https://doi.org/10.15625/2615 - 8965/22310201

Validation of the progression of chemical identity thinking using Rasch analysis: A response to challenges in educational measurement

Jonathan M. Barcelo
jmbarcelo@slu.edu.ph
School of Nursing, Allied Health,
and Biological Sciences,
Saint Louis University
(Philippines)

ABSTRACT: *Objective measurement of students' reasoning competencies in science contexts is relevant in promoting educational reforms in science education. Over the years, progression-based assessments have been prompted by applying Rasch analysis in developing research instruments. Towards contributing to this reform, this study describes the development and validation of the Chemical Identity Thinking Instrument, which is anchored on a hypothesized progression of chemical identity thinking among pre-medical students. Item content validity indices were determined to inform item restructuring and revision. The research instrument was administered to 362 first-year to third-year pre-medical students from five higher education institutions in the Philippines. A scoring rubric was utilized to evaluate the accuracy of answers, the accuracy of explanations, and the type of explanations of students. Based on these parameters, the level of chemical identity thinking of the students was determined. By applying the Rasch rating scale model, evidence of reliability and construct validity was obtained. The applications of the instrument and the scoring rubric were discussed in the context of assessment and evaluation in pre-medical programs. The interpretation of changes in student abilities along the hypothesized levels of chemical identity thinking in all items was also described.*

KEYWORDS: **Chemical identity, chemical identity thinking, pre-medical students, Rasch analysis, reasoning patterns**

➔ Received 16/03/2023 ➔ Revised manuscript received 14/05/2023 ➔ Published 30/06/2023.

1. Introduction

In chemistry courses, students are trained to compare substances based on their properties and utilize relevant chemical information in decision-making activities that involve chemical analysis. The ability to compare substances is built on the knowledge of chemical identity, an "*attribute of a substance which makes it different from other substances*" (Ngai & Sevian, 2016). Reasoning competencies that invoke the comparison of chemical identity are relevant in various health-related contexts such as nutrition, diagnostics, or pharmacology. In pre-medical programs, knowledge of chemical identity empowers students in evaluating the safety, efficacy, and function of the substances, thereby supporting the development of clinical reasoning that is anchored on an evidence-based framework. Chemical identity is one of the cross-cutting concepts that facilitate the development of chemical thinking (Sevian & Talanquer, 2014) which is necessary to apply chemistry concepts to various contexts. In order to compare substances, an individual should possess adequate chemical identity thinking, a thinking process that is anchored on the ability to compare or differentiate substances (Ngai & Sevian, 2016). Knowledge of chemical identity allows the correct comparison of properties or characteristics of substances (Enke, 2001), and adequate chemical identity thinking is necessary to guide decision-making. It is therefore logical to promote the progression of chemical identity thinking if high levels of decision-making skills involving chemistry concepts are desired.

The progression of chemical identity thinking requires the application of chemical knowledge and cognitive skills that are necessary to classify substances (Ngai, 2017). Choosing a unique and appropriate property as a basis for comparing substances requires a wide range of knowledge on explicit and implicit properties, molecular

https://doi.org/10.15625/2615 - 8965/22310201

composition, and molecular interactions. Hence, it is recognized that chemical identity thinking is an important facet to be developed especially if students are expected to develop science-based literacy that involves chemical knowledge. While knowledge of chemical identity is an important aspect of chemical identity thinking, it is the sophistication of chemical identity thinking that predicts the ability of an individual to apply chemical identity in actual decision-making activities. As a part of the chemical thinking framework, chemical identity thinking is concerned with answering the question "*what is this substance*" (Banks et al., 2015; Ngai & Sevian, 2016). At present, the recent advances in the educational landscape of pre-medical programs require the development and progression of chemical identity thinking among students to link relevant chemistry concepts to clinical decisions.

The concept of chemical identity has long been integrated into chemistry courses offered in pre-medical programs, such as general inorganic chemistry, organic chemistry, and biochemistry. Hence, it is expected that pre-medical students possess various levels of chemical identity thinking. In these chemistry courses, substances are often classified based on their molecular structures, properties, or composition. By choosing a unique property of a substance, it is possible to detect a substance in a complex mixture or design methods that can help in the quantification of substances in complex matrices such as biological fluids. By determining the quantity, it is possible to make decisions that are context-bound. This line of thinking is also applied when detecting biomarkers in biological samples as a basis for diagnosis, or describing the effectiveness of a product (i.e., medicine, foodstuff, herbal products) based on the type and amounts of active ingredients. In addition, pharmacological agents are designed and developed based on their intended therapeutic functions. With the advent of newer technologies in drug analysis, chemical structure is considered one of the important characteristics that premedical students should learn, and chemical identity thinking that utilizes molecular structures and interactions is desired. With these applications, the development of an adequate level of chemical identity thinking is desired among pre-medical students.

Objective educational measurement is necessary to measure the progression of chemical identity thinking among pre-medical students. The measurement of chemical identity thinking needs to be based on a linear scale to provide an empirical basis for justifying curricular changes and educational reforms. In previous studies, chemical identity thinking was evaluated using a qualitative analysis of reasoning patterns to ascertain types of chemical identity thinking (Ngai et al., 2014; Ngai & Sevian, 2016; Ngai & Sevian, 2018). The progression of chemical identity thinking was based on the conceptualization of matter types, types of properties used in making decisions about chemical identity, and major reasoning patterns such as *objectivization, principlism, compositionism,* and *interactionism* (Ngai et al., 2014). While qualitative analysis of chemical identity thinking provides robust descriptions of the thinking processes of students, measuring the progression of chemical identity thinking using this method may be difficult to interpret. Measuring the progression of chemical identity thinking is also necessary for introducing teaching interventions that support the achievement of *structure-property* thinking, which is a desired thought process among students who have completed their chemistry courses. Notably, the chemical identity progression involves a shifting of reasoning using the explicit properties to molecular composition, then finally, to the interaction of molecular components that are linked to emergent properties. While the progression of chemical identity thinking was hypothesized in this study to follow a linear pattern, learning is often characterized to be non-linear (Reed & Wolfson, 2021), suggesting a possible heterogeneity of the increase of chemical identity thinking when laid out in a linear progression-based framework.

Chemical identity thinking is relevant in promoting the development of *structure-property* thinking among pre-medical students, as the progression of chemical identity thinking invokes knowledge of molecular composition and interactions. While it is uncommon for health science students to use physical and chemical properties when performing clinical roles, a high level of chemical identity thinking is required to understand chemical processes in metabolic pathways and disease pathogenesis. Similarly, health advice on nutrition and diet modification may not involve molecular composition and interactions but a high level of chemical identity thinking is required when understanding food-drug interactions. In pre-medical programs, substances are classified based on their intended health effects or functions, and sophisticated chemistry explanations may be less prioritized. But it must be emphasized that biological systems are complex, and the chemical environment of a cell has its own set of emergent properties (Van Regenmortel, 2004), requiring a good grasp of what the chemical identity is. A chemical identity thinking that requires knowledge of molecular structures and interactions is necessary if pre-medical students will conduct research related to pharmacology or diagnostics as part of their terminal requirements in their programs. This justifies why the progression of chemical identity thinking among pre-medical students is an important outcome of chemistry education in health science programs.

Measuring the progression of chemical identity thinking resonates with the need to respond to the challenges of assessment and evaluation in science education. Providing an empirical point of reference to determine the progression of chemical identity thinking is possible if there is a linear scale that can be utilized along a hypothesized progression of the construct. With the limitations of the literature on how to measure and report chemical identity thinking, there is a need to develop a scoring rubric that is based on the hypothesized progression described by Ngai & Sevian (2016) and evaluate the validity of the progression using Rasch analysis, a psychometric technique that is utilized to improve the precision of research instruments (Boone, 2016). This study applied Rasch analysis in the development and psychometric evaluation of the *Chemical Identity Thinking Instrument*, an instrument that is designed to measure the chemical identity thinking of students using a rubric that measures the accuracy and reasoning patterns when differentiating substances. This study was guided by the following research questions:

(1) What are the item content validity indices of the *Chemical Identity Thinking Instrument*?

