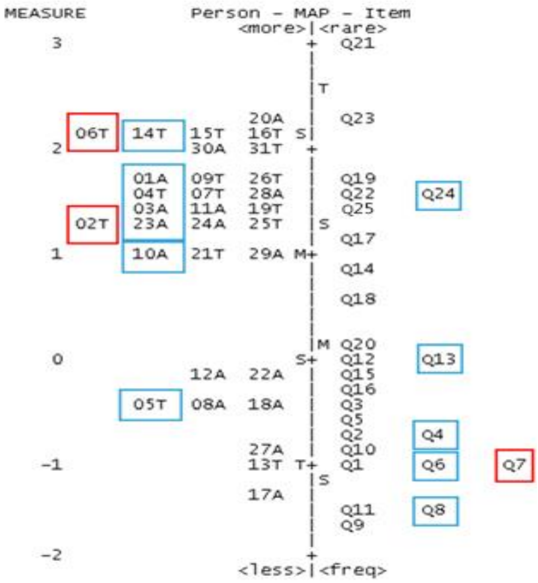

Figure 4
Scalogram of CTD-AB instrument

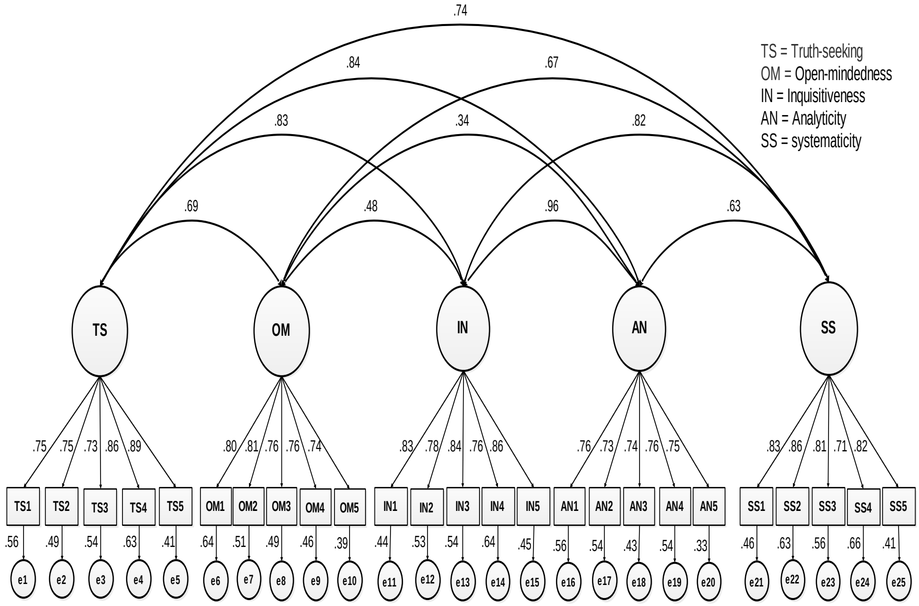

Figure 5
Confirmatory factor analysis (CFA) of CTD-AB instrument

Figure 4 shows the scalogram result from the instrument. In the left side is the distribution map of the participants which have an initial A for "*Akang*" and initial T for "*Teteh*" to differentiate male and female participants. In the right side of the scalogram, there is the distribution of question items. There is a relationship between items of critical thinking disposition and participants. Q7 (red border) is an item that is out of the line (outfit). The items that go off track state that the item does not fully measure CTD. Apart from the item, participants can also be categorized as outfit participants (red border). Outfit participants on the scalogram were 02T and 06T. These participants have the ability to perform above the others (except 20A). The participants left the pattern because they tended to respond correctly in difficult items and to answer incorrectly in the easy items. This is possible because they might have been cheating or were influenced by other factors, such as misconceptions. However, identifying the corresponding misconception would be a different study that could also be subject to Rasch-analysis as done, e. g., by Adimayuda et al. (2020) or Aminudin et al. (2019).

