Muslim Education Review
Vol. 5 No. 1, 2026, 203-242
https://doi.org/10.56529/mer.v5i1.586

Systems Thinking Approach in Chemistry Education: How to Design Its Measurement Test

Asih Widi Wisudawati, Septy Nur Fadhilah
Chemistry Education Department UIN Sunan Kalijaga Yogyakarta, Indonesia
Corresponding E-mail: asih.wisudawati@uin-suka.ac.id

Abstract

The systems thinking approach in education requires analytical skills to examine complex problems and systems across multidisciplinary subjects. This approach focuses on how students analyze the components within systems and the interactions between subsystems that shape dynamic behavior. In education, systems thinking can engage multiple disciplines including science, social studies and religious studies. In chemistry education, however, this approach is recommended at the molecular level, reflecting the particulate nature of matter. While there are various systems thinking research reports in chemistry education, assessment tools to analyze the molecular basis of sustainability using this approach remain limited. This paper is aimed to design a chemistry test framework aligned with the systems thinking strands. Accordingly, this paper explains how to develop a construct map and design items, following the cyclical assessment development from Wilson's four building blocks models. This resulted in a construct map assessment design that follows the four stages of systems thinking—identifying components, understanding their interactions within a dynamic system, calculating the involvement of these components based on the molecular basis of sustainability, and encouraging students' mechanistic reasoning. According to the item design process, this paper merely reports the results of a pilot project and calculates content validity index using Aiken's V. Here, we involved two experts, two teachers and 102 students who completed the test draft. The results show that the pattern of students' answers follows the rubrics, and the test design is good. This study highlights the novel framework and its rubrics, but they need to be extended in the subsequent cycle of constructing measurement.

Keywords: *systems thinking approach, assessment tool, chemistry education*

Introduction

The systems thinking approach is a holistic way of examining complex, real-world systems, where the focus is not on individual components but on the dynamic interrelationships between them and on the patterns and behaviors that emerge from these interactions to enhance students' analytical thinking. The integration of systems thinking (ST) into chemistry education has emerged as a powerful approach to equip students with the skills necessary to address complex global challenges, particularly those related to sustainability and the United Nations' Sustainable Development Goals (SDGs) (D'eon & Silverman, 2023; Talanquer & Szozda, 2024). Here, the role of systems thinking in chemistry education research focuses on how it fosters young students as systems thinkers. Furthermore, the question arises of how to assess students who hold a systems thinker's skill set. They are recommended to have competencies such as the ability to recognize isolated components, identify some relationships, explain feedback loops, calculate their effect, predict systemic outcomes, and leverage points. Those competencies are needed to be systems thinkers who are ready to tackle complex real-world problems.

Historically, systems thinking emerged in 1960, originally called systems dynamics by engineering, which introduced the concept of feedback loops—both positive and negative—to understand how systems behave. Therefore, Donella Meadows, in her book *Thinking in systems: A primer*, provides a way to understand complex interactions and behaviors within systems (Meadows, 2008). These variables are potentially considered as expected abilities for a systems thinker. Moreover, based on the World Economic Forum's Future of Jobs Survey 2024, systems thinking is anticipated to be a core skill by 2030 (see Figure 1). Here, the transformation from engineering to the education fields offers great potential to solve complex environmental problems.

To train youth in the ability to solve complex problems, the education system is responsible for promoting its students to think holistically and analytically within a system. Therefore, once the focus is on formal school systems, the design of learning and its assessment should be prepared to achieve those skills. In terms of assessment in formal education, particularly in chemistry learning, problem-solving skills for real-world problems in a holistic way are suggested using systems thinking frameworks (MacDonald *et al.*, 2025; Mahaffy *et al.*, 2019; Szozda *et al.*, 2023; York & Orgill, 2020). This paper merely focuses on designing an assessment test for the concept. That is an acid-base concept that is relevant to the students' daily lives.