(2) What is the evidence of reliability and construct validity of the *Chemical Identity Thinking Instrument* based on Rasch analysis?

2. Literature review

2.1. Studies on evaluation of chemical identity thinking

At present, there are only a few studies that have investigated the chemical identity thinking of undergraduate students. In these studies, the evaluation of the chemical identity thinking of students is based on a rigorous qualitative analysis of explanations in a survey that elicits various types of reasoning patterns related to chemical identity thinking. In the study of Ngai and Sevian (2016), a survey was conducted among students from grade 8 to university students who are enrolled in a general chemistry course. Items that were included in the survey were those which can elicit reasoning patterns that are related to *objectivization*, *principlism*, *compositionism*, and *interactionism*. The explanations of students were analyzed qualitatively to classify the responses. Based on their findings, students who have limited background in chemistry used reasoning patterns based on *objectivization*. In addition, explanations related to *objectivization* and *principlism* were more common among university students enrolled in their chemistry courses.

The effect of a teaching intervention on chemical identity thinking has been investigated using a qualitative evaluation of student explanations. Sanah et al. (2019) implemented problem-based learning in teaching redox and electrochemistry concepts to senior high school students. Based on the results of their teaching intervention, there was an increase in the number of students who utilized *principlism*, *compositionism*, and *interactionism* among students after the intervention, while the number of students who previously utilized *objectivization* decreased after the teaching intervention. While the context of chemical identity thinking was only limited to redox and electrochemistry concepts, it was demonstrated that a progression in chemical identity thinking is possible if students had prior knowledge about the concepts, and if the teaching approach involved investigation of a problem.

The chemical identity thinking of biochemists and biochemistry students was also investigated by Ngai and Sevian (2018). In the study, it was found that individuals with higher expertise in biochemistry, particularly those who were using liquid chromatography-mass spectroscopy, regarded chemical identity thinking as important in the analysis of biochemicals. Among students, the researchers found out that the theme "*composition and structure*" were elicited in the explanations when molecular structures were provided. Furthermore, it was also observed that when students were not explicitly required to identify and differentiate substances, they are less likely to utilize certain types of explanations. Lastly, familiarity with the substances was observed to influence the utilization of various types of explanations. These findings imply that eliciting the progression of chemical identity thinking is based on the type of task, context, and expertise with molecular structures and properties.

2.2. Application of Rasch analysis in validating the progression of chemical identity thinking

The development of chemical identity thinking follows a progression or sophistication of reasoning, from *objectivization* to *principlism*, to *compositionism*, and *interactionism* (Ngai, 2017). Along these reasoning patterns, a high level of chemical identity thinking is characterized by the ability to describe the molecular composition and interactions in diverse chemical systems, promoting the development of *structure-property* thinking. The most advanced level of chemical identity thinking (*interactionism*) relates well to the analysis of molecular interactions to describe the properties of substances. Such a thinking process is important in understanding the role of molecular interactions in understanding properties such as boiling point, viscosity, and polarity (Sevian & Talanquer, 2014; Talanquer, 2018). It was hypothesized by Ngai et al. (2014) that chemical identity thinking should progress as students learn more about chemistry concepts, particularly those that are related to molecular structures and molecular properties. Furthermore, several factors, such as the presence of cues, familiarity, and prior conceptual knowledge all play significant roles in the progress of chemical identity thinking (Ngai & Sevian, 2018). While qualitative studies are commonly used to describe the level of chemical identity thinking, no study has explicitly validated the progression of chemical identity thinking.

The application of Rasch analysis in validating the progression of chemical identity thinking was based on the previous studies that have utilized Rasch models in supporting the construct validity of hypothesized progression of reasoning competencies in various contexts (Clark & Watson, 2019; Fiedler et al., 2019; Maeng, 2020; Yao & Guo, 2017). To the author's knowledge, no attempt has been done to validate the progression of chemical identity thinking, despite a robust description of reasoning patterns. Rasch analysis has the potential to provide evidence on the validity of the progression of chemical identity thinking based on types of reasoning patterns and allow the comparison of the level of the chemical identity of different cohorts. In addition, utilizing a Rasch model in validating a progression of chemical identity thinking allows the determination of a teaching intervention when attempting to improve students' ability to differentiate substances.

Rasch theory serves as a guide in formulating an instrument that represents a range of "test-item difficulty" to its respondents (Boone, 2016). With Rasch analysis, validated instruments can be developed to evaluate the effectiveness of the learning progressions of students (Herrmann-Abel & De Boer., 2017). Since the progression of chemical identity thinking has been documented rigorously, this study will adopt the types of reasoning patterns – *objectivization*, *principlism*, *compositionism,* and *principlism* – and incorporate them in a scoring system to identify the level of chemical identity thinking of the students in various types of substances and materials. Performing Rasch analysis provides evidence of construct and validity and aids in the refinement of research instruments that can be used in the assessment and evaluation of student abilities in several contexts. Furthermore, Rasch analysis converts raw scores into linear measures, allowing the comparison of student abilities with respect to item difficulties. With these advantages, it is not surprising that several chemistry education studies have utilized Rasch analysis to give a more objective measurement of a construct, including this present study.

3. Methodology

3.1. Development of the Chemical Identity Thinking Instrument

When developing a research instrument that is anchored on a hypothesized progression, item difficulty can be hypothesized based on an existing framework. As shown in Table 1, the *Chemical Identity Thinking Instrument* was composed of 20 items involving the differentiation of substances or materials. The first part was composed of 10 items, and the task was to evaluate whether the substances or materials were similar or different. In these items, no cues were provided to students to elicit different types of reasoning patterns when providing explanations. The second part of the instrument was composed of 10 situations that were presented with cues (MC = molecular structure, P = property, S = source, or F = function). In these items, students were tasked to compare the substances or materials. Eight chemistry teachers who had at least five years of teaching chemistry courses to pre-medical students evaluated the difficulty of the items based on the hypothesized progression of chemical identity thinking. The difficulty of the items was hypothesized based on three criteria: the presence of cues, type of task, and familiarity with substances, in relation to the findings of Ngai and Sevian (2018). It was hypothesized

Table 1. Summary of assessment items in the Chemical Identity Thinking Instrument

Item	Code*	Topic	Hypothesized difficulty
Item 01	HM1	Ethyl alcohol vs. rubbing alcohol	Easy
Item 02	HM2	Fat vs. oil	Easy
Item 03	B1	Starch vs. flour	Easy
Item 04	HM3	Albumin vs. egg white	Easy
Item 05	B2	Butter vs. margarine	Easy
Item 06	HM4	Rubber vs. plastic	Moderately difficult
Item 07	B3	Dairy milk vs. soya milk	Moderately difficult
Item 08	HM5	Glass vs. fiberglass	Easy
Item 09	GIOC1	Copper vs. bronze	Moderately difficult
Item 10	GIOC2	Aluminum vs. silver	Moderately difficult
Item 11	B4	Amino acid vs. protein powder (P)	Moderately difficult
Item 12	B5	Tryptophan vs. adenosine (MC)	Difficult
Item 13	B6	Vitamin A vs. Isotretinoin (S)	Difficult
Item 14	B7	Aspartame vs. sugar (MC)	Difficult
Item 15	HM6	Boiled egg vs. scrambled egg (P)	Moderately difficult
Item 16	B8	Nucleotide vs. dietary nucleotide (F)	Difficult
Item 17	GIOC3	Ethyl alcohol vs. acetic acid (S)	Difficult
Item 18	B9	Sucrose vs. lactose (MC)	Difficult
Item 19	B10	Starch in pasta vs. starch in potatoes (MC)	Difficult
Item 20	HM7	Uncooked rice vs. steamed rice (P)	Difficult

Note: HM = household materials; GIOC = general inorganic and organic chemistry; B = biochemistry that items that were more familiar were easier to describe and differentiate. If cues are presented, it was hypothesized that items that were presented with molecular structures were more difficult.