CFA analysis was performed on each indicator / observed variable of CTD (Figure 5). From each observed variable, there are five latent variables. The purpose of this CFA is to analyze the relationship between observed variables and latent variables. The relationship between the observed variable and the latent variable is called factor loading (Hu & Bentler, 1999). This factor loading shows the quality of an item on its construct. The results of the CFA modeling show that the quality of the items in representing the construct is in a good way. This is based on the minimum factor loading criteria, which describes the quality of a good item that must be greater than 0.5 (Hair et al., 2010). Factor loading shows the large correlation between the indicator and its latent construct (Hu & Bentler, 1999). Indicators with a high factor loading have a higher contribution to explain the latent construct. On the other hand, indicators with low loading factors have a weak contribution to explain the latent construct.

The modelling feasibility was tested with various kinds of tests (Table 4) all due diligence developed to have a very good index, namely the chi square tes (χ^2)/ degree of freedom (df), probability, minimum sample discrepancy function divided by its degrees of freedom (Cmin/df), Adjusted Goodness of Fit Index (AGFI), Tucker-Lewis Index (TLI), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA). An exception is the Groningen Frailty Indicator (GFI), which has a value of 0.756. This value is smaller than the cut of the value of the GFI. However, the other feasibility tests have met the criteria of the expected model development. In this section, the GFI value can be ignored because the due diligence is fulfilled from the others (Syahfitri et al., 2019).

Table 4
Cut-off criteria for Feasibility index of TTCD-AB (Hu & Bentler, 1999)

Fit Index	Cut of value	Test value
χ^2 /df	≤3	2.725
Probability	≥0.05	0.078
Cmin/df	≥2	2.780
GFI	≥0.90	0.756
AGFI	≥0.90	0.902
TLI	≥0.95	0.906
CFI	≥0.90	0.956
RMSEA	≤0.08	0.044

CONCLUSIONS

Our study explored a new evaluation method for measuring CTD in the context of acid-base-chemistry. The development of the CTD-AB instrument in acid-base-chemistry learning can be an alternative for researchers in measuring the increase in students' CTD. Instrument was developed with three tier open-ended form. The first tier is a multiple-choice question with five choices, the second tier is open-ended question that students’ answer from the first tier and the third tier is students’ confidence after answering the first tier and second tier with two choices "Sure" and "Not Sure". The instrument was developed based on CVR, Rasch analysis and CFA. Rasch analysis shows that the instrument has good reliability (item reliability and person reliability). Person reliability shows the level of consistency of the participants in answering the questions, while item reliability shows the level of consistency of the items when measured by different research settings. Overall, Rasch-analysis indicated that the CTD-AB instrument satisfies IRT test standards. Meanwhile, CFA shows that the construct validity of the instrument is good. The model developed has met the feasibility index, except for the GFI value. A total of 26 items developed, there are 25 items that meet the CVR, Rasch analysis and CFA standard. The results of this study can be an alternative instrument that can be used to analyze the impact of learning innovations on students'

CTD, especially on acid-base chemistry concept. In addition, the results of this study can also be used as a guide for instrument development by combining CTT and PTT. However, the study is limited in a comparably small sample. Future research should reveal whether the promising findings in this study can be replicated with larger samples (N≥100) and in different educational environments.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Alma Ata University, Bremen University and the University of Paderborn for support to this research.

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

Adimayuda, R., Aminudin, A. H., Kaniawati, I., Suhendi, E., & Samsudin, A. (2020). A multitier open-ended momentum and impulse (MOMI) instrument: Developing and assessing quality of conception of 11th grade sundanese students with rasch analysis. *International Journal of Scientific and Technology Research*, *9*(2), 4799–4804.

Aminudin, A. H., Adimayuda, R., Kaniawati, I., Suhendi, E., Samsudin, A., & Coştu, B. (2019). Rasch analysis of Multitier Open-ended Light-Wave Instrument (MOLWI): Developing and assessing second-years sundanese-scholars alternative conceptions. *Journal for the Education of Gifted Young Scientists*, *7*(3), 557–579. https://doi.org/10.17478/jegys.574524

Behmke, D. A., & Atwood, C. H. (2013). Implementation and assessment of Cognitive Load Theory (CLT) based questions in an electronic homework and testing system. *Chemistry Education Research and Practice*, *14*(3), 247–256. https://doi.org/10.1039/c3rp20153h

Cacioppo, J. T., Petty, R. E., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*(2), 197–253.