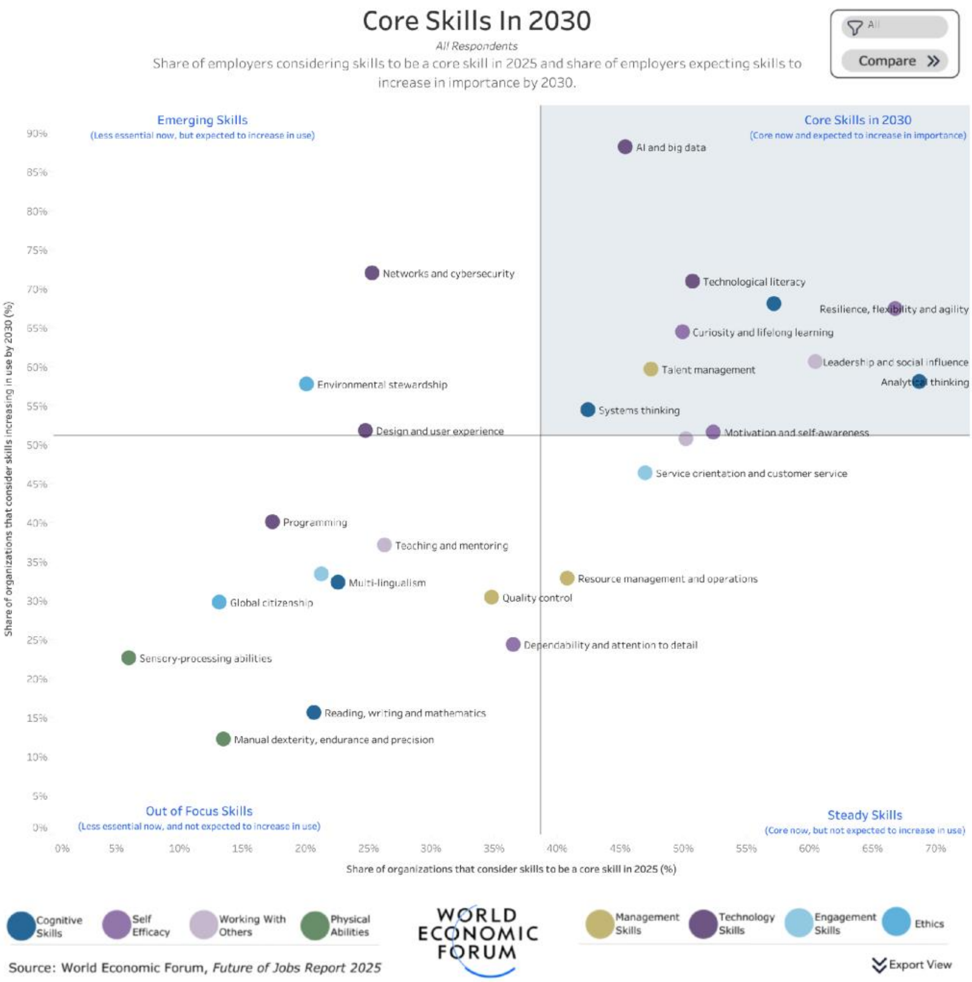
Before addressing acid-base-related problems, it is essential to understand the components within the system and their dynamic interactions within the given chemical context. The foundational element of the systems thinking skill set begins with cognitive skills. Theoretically, the systems thinking approach is an essential cognitive skill , like critical and analytical skills (York *et al.*, 2019). Adopting a systems thinking approach, further, also boosts critical thinking, collaboration, and communication skills(Sarita & Wisudawati, 2024). As shown in Figure 1, the ability to think holistically is categorized as a cognitive skill. This indicates that if young people want to solve a human health problem related to carbonic acid equilibrium, they are advised to have appropriate knowledge about the acid-base concept. It can be started from chemistry learning, and students are trained to think in the system. Hereby, via the chemistry subject, students try to investigate a real phenomenon. This discipline has specific characteristics as a subject matter in school. In a chemistry lesson, students are asked to represent a micro world or particle level for a given macro world using a specific chemical symbol language.