Items that required students to evaluate the similarity of two materials of substances were hypothesized to be easier, while items that require students to analyze cues and apply the cues when differentiating substances are considered more difficult. Lastly, topics that were discussed in general inorganic and organic chemistry (GIOC) were also classified as less difficult compared to topics that are discussed in biochemistry (B). However, when the topics are presented with cues that are involving molecular structures, the items were hypothesized to be difficult as well.

3.1.2. Item structure

The first ten items have three parts – a situation, dichotomous options (same or different), and a blank box for the explanations. In each item, students were required to evaluate whether two substances or materials are the same or different. Then, students were tasked to provide an explanation of why they think the substances or materials were the same or different. An example of an item is shown below.

Item Q01

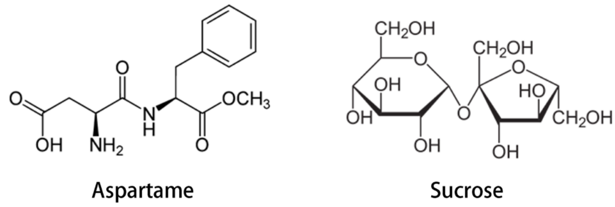
Ethyl alcohol vs. rubbing alcohol

Answer and Explanation

The last ten items had two parts – a situation and a blank box where students wrote their explanations. The situation included the cues that were used as a point of reference to compare substances or materials. The cues were either related to molecular structures, properties, sources, or functions. An example of an item (Item 14) is shown below.

Item Q14

The molecular structures of aspartame and sucrose are shown below. In what way are they different?



Answer and Explanation

3.2. Item content validity of the Chemical Identity Thinking Instrument

Item content validity was determined to evaluate whether the items can be included in the final draft of the research instrument or not. The first draft of the *Chemical Identity Thinking Instrument* was sent to two Ph.D. Chemistry Education professors, one Ph.D. Chemistry professor, and one Ph.D. Biochemistry professor for determining item-content validity, which was obtained using Lawshe Content Validity Index and a researcher-made content validity form. The procedure and interpretation of the item-content validity indices were based on previous studies (Ayre & Scally, 2014; Gilbert & Prion, 2016). At an item level, the item content validity index (I-CVI) was computed as the number of experts evaluating an item as "essential" divided by the total number of experts. Relevancy is based on how the item relates to the construct of the instrument. The validators rated each item as non-essential (NE), useful (U), or essential (E) based on how each item related to the main construct or concept of the test or instrument. The item-content validity indices express the proportion of agreement on the relevancy of each item (which is between zero and one) and were supplemented by the *kappa* statistics, which is a consensus index of inter-rater agreement that adjusts for chance agreement.

To calculate the modified *kappa* statistics, the probability of chance agreement was first calculated for each item by the following formula: PC= [N! / A! (N–A)!]*0.5N]. In this formula, N refers to the number of experts in a panel and A refers to the number of panelists who agree that the item is essential. After calculating item-content validity indices for all items, *kappa* statistics were determined by entering the numerical values of probability of chance agreement (pc) and item-content validity index of each item in the formula: K = (I-CVI – pc) / (1–pc). If the *kappa* statistics is above 0.74, between 0.60 and 0.74, and between 0.40 and 0.59, the values are classified as excellent, good, and fair, respectively (Cicchetti, 2001). Table 2 shows that the item content validity indices ranged from 0.750 to 1.000 and are interpreted as fair to excellent (Polit & Beck, 2006). However, the I-CVI results implied that two items (Item Q16 and Item Q19) need to be revised. These items were restructured based on the recommendations of the evaluators, such as shortening the statements, modifying the molecular structures, and improving the clarity of the presented contexts.

3.3. Hypothesized progression of chemical identity thinking

The progression of chemical identity thinking was described to follow four levels of reasoning patterns. For the purpose of Rasch analysis, the characteristics of levels were determined to the abilities of the target participants for whom the research instrument is going to be used. In pre-medical courses, students have already completed their high school chemistry subjects encompassing general chemistry, basic inorganic chemistry, and basic organic chemistry. In pre-medical programs, the most common chemistry subjects include general inorganic chemistry, general organic chemistry, analytical

Table 2. Item content validity indices and kappa statistics of the Chemical Identity Thinking Instrument

Item	I-CVI	Interpretation of I-CVI	pc	k	Interpretation of k
Item 01	1.000	Appropriate	0.062	1.000	Excellent
Item 02	1.000	Appropriate	0.062	1.000	Excellent
Item 03	1.000	Appropriate	0.062	1.000	Excellent
Item 04	1.000	Appropriate	0.062	1.000	Excellent
Item 05	1.000	Appropriate	0.062	1.000	Excellent
Item 06	1.000	Appropriate	0.062	1.000	Excellent
Item 07	1.000	Appropriate	0.062	1.000	Excellent
Item 08	1.000	Appropriate	0.062	1.000	Excellent
Item 09	1.000	Appropriate	0.062	1.000	Excellent
Item 10	1.000	Appropriate	0.062	1.000	Excellent
Item 11	1.000	Appropriate	0.062	1.000	Excellent
Item 12	1.000	Appropriate	0.062	1.000	Excellent
Item 13	1.000	Appropriate	0.062	1.000	Excellent
Item 14	1.000	Appropriate	0.062	1.000	Excellent
Item 15	1.000	Appropriate	0.062	1.000	Excellent
Item 16	0.750	Revise	0.125	0.714	Good
Item 17	1.000	Appropriate	0.062	1.000	Excellent
Item 18	1.000	Appropriate	0.062	1.000	Excellent
Item 19	0.750	Revise	0.125	0.714	Good
Item 20	1.000	Appropriate	0.062	1.000	Excellent

Note: I-CVI = item-content validity index; k = kappa statistic; pc = chance agreement

https://doi.org/10.15625/2615 - 8965/22310201

chemistry, and biochemistry. In most cases, general inorganic chemistry and general organic chemistry were offered as chemistry prerequisites to biochemistry. Hence, the items that were included in the *Chemical Identity Thinking Instrument* were substances and materials that are commonly discussed in the prerequisite chemistry courses such as representative of household materials and organic compounds.

The lowest level of chemical identity thinking needs to consider the ability of pre-medical students in the Philippine setting. While it is expected that the levels of chemical identity thinking can be linked to four levels directly, a careful justification of the proposed levels of chemical identity thinking using the Rasch model was done to ensure meaningful and contextualized hypothesized progression levels in the intended cohort. Pre-medical students have varied levels of prior knowledge of chemistry concepts, and inaccurate answers or alternative misconceptions were expected due to various factors. It is possible that the explanations may include different types of chemical identity thinking, although the explanations may be inaccurate, unrelated, or focused on superficial or generic concepts only (e.g., substances are made of molecules). Hence, the lowest level (Level 0) was labeled as '*unsupported or inaccurate chemical identity thinking*' and is characterized by grossly inaccurate or unrelated explanations. This was the lowest anchor in the hypothesized progression of chemical identity thinking. This level also included answers which were not supported by an explanation. Based on the literature, the lowest level of chemical identity thinking was *objectivization* (Ngai & Sevian, 2016; Ngai & Sevian, 2018). The next level of chemical identity thinking was presumed to be correct and predominantly anchored on *objectivization* reasoning patterns. A Level 1 chemical identity thinking, or '*chemical Identity thinking based on objectivization*", was hypothesized to be characterized by adequate and correct reasoning patterns that are based on *objectivization*. If students only described the molecular structures based on the atoms present in the cues, the reasoning pattern was still considered *objectivization*, as students are merely describing the presented molecular structures. The third level (Level 3) is '*chemical identity thinking based on principlism*'. In the study of Ngai & Sevian (2018), it was reported that experts often utilize interactionism in their explanations. However, the *Chemical Identity Thinking Instrument* was designed to be used for undergraduate students who are non-experts in chemistry. However, the topics in chemistry often include interactions of molecular components, in relation to the *structure-property* relationship concept. It was expected that students can utilize *compositionism* since molecular compositions were taught in prerequisite chemistry courses. However, it might be uncommon to expect students to utilize a thinking pattern based on *interactionism*, as