Chen, Q., Liu, D., Zhou, C., & Tang, S. (2020). Relationship between critical thinking disposition and research competence among clinical nurses: A cross-sectional study. *Journal of Clinical Nursing*, *29*(7), 1332–1340. https://doi.org/10.1111/jocn.15201

Chen, W. H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*(2), 485–493. https://doi.org/10.1007/s11136-013-0487-5

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO personality inventory and NEO five factor inventory*. Psychological Assessment Resources.

Deng, Y., & Wang, H. (2017). Research on evaluation of Chinese students' competence in written scientific argumentation in the context of chemistry. *Chemistry Education Research and Practice*, *18*(1), 127–150. https://doi.org/10.1039/c6rp00076b

DeVellis, R. F. (2003). *Scale development: theory and applications 3rd edition*. SAGE publication Ltd.

Duarte, P. A. S., Palmeira, L., & Pinto-Gouveia, J. (2020). The three-factor eating questionnaire-R21: a confirmatory factor analysis in a portuguese sample. *Eating and Weight Disorders*, *25*(1), 247–256. https://doi.org/10.1007/s40519-018-0561-7

Ennis, R. H. (1996). Critical thinking dispositions: their nature and assessability. *Informal Logic*, *18*(2), 165–182. https://doi.org/10.22329/il.v18i2.2378

Facione, P. A. (1991). *Using the california critical thinking skills test in research, evaluation, and assessment*. California Academic Press. https://doi.org/10.1111/j.1471-8286.2007.01696.x

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, *20*(1), 61–84. https://doi.org/10.22329/il.v20i1.2254

Farzad, M., MacDermid, J. C., Lu, Z., & Shafiee, E. (2020). Validation of persian version of patient-rated wrist and hand evaluation: Confirmatory factor analysis and Rasch analysis. *Archives of Rehabilitation Research and Clinical Translation*, *2*(4), 1–11. https://doi.org/10.1016/j.arrct.2020.100076

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Pearson Educational, Inc. https://doi.org/10.1016/j.ijpharm.2011.02.019

Hale, L. V. A., Lutter, J. C., & Shultz, G. V. (2016). The development of a tool for measuring graduate students' topic specific pedagogical content knowledge of thin layer chromatography. *Chemistry Education Research and Practice*, *17*(4), 700–710. https://doi.org/10.1039/c5rp00190k

He, P., Liu, X., Zheng, C., & Jia, M. (2016). Using Rasch measurement to validate an instrument for measuring the quality of classroom teaching in secondary chemistry lessons. *Chemistry Education Research and Practice*, *17*(2), 381–393. https://doi.org/10.1039/c6rp00004e

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lee, M., Yun, J., Pyka, A., Won, D., Kodama, F., Schiuma, G., Park, H., Jeon, J., Park, K., Jung, K., Yan, M.-R., Lee, S., & 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), 1–24. https://doi.org/10.3390/joitmc4030021

Lu, H., Jiang, Y., & Bi, H. (2020). Development of a measurement instrument to assess students' proficiency levels regarding galvanic cells. *Chemistry Education Research and Practice*, *21*(2), 655–667. https://doi.org/10.1039/c9rp00230h

Lu, S., & Bi, H. (2016). Development of a measurement instrument to assess students' electrolyte conceptual understanding. *Chemistry Education Research and Practice*, *17*(4), 1030–1040. https://doi.org/10.1039/c6rp00137h

Marsh, H. W., Guo, J., Dicke, T., Parker, P. D., & Craven, R. G. (2020). Confirmatory Factor Analysis (CFA), Exploratory Structural Equation Modeling (ESEM), and Set-ESEM: Optimal balance between goodness of fit and parsimony. *Multivariate Behavioral Research*, *55*(1), 102–119. https://doi.org/10.1080/00273171.2019.1602503

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

Nugroho, P. B., Nusantara, T., As'ari, A. R., Sisworo, Hidayanto, E., & Susiswo. (2018). Critical thinking disposition: Students skeptic in dealing with ill-logical mathematics problem. *International Journal of Instruction*, *11*(3), 635–648. https://doi.org/10.12973/iji.2018.11343a

Qiang, R., Han, Q., Guo, Y., Bai, J., & Karwowski, M. (2020). Critical thinking disposition and scientific creativity: the mediating role of creative self-efficacy. *Journal of Creative Behavior*, *54*(1), 90–99. https://doi.org/10.1002/jocb.347

Randall, J., & Engelhard, G. (2010). Using confirmatory factor analysis and the Rasch model to assess measurement invariance in a high stakes reading assessment. *Applied Measurement in Education*, *23*(3), 286–306. https://doi.org/10.1080/08957347.2010.486289

Rauch, D. P., & Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, *52*(4), 354–379.