According to the cognitive process during learning, the chemistry concept could be figured out from three chemical representations—that is, macroscopic, sub-microscopic and symbolic. Those three representations are potentially utilized either in the learning or assessment processes. Those were launched by Johnstone and often used in chemistry education research (Johnstone, 2000). Here, macroscopic representation in the chemistry lesson means teachers are recommended to start their lesson from something that can be seen, smelled or touched by students. Including in the assessment tools, the test should contain a real-life object picture or video. Further, according to the sub-microscopic representation, it is suggested the chemistry lesson or test should stimulate students to think in the particle concept related to the given real object from the macroscopic session. The last is symbolic; here, teachers are recommended to show the chemical symbol during the learning process or in assessment tests, which should stimulate students to write the correct chemical symbols. Those three chemical representations in the lesson or assessment processes should be on a cycle and an interconnected way of thinking.

Regarding the three chemical representations mentioned above, understanding real-world phenomena fosters students to think beyond the concrete world. It requires students to start learning the particle level as a visualization of the concrete or macro world. They are stimulated to understand the whole system from the connection between the component or particulate nature of matter, which is a part of certain subsystems linked to each other in the system, holistically. However, besides the ideal chemical representation to assess the students' analytical thinking, the actual test merely emphasizes memorizing certain concepts.

The ability to think analytically in the system needs to be assessed in chemistry learning. However, after the systems thinking approach was launched by the IUPAC of the Chemistry Education Division, the assessment tools remain limited. The existing systems thinking assessment tools are related to the visualization in the form of a concept map (Aubrecht *et al.*, 2019) or Kit (MacDonald *et al.*, 2025). Apart from measuring systems thinking skill sets via concept maps, previous research has developed assessment tools to measure the students' development of systems thinking related to motivation and attitude toward chemistry (An *et al.*, 2021; Chen *et al.*, 2025). But the systems thinking measurement tools for diagnostic purposes remain limited.

Figure 1. *Map of Core Skills*



The map of core skills in 2030 shows us the significant position of systems thinking skills to face global challenges. The source of the figure came from https:// www.weforum.org/ publications/ the-future-of-jobs-report-2025/in-full/3-skills-outlook

An analytic assessment tool is necessarily designed for advanced systems thinking. Here, a diagnostic is defined as a tool to track students' cognition. This diagnostic assessment is vital for revealing students' analytical thinking. According to the PISA results 2022, Indonesian students were categorized as having lower-order thinking skills. This study designs novel cognitive tests that refer to systems thinking sequences to stimulate students' analytical thinking. Here, the chemical cognitive test is meant to measure how students use their cognition to identify chemical processes from the real complex world into a particle world that is represented by chemical symbols. Although Indonesian students often face analytical tests like in PISA, or certain HOTS tests in chemistry (Mulyani *et al.*, 2022), there are limited instruments following a systems thinking sequence. The breakthrough systems thinking instrument design can help students gain analytical thinking by identifying simple chemical components and their interactions within chemical reactions. The design of the systems thinking instrument is categorized into a four-tier cognitive test that begins with recognizing system structure, causal relationships, dynamic behavior through mathematical calculations, and mechanistic reasoning. All these tiers reflect the systems thinking sequences. Therefore, we want to address two research questions:

1) How do we design the intended measurement framework and its rubrics using the systems thinking approach?
2) To what extent can content validity be gained on experts', teachers', and students' responses in the pilot test process?

Literature Review

The measurement framework following the systems thinking approach demands specific analytical thinking strands to support the mentioned research gap. This section consists of two parts: (1) the systems thinking approach as the recommended framework in chemistry education; and (2) the integration of systems thinking principles with cognitive assessment. Both theoretical frameworks are underpinned by four building blocks (Wilson, 2005) and Aiken's V calculation (Aiken, 1985), to serve as a milestone for measurement development.

The Framework of Systems Thinking in Chemistry Education

Systems thinking is an approach that involves understanding how components interact within a system, recognizing the dynamic and interconnected nature of these interactions within the system as a part of the whole system (Schultz *et al.*, 2021; Talanquer & Szozda, 2024). In the context of chemistry education, systems thinking encourages students to view chemical processes not in isolation but as part of larger, interconnected systems that impact both the natural world and society (MacDonald *et al.*, 2022). Therefore, in the frame of chemical properties interaction within a system, chemical systems contain their components, called molecular properties. Every chemical process engages the molecular properties, which are named in the systems thinking approach as the molecular basis of sustainability (Mahaffy *et al.*, 2019). The example is carbonate ions, hydronium ions, and carbon dioxide molecules as the molecular basis of sustainability for ocean acidification (Wisudawati & Barke, 2024), and ammonia ions as the molecular basis of sustainability for the Haber-Bosch process (Mahaffy *et al.*, 2019). Despite the mentioned molecular basis of sustainability, the implementation of the systems thinking approach is also conducted by designing sustainable concrete in guided inquiry experiments for schools (Delaney *et al.*, 2022). These mentioned examples are aimed at fostering analytical thinking by understanding the molecular level, which is connected to a complex phenomenon in the environment.