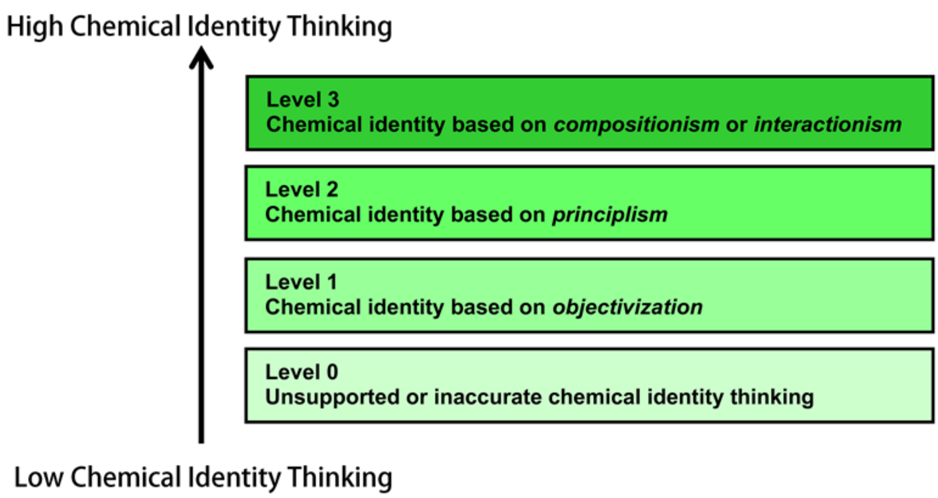


Figure 1. Hypothesized progression of chemical identity thinking

this requires advanced chemistry knowledge and familiarity with the *structure-property* relationship. Because of these reasons, the third level (Level 3) was based on the utilization of correct explanations that are anchored on molecular composition or molecular interactions. The last level was called '*chemical identity based on compositionism and interactionism,*' and was considered the highest level for now. The proposed progression of chemical identity thinking is illustrated in Figure 1.

3.4. Scoring rubric

Table 3 shows the scoring rubric that was used in determining the level of chemical identity thinking of students. The score of the students in each item was determined using three criteria: accuracy of the answer, accuracy of explanation, and type of explanation. The scores ranged from 0 to 3, corresponding to the four levels in the hypothesized progression of chemical identity thinking. The types of explanations were based on the levels of chemical identity thinking described in the study by Ngai & Sevian (2016). The explanation was coded as "0" if it is incorrect, and "1" if it is correct. Then, the type of reasoning pattern was identified and classified into *objectivization*, *principlism*, *compositionism*, or *interactionism*. If the explanations involved two levels of chemical identity thinking, the higher level was used as a reference for determining the type of reasoning pattern.

3.4. Ethics approval

The research protocol of this study was a component of the main research protocol that involved the utilization of the validated *Chemical Identity Thinking Instrument* as one of the research instruments for pretest-posttest comparison after implementing mereology instruction and conventional instruction in the biochemistry course of pre-medical students in a tertiary institution in the Philippines. The main research protocol was reviewed last February 18, 2019, approved last March 25, 2019, and was assigned with Protocol Number SLU-REC 2019-021.

3.5. Data collection

After ethics approval, letters of communication were sent to various tertiary institutions in the Philippines offering pre-medical courses to coordinate the schedule of data gathering. A total of 362 students composed of 101 third-year BS Medical Technology students who completed general inorganic and organic chemistry and biochemistry from one tertiary-level institution in Metro Manila, 24 second-year medical technology students who completed general inorganic and organic chemistry from one tertiary-level institution in Pampanga, 113 second-year BS Medical Technology who completed general inorganic chemistry and biochemistry in a tertiary-level institution in Metro Manila, 91 second year BS Biology students who completed general inorganic chemistry and organic chemistry from one tertiary-level institution in Baguio City, and 33 second-year BS Medical Technology students who completed general inorganic chemistry and organic chemistry from a tertiary-level institution in Pampanga were recruited to answer the research instrument. Data gathering was conducted from March to August 2019.

Table 3. Scoring rubric used in assigning the hypothesized level of chemical identity thinking of students

Accuracy of Answer	**Explanation**			**Level**
	Accuracy	**Type of Reasoning Pattern**		
Correct	Correct	Compositionism or Interactionism	3	Level 3
Correct	Correct	Principlism	2	Level 2
Correct	Correct	Objectivization	1	Level 1
Incorrect	Correct or Incorrect	Objectivization, Principlism, Compositionism, or Interactionism; No explanation	0	Level 0

3.6. Interrater reliability

All answers and explanations were encoded in Microsoft Excel, then sent to two raters who had sufficient knowledge in general inorganic chemistry, organic chemistry, and biochemistry. The first rater graduated with a Bachelor's degree in Chemistry and currently finishing the MS Chemistry program. The second rater was an MS Chemistry graduate. The two raters underwent eight hours of training in using the scoring rubric (Table 3) to classify the types of explanations. Due to the pandemic last 2020, the evaluation of answers was postponed until October 2021. The two raters evaluated a total of 120 explanations (6 explanations per item) independently for two weeks. The interrater reliability was Kappa = 0.87 ($\rho<0.001$), indicating a good agreement between the raters (Landis & Koch, 1977; McHugh, 2012).

3.7. Data analysis

After data entry, Rasch analysis was applied to determine the validity of the hypothesized progression of chemical identity thinking using the hypothesized levels in the study of Ngai & Sevian (2016) and to evaluate the item functioning. In this study, the Rasch rating scale model was utilized, since the progression of chemical identity thinking was hypothesized to be similar in all items. In easier items (familiar items), achieving the progression of chemical identity was hypothesized to be less difficult compared to the progression in difficult items (e.g., less familiar items with cues related to molecular structures). It is argued that the ability of students to exhibit higher levels of chemical identity thinking is influenced by prior knowledge about substances and materials, and the presence of available cues. All answers were rated using the same scoring system. The Rasch rating scale model was deemed appropriate since each item has the same threshold for the rating scale categories (Planinic et al., 2019). Other studies have also applied the rating scale model in assessing reasoning competencies (Fiedler et al., 2019) and rubric-based scoring (Li & Wang, 2021), which used a similar approach to this study.

Rasch analysis was performed using Winsteps 4.4.5. Rasch indices such as person and item reliability and person and item separation were evaluated to determine the instrument quality and item hierarchy. To determine item fit, the mean square (MNSQ) and standardized values (ZTSD) outfit and infit summary statistics were also investigated. MNSQ values between 0.70 and 1.30 and ZTSD values below between -2.0 to +2.0 were considered acceptable for this instrument, based on the recommendation of Bond and Fox (2007). To determine whether an item contributed to the construct that is being measured, the point-measure correlation was also determined (Arnold et al., 2018). Unidimensionality was evaluated by using principal component analysis (PCA) on standardized residuals. The unexplained variance by the first contrast should be below 5% (less than 2 eigenvalues) if the items are exhibiting unidimensionality (Linacre, 2012). A high-quality research instrument should show evidence for unidimensionality, adequate fit statistics of items and persons, proper targeting, and item difficulties that are within the range of person abilities. For rating scale models, the category response curves should also exhibit a proper ordering of the hypothesized levels. In this case, the order of categories should be from Level 0 (*Unsupported or inaccurate chemical identity thinking*), followed by Level 1 (*chemical identity thinking based on objectivization*), then Level 2 (*chemical identity thinking based on principlism*), and lastly, Level 3 (*chemical identity based on compositionism or interactionism*).