Robillos, R. J. (2022). Impact of LoiLooNote Digital Mapping on University Students' Oral Presentation Skills and Critical Thinking Dispositions. *International Journal of Instruction*, *15*(2), 501–518. https://doi.org/10.29333/iji.2022.15228a

Sakib, N., Bhuiyan, A. K. M. I., Hossain, S., Al Mamun, F., Hosen, I., Abdullah, A. H., Sarker, M. A., Mohiuddin, M. S., Rayhan, I., Hossain, M., Sikder, M. T., Gozal, D., Muhit, M., Islam, S. M. S., Griffiths, M. D., Pakpour, A. H., & Mamun, M. A. (2020).

Psychometric validation of the Bangla fear of COVID-19 scale: Confirmatory factor analysis and rasch analysis. *International Journal of Mental Health and Addiction*, 1–12. https://doi.org/10.1007/s11469-020-00289-x

Seçer, E. (2022). Investigation of the Effect of Learning Styles on Critical Thinking Dispositions of Physiotherapy and Rehabilitation Students. *Archives of Health Science and Research*, *9*(2), 116–122. https://doi.org/10.54614/ArcHealthSciRes.2022.21111

Sosu, E. M. (2013). The development and psychometric validation of a Critical Thinking Disposition Scale. *Thinking Skills and Creativity*, *9*, 107–119. https://doi.org/10.1016/j.tsc.2012.09.002

Syahfitri, J., Firman, H., Redjeki, S., & Sriyati, S. (2019). Development and validation of critical thinking disposition test in biology. *International Journal of Instruction*, *12*(4), 381–392. https://doi.org/10.29333/iji.2019.12425a

Ulger, K. (2018). The Effect of Problem-Based Learning on the Creative Thinking and Critical Thinking Disposition of Students in Visual Arts Education. *The Interdisciplinary Journal of Problem-Based Learning*, *12*(1), 3–6.

Vaingankar, J. A., Abdin, E., Van Dam, R. M., Chong, S. A., Tan, L. W. L., Sambasivam, R., Seow, E., Chua, B. Y., Wee, H. L., Lim, W. Y., & Subramaniam, M. (2020). Development and validation of the Rapid Positive Mental Health Instrument (R-PMHI) for measuring mental health outcomes in the population. *BMC Public Health*, *20*(1), 1–12. https://doi.org/10.1186/s12889-020-08569-w

Walsh, C. M., Seldomridge, L. A., & Badros, K. K. (2007). California critical thinking disposition inventory: Further factor analytic examination. *Perceptual and Motor Skills*, *104*(1), 141–151. https://doi.org/10.2466/PMS.104.1.141-151

Wang, X., Sun, X., Huang, T., He, R., Hao, W., & Zhang, L. (2019). Development and validation of Critical Thinking Disposition Inventory for Chinese medical college students (CTDI-M). *BMC Medical Education*, *19*(200), 1–14.

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. https://doi.org/10.1177/0748175612440286

Wren, D., & Barbera, J. (2014). Psychometric analysis of the thermochemistry concept inventory. *Chemistry Education Research and Practice*, *15*(3), 380–390. https://doi.org/10.1039/c3rp00170a

Zhou, Q. (2012). Integrating webquest into chemistry classroom teaching to promote students' critical thinking. *Creative Education*, *3*(3), 369–374. https://doi.org/10.4236/ce.2012.33058

Zhou, Q., Yan, C., Zhao, S., Liu, L., & Xing, L. (2012). A preliminary investigation into critical thinking of in-service and pre-service middle school chemistry teachers in Shaanxi province of China. *Asia-Pacific Forum on Science Learning and Teaching*, *13*(2), 1–13.