Our environment faces tremendous global challenges caused by the climate crisis, ocean acidification, resource depletion, and environmental degradation, necessitating the development of systems thinkers who can approach problems from a holistic perspective

(Talanquer & Szozda, 2024). Chemistry, as a discipline, plays a critical role in addressing these challenges, making it essential to integrate systems thinking into chemistry education (Aubrecht *et al.*, 2019; D'eon & Silverman, 2023).

The systems thinking framework had been launched through the Systems Thinking in Chemistry for Sustainability: 2030 and Beyond (STCS 2030+) project (Mahaffy, n.d.; Talanquer & Szozda, 2024). This framework provides educators in chemistry with resources and tools to plan, implement and assess chemistry units using a systems thinking approach. According to the assessment using systems thinking, there are the SOCME or Systems-Oriented Concept Map Extension or SOCkit or SOCMEs Online Construction Kit (Aubrecht *et al.*, 2019; MacDonald *et al.*, 2025; Reynders *et al.*, 2023). Via SOCME or SOCKit, students could effectively build their knowledge through identifying the correlation between the subsystem and its influences. However, administering SOCME and SOCKit in the evaluation process poses a challenge in scoring the students' achievement. Furthermore, another systems thinking tool, like CheMIST, remains focused on the framework without the testing tools. CheMIST is designed to facilitate the integration of the principles of systems thinking into chemistry education through the essential principles (York & Orgill, 2020).

The systems thinking approach in this study is based on scaffolding cognitive sequences following the essential principles from the CheMIST sequences. Here, the type of assessment is aimed at diagnosing students' analytical thinking regarding daily acid-base phenomena. Typically, diagnostic assessment tools are similar to misconceptions diagnostic tools in four tiers (Jumilah & Wasis, 2023; Prodjosantoso *et al.*, 2019). Therefore, cognitive assessment aimed at revealing analytical skills can potentially be integrated with systems thinking in chemistry education. Here, analytical skills are essential for encompassing the cognitive component, including the ability to produce alternative solutions, step-by-step conceptualization, and cause-and-effect analysis (Eskin, 2013).

Integration Between Systems Thinking Principles and Cognitive Assessment

Integrating both systems thinking principles and cognitive assessment promotes a novel assessment tool pattern in chemistry education. Three different attributes need to be revealed based on the existing theory and previous research so the position of every variable in this study would be clear and precise. Those variables are systems thinking principles, cognitive taxonomy, and assessment tools. The systems thinking principles in this study are articulated from the terminology of essential characteristics (York & Orgill, 2020), transcribed from systems thinking skills and attitudes (Reynders *et al.*, 2023) and major attributes of a system and associated systems thinking skills (Talanquer & Szozda, 2024).