4. Results

In this study, Rasch analysis was utilized to test the construct validity of the *Chemical Identity Thinking Instrument*, an instrument that was designed to measure the progression of chemical identity thinking through the use of a scoring rubric. The progression of chemical identity was hypothesized to be characterized primarily by reasoning patterns: *objectivization*, *principlism*, *compositionism*, and *interactionism* (Ngai & Sevian, 2016). The research instrument was designed to elicit different reasoning patterns from students when they differentiate various substances and materials. In this section, the results of Rasch analysis are presented.

4.1. Separation and reliability

The *Chemical Identity Thinking Instrument* showed adequate reliability (Pr = 0.82, Ir= 0.98), and separation (Ps= 2.16, Is= 7.51). Ideally, the person separation index should be greater than 2, the item separation index should be 3 while person reliability should be greater than 0.80, and while item reliability should be greater than 0.90 (Linacre, 2011).

4.2. Item fit and item difficulty

Table 4 shows that all MNSQ infit and outfit values of items in the *Chemical Identity Thinking Instrument* were within 0.70 to 1.30, and the ZSTD values were within -2.0 to +2.0. An item with MNSQ that is greater than 1.30 means that there is more than 30% variation than what Rasch modeling would predict (underfitting), while an MNSQ that is lower than 0.70 means that there is greater than 30% less variation than what is predicted by the model (overfitting) (Bond & Fox, 2003). The point measure correlations ranged from 0.41 to 0.54, indicating that all items contribute to the measurement of the construct chemical identity thinking. When the items fit the Rasch model, it can be assumed that all items contribute to the measurement of the underlying construct (Oon et al., 2016). Furthermore, all items had point measure correlations ranging from 0.42 to 0.51, indicating that all items contribute to the measurement of the underlying construct, chemical identity thinking.

Based on mean item difficulty, the most difficult items were related to biochemistry (M = 0.31, SD = 0.49), followed by items related to general inorganic and organic chemistry (M = 0.17, SD = 0.5), then items

Table 4. Item fit of the Chemical Identity Thinking Instrument

Item	Measure	SE	Infit		Outfit		Pt. Meas. Correl.
			MNSQ	ZSTD	MNSQ	ZSTD	
Item 01	0.08	0.08	1.04	0.57	1.03	0.38	0.51
Item 02	-0.26	0.07	0.92	-1.20	0.92	-1.14	0.51
Item 03	0.04	0.07	1.07	1.06	1.08	1.11	0.41
Item 04	-1.19	0.07	1.02	0.29	1.03	0.53	0.51
Item 05	-0.39	0.07	0.95	-0.75	0.95	-0.74	0.50
Item 06	-0.77	0.07	1.04	0.65	1.05	0.78	0.50
Item 07	0.22	0.08	1.07	0.99	1.08	1.10	0.43
Item 08	-0.65	0.07	0.99	-0.14	0.99	-0.15	0.49
Item 09	-0.12	0.07	1.09	1.24	1.10	1.39	0.49
Item 10	0.33	0.08	1.00	0.05	1.00	-0.03	0.45
Item 11	-0.36	0.07	1.01	0.14	0.99	-0.09	0.50
Item 12	0.81	0.08	1.00	0.03	0.98	-0.22	0.49
Item 13	0.13	0.08	1.06	0.80	1.05	0.73	0.45
Item 14	0.87	0.08	0.96	-0.59	0.94	-0.74	0.45
Item 15	-0.90	0.07	1.02	0.38	1.02	0.27	0.49
Item 16	1.06	0.08	0.90	-1.41	0.93	-0.90	0.42
Item 17	0.29	0.08	0.96	-0.61	0.95	-0.74	0.54
Item 18	0.33	0.08	0.96	-0.53	1.00	0.06	0.45
Item 19	0.36	0.08	0.89	-1.49	0.92	-1.06	0.45
Item 20	0.11	0.08	0.96	-0.54	0.96	-0.53	0.53

related to household materials (M = -0.51, SD = 0.49). It can be surmised that the students can differentiate household materials easier than the substances that are discussed in their chemistry subjects because these materials are commonly used, and students are more familiar with the surface features of these materials. The most difficult item in the instrument is Item 16 (*nucleotide vs. dietary nucleotide*), followed by Item 14 (*aspartame vs. sucrose*), and Item 12 (*tryptophan vs adenosine*). These difficult items were presented with molecular structures and were commonly discussed in biochemistry contexts.

4.2. Unidimensionality

The unexplained variance by the first contrast should be below 5% or less than 2 eigenvalues to ensure unidimensionality (Linacre, 2012). Based on the results of Rasch analysis, the first contrast had an eigenvalue = 1.57, which satisfied the condition for unidimensionality. Unidimensionality is a major requirement of the Rasch model, as it ensures that only one latent trait is being measured (Planinic et al., 2019).

4.3. Item category function

The four levels of the *Chemical Identity Thinking Instrument* were ordered in correspondence to the rating categories (Figure 2). Since the thresholds increased monotonically for adjacent categories, it means that the hypothesized progression of levels of chemical identity thinking are functioning as intended in all items related to household materials, general inorganic and organic chemistry, and biochemistry. The average observed measures should increase monotonically as the rating scale increases. The distinct peaks on the category probability curve indicate the most probable level along the continuum.

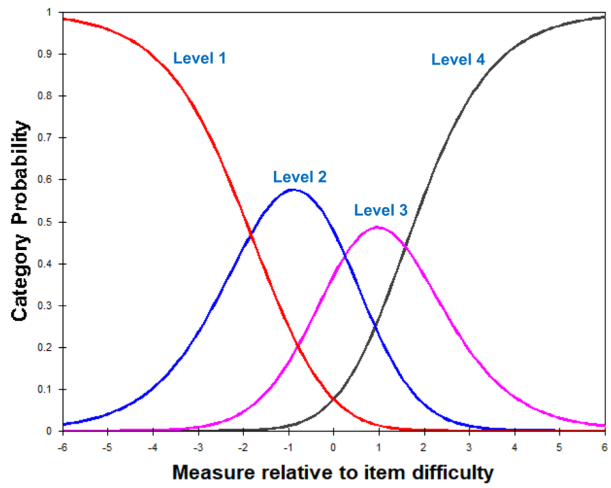


Figure 2. Category probability curve of the levels of Chemical Identity Thinking Instrument

4.4. Validity of progression of chemical identity thinking

Table 5 shows that the four categories, corresponding to the four levels of chemical identity thinking, have acceptable infit and outfit MNSQ. With an increase in the category score (0 to 1 to 2 to 3), the category measures also increased from -3.02 to 2.85, indicating that the rating scale is functioning as intended. The results imply that the hypothesized progression of chemical identity thinking based on the levels reported by Ngai & Sevian (2016) is valid. However, it must be noted that the levels in this study involved the merging of reasoning patterns anchored on *compositionism* and *interactionism* at the highest level.

Table 5. Summary of the rating scale category

Category	Observed Count	Observed %	Infit MNSQ	Outfit MNSQ	Category Measure
0	1484	20	1.00	1.00	-3.02
1	3375	46	1.01	1.02	-0.87
2	1920	26	0.98	0.98	0.97
3	521	7	1.01	1.01	2.85

Note: Category 0 stands for "Level 1"; Category 1 stands for "Level 2"; Category 2 stands for "Level 3"; Category 3 stands for "Level 4".

4.5. Variable map of Chemical Identity Thinking instrument

Figure 3 shows the variable map of the items of the *Chemical Identity Thinking Instrument* based on the results of Rasch analysis. The variable map provides a useful reference to the threshold boundaries of each hypothesized level of chemical identity thinking progression in all items. In addition, the items are also arranged based on item difficulty, allowing a straightforward reference to the level of chemical identity thinking of a student in all items, based on their ability. The items are arranged based on increasing difficulty from bottom to top. This arrangement of items implies that the progression of chemical identity thinking is easier to be achieved for items at the bottom. Based on the item hierarchy, Item 04 is the easiest item. This means that biomedical students can easily exhibit a progression of their chemical identity thinking when differentiating albumin and egg white, provided that a teaching intervention that explicitly promotes chemical identity thinking progression is implemented. Furthermore, the relative location of the category thresholds provides information on whether a student has progressed from a lower level to a higher level along the hypothesized progression of chemical identity thinking. In contrast, the progression of chemical identity thinking of biomedical students is less likely to happen in difficult items (e.g., Item 16). It can be observed that Rasch analysis supported the hypothesized difficulty of all items along the progression of reasoning patterns when differentiating substances. The item hierarchy provides evidence of the construct validity of the hypothesized progression of chemical identity thinking in this study.