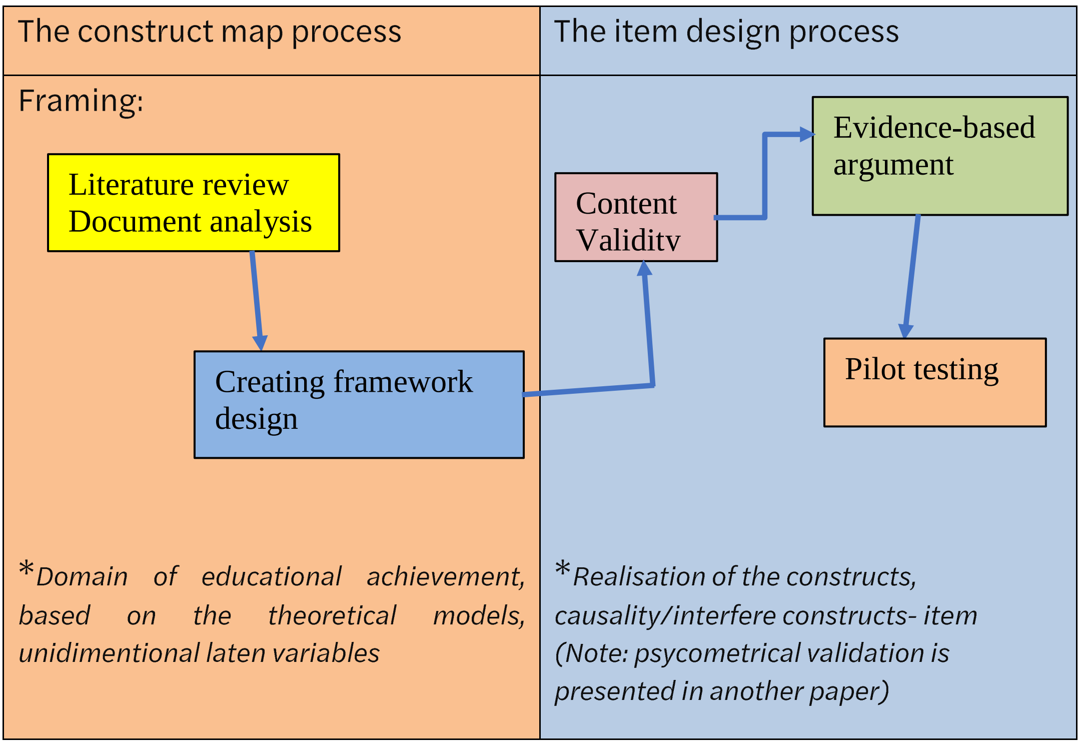
First, as described above, systems thinking is a cognitive approach to understanding complex real-world problems (Reynders *et al.*, 2025), which is defined by the appearance of components, interconnections and purposes (Arnold & Wade, 2015). Furthermore, York and Orgill (2019) defined the operation of systems thinking skills in terms of five principles: (1) systems as a whole, (2) relationships between parts, (3) causal variables, (4) behavior over time, and (5) interaction with the environment. Those five attributes were simplified into (1) system composition, (2) system structure, (3) system behavior, and (4) system effects (Talanquer & Szozda, 2024). According to the implementation of the mentioned systems thinking principles, there are two possibilities for the designs: either the learning environment or the assessment tool. This study focuses on designing an assessment tool following the stages in the four building blocks of measurement (Wilson, 2005) that is limited to the two early stages. Therefore, according to the systems thinking cognitive skill set, the assessment framework will be designed based on the Bloom's taxonomy.

Regarding the sequences of test items, this study utilizes Bloom's taxonomy, where the first tier is designed for memorizing concepts, the second tier for understanding concepts, the third for implementing concepts, and the fourth for analyzing concepts. The main reason for using Bloom's taxonomy is to adjust the Indonesian curriculum for lesson plans and assessments. Moreover, the implementation of systems thinking principles remains open to adapting the local curriculum. So, this is a promising opportunity, particularly for scaffolding assessment tools. Indonesian teachers find it challenging to develop and apply the ideal instruments, complete with their rubrics. However, the study reported that the use of rubrics in assessing systems thinking skills is effective in both formative and low-stakes assessment contexts(Szozda *et al.*, 2023).These rubrics can help identify areas where students may need additional support in developing their systems thinking cognitive skill sets, while also providing a means of tracking progress over time. The essential attribute in the rubrics involves identifying the composition and chemical structure, as well as analyzing the interaction between particles and their behavior. This is relevant to Bloom's taxonomy, especially for the fourth level of analytical skills (Bloom, 1986). The complexity of test items for each stimulation text is considered in terms of understanding or comprehension, application, and analysis. These four domains focus on cognitive processes, which are integrated into analytical thinking through the systems thinking approach in chemistry education.