4.6. Discussion

In educational measurement, studies commonly use the Classical Test Theory, in which, raw scores are used to measure student abilities. Based on the available literature regarding the presentation of results of chemical identity thinking, the frequency of students using *objectivization*, *principlism*, *compositionism*, and *interactionism* is a possible method of

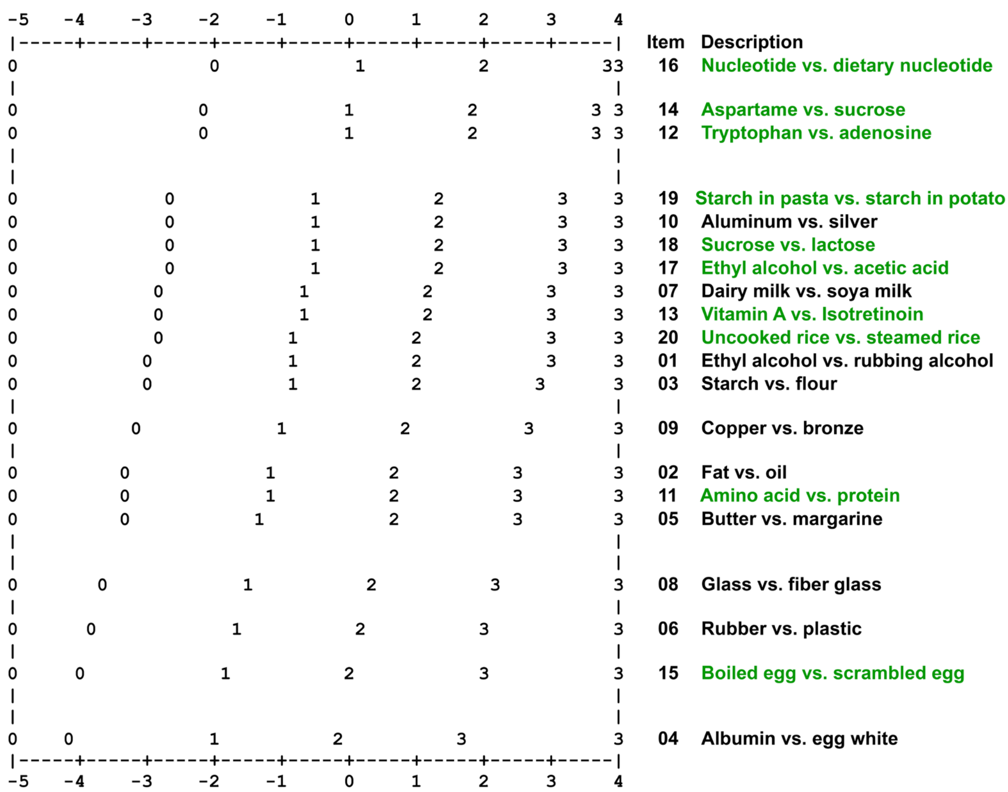


Figure 3. Variable map of items in Chemical Identity Thinking Instrument

Note: Items in green color were presented with cues presenting results. However, there is a call for objective educational measurement when designing a curriculum, assessing the impact of a teaching intervention, and determining the baseline competencies of students. In the past years, Rasch analysis was used to support the construct validity of research instruments in science education that is good in agreement with the theory and basic requirements of objective measurement (Planinic et al., 2019). In this study, Rasch analysis was utilized to test the construct validity, reliability, and item functioning of the *Chemical Identity Thinking Instrument,* in relation to a scoring rubric that determines the level of students' chemical identity thinking. The construct, chemical identity thinking, was based on the hypothesized progression of the trait among pre-medical students. The progression of chemical identity thinking was based on the four types of thinking that are rigorously described by Ngai & Sevian (2016) – *objectivization*, *principlism*, *compositionism*, and *interactionism*. By applying a Rasch rating scale model, the hypothesized progression of chemical identity thinking into four levels was validated using the proposed scoring rubric. In addition, the *Chemical Identity Thinking Instrument* showed adequate item fit and unidimensionality, ensuring an objective measurement of students' chemical identity thinking.

The analysis of chemical identity thinking involved a qualitative method. While this was also used in this study, the use of a rubric provided a definite classification of the level of chemical identity thinking that is exhibited by students in each item. Qualitative analysis of students' explanations is a rigorous and cumbersome method of determining chemical identity thinking. The rubric tool provided a more empirical basis for determining the level of chemical identity thinking, which was analyzed using the Rasch rating scale model. Several items in the *Chemical Identity Thinking Instrument* elicited specific types of reasoning patterns (i.e., molecular structures elicit a reasoning pattern anchored on *compositionism*; cues related to function elicited a reasoning pattern based on *objectivization*), suggesting the possible role of cues in influencing the level of students' chemical identity thinking. By designing items based on the type of task, situations, and cues, and applying the scoring rubric, the measurement of chemical identity thinking can become more objective and meaningful for educational measurement. When Rasch analysis supports the progression-level validity, reliability, and item fit of a research instrument and a rubric tool, the underlying construct can be objectively measured, thereby extending the application of the measurement to assessment and evaluation among pre-medical students.

The decision to utilize Rasch analysis in validating the progression of chemical identity thinking was supported by several advantages. First, the Rasch model explains how a person's performance regarding a specific trait, can predict that person's response in a particular test item involving that trait (Boone & Scantlebury, 2006). In educational measurement, studies commonly use the classical test theory, in which raw scores are used to measure changes in student abilities. However, raw scores are not linear (Bond & Fox, 2007; Planinic et al., 2019), and the differences between any two consecutive raw scores cannot be assumed to represent equal intervals (Boone & Notelmeyer, 2017). The sum of raw scores cannot be used to compare student performance because item difficulties are different in research instruments (Boone, 2016). Subjecting non-linear measures to statistical tests also produces distorted results (Planinic et al., 2019). The Rasch theory also serves as a guide in formulating an instrument that represents a range of "test-item difficulty" to its respondents (Boone, 2016). Using Rasch analysis, validated instruments can be developed to evaluate the effectiveness of a curriculum and the learning progress of students (Herrmann-Abel et al., 2017). With these advantages, it is not surprising that several chemistry education studies have applied Rasch analysis. A few studies in chemistry education have utilized Rasch analysis for various purposes. Rasch models have been used in chemistry education to develop a scoring rubric (Deng & Wang, 2017), measure conceptual understanding of chemistry concepts (Wei et al., 2012), develop concept inventories (Nedungadi et al., 2019) or evaluate the psychometric properties of an existing chemistry concept inventory (Barbera, 2013).

It is tempting to assign a score based on the type of chemical identity thinking level that is used. However, using the sum of scores as a basis for measuring chemical identity thinking is difficult to interpret, and may lead to erroneous inferences. It is also possible to simply use the rubric for evaluating chemical identity thinking without applying Rasch analysis. However, several problems may arise from utilizing raw scores and these could compromise the objective measurement of the underlying construct. Claiming that raw scores represent an objective measurement of the construct is inaccurate, as raw scores are nonlinear and the differences between raw scores cannot be assumed to represent equal intervals (Boone & Noltemeyer, 2017). Hence, raw scores do not accurately measure a person's actual ability in a particular construct underlying a research instrument. If science education researchers do not convert raw scores to a linear scale, an incorrect conclusion may be reached when using raw scores when using a parametric test (Boone & Scantlebury, 2006). This problem is addressed by Rasch analysis, as the underlying construct can be measured on a linear scale of log-odd unit or logit.

The Rasch rating scale model was utilized in this study because it was hypothesized that the progression of chemical identity thinking is more achievable if the context is familiar to students, highlighting the role of prior knowledge or familiarity with substances or materials that are being compared. The results indicate that the *Chemical Identity Thinking Instrument* has adequate reliability, with valid construct validity and unidimensionality. The results suggest that the instrument, along with the scoring rubric, can be used in various educational applications, such as evaluating the effects of teaching interventions in pre-medical programs or determining the progression of chemical identity thinking in various cohorts of pre-medical students. Based on the item hierarchy, the progression of chemical identity thinking was indeed easier to achieve in items that are more familiar (e.g., household materials). Items which were presented with molecular structures were more difficult for students, such as Item 12 (*tryptophan vs. adenosine*) and Item 14 (*aspartame vs. sucrose*). The inclusion of Item 16 (nucleotide vs. dietary nucleotide) among the difficult items indicates that topics on nucleic acids were perceived to be difficult for pre-medical students. In chemistry courses, molecular structures are taught to promote a conceptual understanding of molecular properties (Zarkadis et al., 2017) and emphasize *structure-activity* relationships (Carvalho et al., 2005). Students who find items difficult when presented with molecular structures may have experienced cognitive load, as decoding molecular structures is an arduous mental task.

Of particular importance is the utilization of the item variable map in assessing and evaluating the progression of chemical identity thinking among pre-medical students. Pre-test and post-test comparisons can be done per student based on ability logits, while interpretations on the level of chemical identity thinking can be done by plotting the student's ability across all items in a variable map. The student's level of chemical identity thinking can be determined in easy and difficult items, and a robust interpretation of the effects of the teaching intervention can be discussed in relation to the progression of chemical identity along specific contexts. Figure 4 shows the initial and final ability of a student with respect to the category thresholds of all items in the variable map. An initial ability of -0.33 logit indicates that the student exhibited Level 1 chemical identity thinking (*chemical identity based on objectivization*) in all items, except for Item 12, Item 14, and Item 16. If the student has been exposed to a teaching intervention that promotes the progression of chemical identity thinking, it is presumed that the ability will increase. If the final student ability was 1.90 logits, then the student is likely to exhibit Level 1 chemical identity thinking in Item 16 (most difficult item), Level 2 chemical identity thinking (*chemical identity thinking based on principlism*) in 18 items, and Level 3 chemical identity thinking (chemical identity thinking based on compositionism or interactionism) in Item 04 (easiest item). The variable map serves as a guide in interpreting changes in chemical identity thinking and allows comparison between groups or gender.

There are several limitations of this study. First, the hypothesized progression of chemical identity thinking was based on the reasoning patterns of first-year to third-year pre-medical students in the Philippines. While the types of chemical identity may be based on the four levels reported by Ngai and Sevian (2016), the items included in the *Chemical Identity Thinking Instrument* were contextualized to the curriculum of the target respondents. Other substances and materials that are relevant to other contexts, such as diagnostics, nutrition, and pharmacology can be explored using the described progression of chemical identity thinking in this study. Second, the progression of chemical identity thinking was based on the reasoning patterns that were described in the available literature. As observed in the results, students exhibited heterogeneous reasoning patterns, and modifications in the hypothesized progression may be necessary to include the analysis of explanations that utilize different types of reasoning patterns. While pre-medical students may be required to develop reasoning patterns that are anchored on *compositionism* or *interactionism* in selected chemistry topics, students may not always use *interactionism* when differentiating common materials that are found in households, as these are often characterized based on their function or source. This observation warrants a modification of the types of items included to explicitly elicit higher levels of chemical identity thinking. But in contexts where students should be able to explain molecular structures in relation to health contexts such as pharmacology and diagnostics, there is a need to improve the scoring rubric because the topics may involve more complex

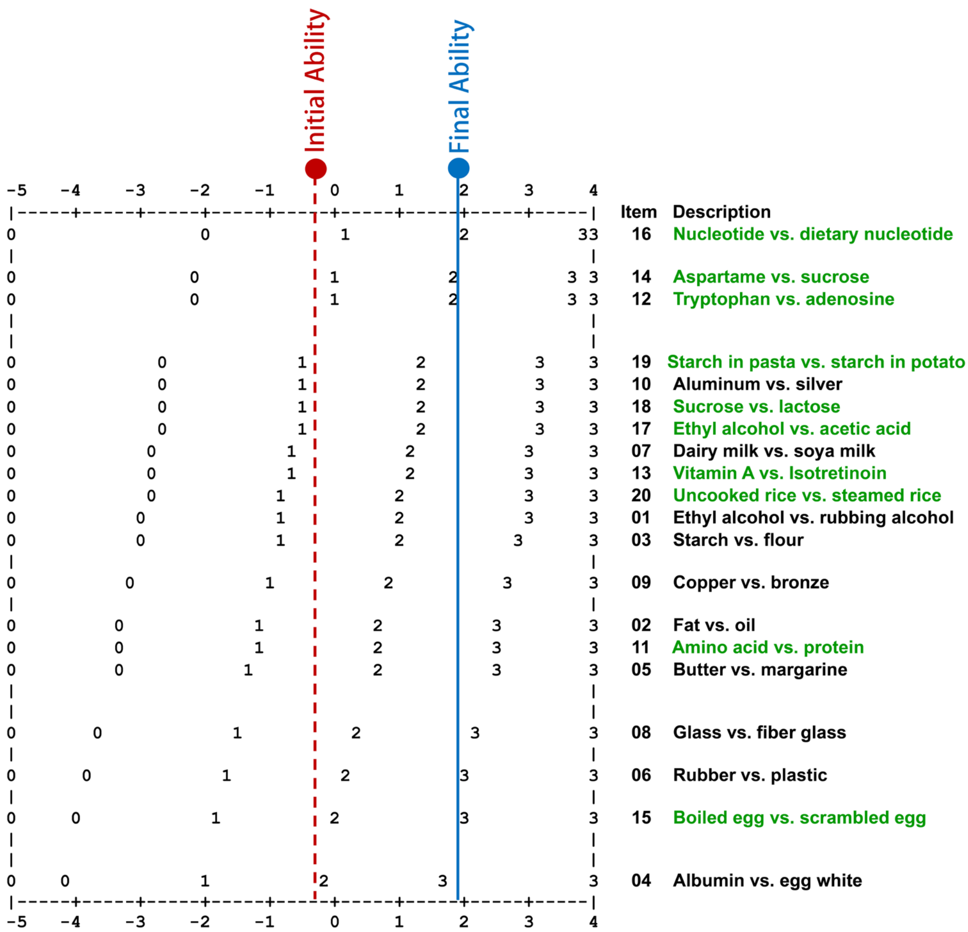


Figure 4. Application of the Chemical Identity Thinking Instrument variable map in determining the progression of chemical identity thinking

biological matrices that require multidisciplinary perspectives. Lastly, additional items may be included in succeeding versions of the *Chemical Identity Thinking Instrument* to increase the discrimination of student abilities and improve the construct validity of the research instrument.

5. Conclusions

Rasch analysis ensures an objective measurement of chemical identity thinking. Guided by the hypothesized levels of chemical identity thinking and the scoring rubric for specific contexts related to chemistry subjects offered in pre-medical programs, the progression of chemical identity thinking was validated in this study. In addition, the *Chemical Identity Thinking Instrument* is suitable for measuring the progression of chemical identity thinking of biomedical students in various items. Items which are based on molecular structures and biochemistry topics are more difficult and may require chemistry educators to design teaching interventions to improve chemical identity thinking, especially since biochemistry is a prerequisite of professional subjects in pre-medical programs. Overall, the Rasch rating scale model provides ample evidence of the reliability and construct validity of the research instrument.