Method

The development of an assessment tool, especially a cognitive test, requires a rigorous method to ensure the test's credibility. This paper utilizes Wilson's four building blocks for the measurement framework, which contains a construct map and item design (see Figure 2), but the outcome space and measurement model are depicted in another paper. According to the framework design, this paper claims that need to be thoroughly explained to create a robust assessment design. First, the construct map defines the cognitive construct and its rubrics that use qualitative methods like literature review, expert interviews, and document analysis. Those help to identify the dimensions of cognition and the level of systems thinking abilities. Second is the item designed to gain content validity after the first test draft is ready. This stage is aimed at getting an evidence-based argument (Yu *et al.*, 2026) by a quantitative approach from involved experts, teachers and students. Therefore, mixed methods research is chosen to address the research questions. Here, the mixed methods research design is part of the primary design (Creswell, 2014).

Figure 2: *The Starting Building Blocks for Designing the Test Instrument*



This study utilizes the mixed methods approach to adhere to the four building blocks of measurement (Wilson, 2005). Two stages are depicted in the following short design:

1. *The construct maps*: A qualitative approach is applied through theoretical reviews to create a definition for each systems thinking ability. The significant point in developing the assessment tools in this study is to develop a novel framework. This framework articulates a systems thinking approach as a part of analytical cognitive skills in the chemistry content. This study limits the systems thinking framework: recognizing systems as a whole, not just a collection of parts; examining the relationship between components in the system; identifying variables which cause system behavior via mathematical calculation for existing particles; and examining how system behavior changes over time via revealing mechanistic reasoning (York & Orgill, 2020). Translation of those mentioned systems thinking abilities demands the specific mental model that is embedded in the test items.
2. *Item design*: The main activities were continued by item writing that was prepared to be readable by students. Before the advanced psychometric acceptance for validity and reliability, the quality control in this study was started from the calculation of content validity from expert judgment and users. It was aimed to ensure the quality of instruments from experts' and users' perspectives. Here, consideration of the experts' background had been identified carefully. This is an important stage to ensure the quality of the instrument test that has never been developed before. Regarding the quantitative approach, a questionnaire was used to reveal experts', teachers' and students' responses. In this approach, quantitative data were calculated to achieve content validity based on Aiken's V calculation (Aiken, 1985). Data gathering was conducted in the acid-base session for X grade in Senior High School, which is taught at the end of the second semester of

2024/2025. This study adheres to the iterative cycle for developing measurement, leveraging the cycle of the four building blocks (Wilson, 2023).

Research Ethics in the Preparation Stage

Besides the structure of assessment tools, the identification of participants engaged in this study was considered from two perspectives. First is the category of participants, and second is about research ethics. According to the ethics of research with human subjects, this study put trust as the foundation to protect the participants' privacy and promote research integrity (Resnik, 2024).

This research ethics was ensured by obtaining participants' agreement. Oral permission was obtained from experts, while a formal letter was used for participants from schools, including teachers and students. Before sending the letter, we considered the category of schools: at the highest, moderate and lowest levels of cognitive achievement. However, due to the rejection of the permission letter from one of the schools, this study only presents results from the high and moderate schools. There is no manipulation or intervention involved with participants. We used purposive sampling techniques guided by teachers in the target school's category. Here, the main research subject is the developed assessment tool, and participants engaged have their role to make responses.

The Participants

The participants in this study are defined as individuals involved in the item design stage to calculate content validity. The participants involved are chosen based on their role in developing the measurement instrument. They contributed to various tasks: two experts evaluated the content and construction of test items based on the intended indicators;

two teachers helped assess how the assessment could be practically implemented in schools. Then, a third participant, students, provided their responses on the assessment tools. The number of students involved was one hundred and two (102), drawn from two different school levels—the first school is categorized as a high-level school, and the second school is categorized as a moderate-level school. School level is based on their achievement of the summative evaluation from the Special Region of Yogyakarta city education office. According to the eligibility criteria for the experts reviewing the assessment tools, they hold at least a master's degree in development evaluation and/or a master's in chemistry content. The teachers are classified as senior teachers. This study included students who had already learned the acid-base concept and were in the eleventh grade of senior high school. Data collection from participants mentioned above was finished in two months, from April to May 2025, in the Special Region of Yogyakarta (DIY) city.

Research Design

The research design is divided into two approaches. According to the exploratory sequential design, the first stage is qualitative to scaffold the construct map containing the main framework, and the second is quantitative descriptive research to calculate the content validity limited by Aiken's V. Further, the subsequent stage, the reinforcement phase, was conducted by a qualitative approach to underline the revised instrument. Here, this study did not manipulate the subject of research. The research focus is on developing an instrument test, so participants are asked to give their responses to the designed questionnaire. The principle of measurement used here is an evidence-based validity argument, whether coming from either the construct map stage or the item design stage.