Acknowledgments

The author would like to acknowledge the Department of Science and Technology Science Education Institute (DOST-SEI) through the Capacity Building Program in Science and Mathematics Education (CBPSME) for providing financial support during the conduct of the study. The author would also like to thank Dr. Nona Marlene B. Ferido for her valuable insights and suggestions.

References

Arnold, J., Boone, W., Kremer, K., & Mayer, J. (2018). Assessment of competencies in scientific inquiry through the application of Rasch measurement techniques. *Education Sciences, 8*(4), 184. https://doi: 10.3390/educsci8040184

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development, 47*(1), 79-86. https://doi.org/10.1177/0748175613513808

Banks, G., Clinchot, M., Cullipher, S., Huie, R., Lambertz, J., Lewis, R., Ngai, C., Sevian, H., Szteinberg, G., Talanquer, V., & Weinrich, M. (2015). Uncovering chemical thinking in students' decision making: A fuel-choice scenario. *Journal of Chemical Education, 92*(10), 1610-1618. https://doi.org/10.1021/acs.jchemed.5b00119

Barbera, J. (2013). A psychometric analysis of the Chemical Concepts Inventory. *Journal of Chemical Education, 90*(5), 546–553. https://doi.org/10.1021/ed3004353

Bond, T. G., & Fox, C. M. (2003). Applying the Rasch model: Fundamental measurement in the human sciences. *Journal of Educational Measurement, 40*(2), 185–187. https://doi.org/10.1111/j.1745-3984.2003.tb01103.x

Bond, T. G., & Fox, C. M. (2007). Applying the Rasch model: Fundamental measurement in the human sciences (2nd ed.). Lawrence Erlbaum Associates Publishers.

Boone, W. (2016). Rasch Analysis for instrument development: Why, when, and how? *CBE-Life Sciences Education, 15*(4), rm4. https://doi: 10.1187/cbe.16-04-0148

Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education, 4*(1). https://doi.org/10.1080/2331186x.2017.1416898

Boone, W. J., & Scantlebury, K. (2006). The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Science Education, 90*(2), 253-269. https://doi.org/10.1002/sce.20106\

Carvalho, I., Borges, Á. D. L., & Bernardes, L. S. C. (2005). Medicinal chemistry and molecular modeling: An integration to teach drug structure-activity relationship and the molecular basis of drug action. *Journal of Chemical Education, 82*(4), 588. https://doi.org/10.1021/ed082p588

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31*(12), 1412–1427. https://doi.org/10.1037/pas0000626

Cicchetti, D. V. (2001). Methodological commentary the precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology, 23*(5), 695–700. https://doi.org/10.1076/jcen.23.5.695.1249

Davidowitz, B., & Potgieter, M. (2016). Use of the Rasch measurement model to explore the relationship between content knowledge and topic-specific pedagogical content knowledge for organic chemistry. *International Journal of Science Education, 38*(9), 1483–1503. https://doi.org/10.1080/09500693.2016.1196843

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

Enke, C. G. (2001). The science of chemical analysis and the technique of mass spectrometry. *International Journal of Mass Spectrometry, 212*(1-3), 1–11. https://doi.org/10.1016/s1387-3806(01)00500-0

Fiedler, D., Sbeglia, G. C., Nehm, R. H., & Harms, U. (2019). How strongly does statistical reasoning influence knowledge and acceptance of evolution? *Journal of Research in Science Teaching, 56*(9), 1183–1206. https://doi.org/10.1002/tea.21547

Gilbert, G. E., & Prion, S. (2016). Making sense of methods and measurement: Lawshe's content validity index. *Clinical Simulation in Nursing, 12*(12), 530-531. https://doi.org/10.1016/j.ecns.2016.08.002

Herrmann-Abell, C. F., & DeBoer, G. E. (2017). Investigating a learning progression for energy ideas from upper elementary through high school. *Journal of Research in Science Teaching, 55*(1), 68–93. https://doi:10.1002/tea.21411

Landis, J., & Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics, 33*(1), 159. https://doi: 10.2307/2529310

Li, J., & Wang, Q. (2021). Development and validation of a rating scale for summarization as an integrated task. *Asian-Pacific Journal of Second and Foreign Language Education, 6*(1). https://doi.org/10.1186/s40862-021-00113-6

Linacre, J. M. (2011). *A user's guide to WINSTEPS/MINISTEP: Rasch-model computer programs.* Chicago, IL: Winsteps.com.

Linacre, J. M. (2012). *Many-Facet Rasch Measurement: Facets Tutorial.* Retrieved on 15th February 2023 from http://winsteps.com/tutorials.htm

Maeng, S. (2020). Investigating a learning progression for reasoning practices of geocognition using geomapapp-based assessment. *Asia-Pacific Science Education, 6*(2), 584–614. https://doi.org/10.1163/23641177-bja10009

McHugh, M. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 276-282. https://doi: 10.11613/bm.2012.031

Nedungadi, S., Paek, S. H., & Brown, C. E. (2019). Utilizing Rasch analysis to establish the psychometric properties of a concept inventory on concepts important for developing proficiency in organic reaction mechanisms. *Chemistry Teacher International.* https://doi.org/10.1515/cti-2019-0004

Ngai, C., Sevian, H., & Talanquer, V. (2014). What is this substance? What makes it different? Mapping progression in students' assumptions about chemical identity. *International Journal of Science Education, 36*(14), 2438–2461. https://doi.org/10.1080/09500693.2014.927082

Ngai, C. L. (2017). An investigation of chemical identity thinking. Graduate Doctoral Dissertations. 336. http://scholarworks.umb.edu/doctoral_dissertations/336

Ngai, C., & Sevian, H. (2016). Capturing chemical identity thinking. *Journal of Chemical Education, 94*(2), 137–148. https://doi.org/10.1021/acs.jchemed.6b00387

Ngai, C., & Sevian, H. (2018). Probing the relevance of chemical identity thinking in biochemical contexts. *CBE-Life Sciences Education, 17*(4), ar58. https://doi.org/10.1187/cbe.17-12-0271

Oon, P.-T., Spencer, B. & Kam, C.C. (2016). Psychometric quality of a student evaluation of teaching survey in higher education. *Assessment and Evaluation in Higher Education, 42*(5), pp. 788–800. https://doi.org/10.1080/02602938.2016.1193119.

Planinic, M., Boone, W., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research, 15*(2). https://doi: 10.1103/physrevphyseducres.15.020111

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi:10.1002/nur.20147

Reed, C. R., & Wolfson, A. J. (2021). Concept inventories as a complement to learning progressions. *CBE—Life Sciences Education, 20*(2). https://doi.org/10.1187/cbe.20-09-0208

Sanah, N.U., Ridwan, A., & Rahmawati, A. (2019). Analysis of chemical identity thinking through problem-based learning based on redox and electrochemistry concepts. In *Empowering Science and Mathematics for Global Competitiveness.* (pp363-368). CRS pess.

Sevian, H., & Talanquer, V. (2014). Rethinking chemistry: A learning progression on chemical thinking. *Chemistry Education Research and Practice*, 15(1), 10-23. https://doi.org/10.1039/c3rp00111c

Wei, S., Liu, X., Wang, Z., & Wang, X. (2012). Using Rasch measurement to develop a computer modeling-based instrument to assess students' conceptual understanding of matter. *Journal of Chemical Education, 89*(3), 335–345. https://doi.org/10.1021/ed100852t

Van Regenmortel, M. H. (2004). Reductionism and complexity in molecular biology. *EMBO reports, 5*(11), 1016-1020. https://doi.org/10.1038/sj.embor.7400284

Yao, J.-X., & Guo, Y.-Y. (2017). Validity evidence for a learning progression of scientific explanation. *Journal of Research in Science Teaching, 55*(2), 299–317. https://doi.org/10.1002/tea.21420

Zarkadis, N., Papageorgiou, G., & Stamovlasis, D. (2017). Studying the consistency between and within the student mental models for atomic structure. *Chemistry Education Research and Practice, 18*(4), 893–902. https://doi.org/10.1039/c7rp00135e