Designing the assessment tools (Construct map) is the first of the building blocks (Wilson, 2023), which depicts the systems thinking approach with some adaptation. The additional competencies in the design instrument framework are mathematics skills, which are articulated in the third question. The designed test incorporates four different tiers for each systems thinking principle. The questions are organized from the first to the fourth competencies, which refer to the main systems thinking foundations, gradually. The gradual cognitive thinking is intended to stimulate students to think from the first systems thinking ability (ST-1) to ST-2 and ST-3, as shown in Table 1.

Table 1: *Construction of an Instrument for integrating Systems Thinking and Bloom's Taxonomy*

Integration between ST and Bloom's Taxonomy (Code)	Construction of instruments (Code)
Identifying component of the systems and memorizing level. (ST-1 and C1)	Chemical knowledge about the mentioned particle from the given picture as test stimulation: 1. chemical formula 2. the properties of matter 3. the chemical phase changing, 4. the involved particles in the chemical reactions. (ST-B 1)
Interaction between components and analysis level (ST-2 and C2 & C4)	Students' ability to predict the interaction between components in meaningful ways: 1. the involved particles in the form of chemical reactions, 2. the chemical phase changing for mentioned chemical reactions 3. The effects of interaction between particles 4. The pH graph changing interpretation. (ST-B 2)
	Students' communication skills for delivering their reasons for explaining the interaction between particles in the systems: 1. students' prediction about the effects of interactions 2. students' ability to think beyond the number (ST-B 4)
Identifying variables which cause system behavior via mathematical calculation for existing particles and implementation level (ST-3 and C3)	Students' ability to calculate the involved particles in the certain chemical reactions. This item measures the ability for analyzing cause and effect of interaction between particles in the limited systems and further broad systems. (ST-B 3)

A quantitative approach (item design validated using Aiken's V for content validity) was conducted by calculating the content validity index from the results of the completed questionnaire by two experts. Here, the questionnaire uses the Likert rating scale average for three main indicators. The following formula is used to calculate the content validity (V) coefficient based on Aiken's V(Aiken, 1985; An Nabil *et al.*, 2022).

$$V = \frac{S}{(n(c-1))} \quad \text{(eq.1)}$$

$$S = r - Lo \quad \text{(eq.2)}$$

Information:

V: content validity coefficient, or called the Aiken index

S: r-lo if Lo < Hi and lo-r if Hi< Lo (Hi is the higher validity category)

r: the rater's validity rating of the item

Lo: lowest validity category (1)

c: maximum scale of integers (3)

n: number of raters

According to a quantitative description, the given categories in the instrument have already been set up following the standard for developing test instruments. These categories are assigned to expert validators, teachers and students in different language versions. The categories for experts include three aspects: the content of subject matter, test construction, and language. The three aspects mentioned earlier are similar to the questionnaire for teachers, but researchers structured them into different statements. Conversely, the students' questionnaire was created with a simple statement. All participants completed their questionnaires after reading the entire test, and they were asked to answer them. The main point in this stage was collecting arguments from the involved research team to provide the scientific argument in terms of continuing the subsequent instrument design.

The standards of aspects guide the content validity of instruments. Table 2 depicts the indicators to ensure the quality of the instrument. Still, participants are assigned to fill out the questionnaire with different indicators and aspects. Experts in evaluation give feedback for the construction of test items; a chemistry expert assesses the content of chemistry; and teachers are assigned to respond to the selected three aspects. Apart from the previous participants, students give their responses for simple indicators, which are depicted in Table 3.

A quantitative approach was employed after participants had filled out the questionnaire. This approach was aimed at ensuring that the developed test items are relevant to the intended indicators and reach a saturated agreement for content and construct validity. While developing instruments based on a theoretical framework as part of a qualitative approach, content validity requires special qualitative methods that involve expert judgment or argument. Through the panel approach, content validity is assessed by professional expertise to review the scale (Beck & Gable, 2001) and define the domain of knowledge (Crocker, 2015).

Table 2: *The Indicators of the Quality of Instruments*

Aspects	Indicators of the quality of test items
Content of chemistry	• Suitability of questions to the curriculum used, • The questions presented follow the indicators, • The depth of the material presented in the stimulus, • The chemical concepts presented follow the theory, • Material/substance according to grade level and class level, • Suitability of material/substance to measurement objectives (diagnostic tests), • The questions are presented for the acid-base calculation, • Conformity between the stimulus and the questions presented, • Accuracy of use of notation/ chemical symbols, • The accuracy of sentence structure according to the rules.
Test item construction	• Declarative sentences used in logical questions, • The stimulus, image or chemical symbol in the question is clear and functional, • Presentation of questions follows the characteristics of HOTS questions, • Presentation of questions follows the characteristics of diagnostic instruments, • The presentation of questions follows the characteristics of the systems thinking approach.

Language aspect	• The language used in the questions is easy for students to understand, • The language used in the questions complies with the General Guidelines for Indonesian Spelling (PUEBI), • Instructions for completing the questions are clearly stated and easy to follow. • The allotted time is appropriate for the number of test items (questions), • The font is suitable and comfortable to read, • The font size is appropriate and easy to read, • Stimuli/discourses, tables, and images are presented clearly and attractively, • The instructions within the questions are easy to understand, • The questions include real-life problems encountered daily, • The questions are designed to challenge students to solve them.

Table 3: *Indicator for Students' Responses (SR)*

Code	The questionnaire indicators
SR-1	Instructions for working on the questions are stated clearly and are easy to understand.
SR-2	The time provided is according to the number of test items (questions) given.
SR-3	The font is appropriate and comfortable to read
SR-4	The font size is appropriate and comfortable to read
SR-5	Stimulus/discourse, tables and images are presented clearly and attractively.
SR-6	The instructions given in the questions are easy to understand
SR-7	The questions use problems encountered in everyday life
SR-8	The questions presented challenge students to work on them

Results and Discussion

A. The Designed Assessment

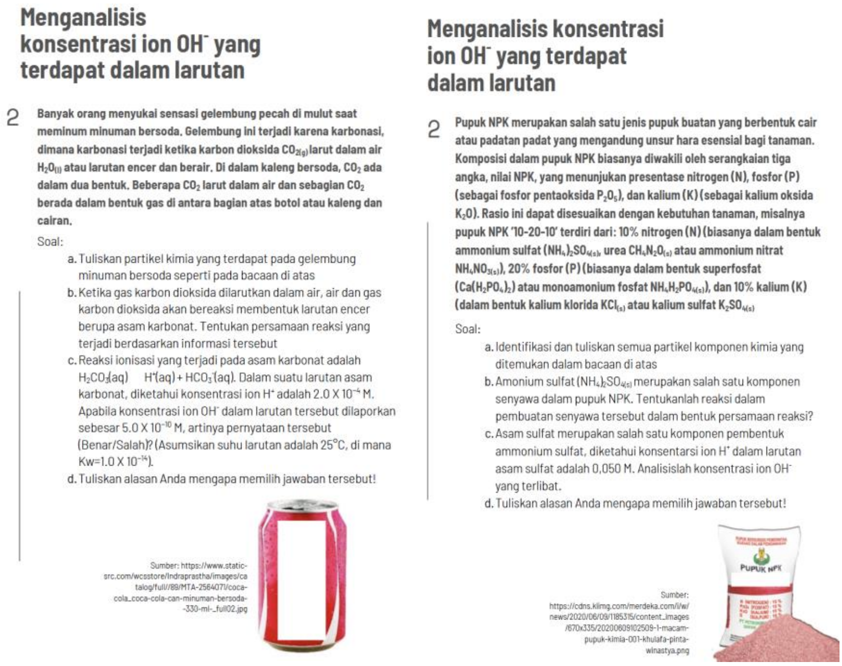
The developed instrument follows a systems thinking approach to enhance analytical thinking skills, beginning from the molecular basis of sustainability (Mahaffy *et al.*, 2019). Here, the assessment tool should engage the real phenomena to stimulate students to think analytically. The structure of intended tests is aimed at implementing the latticework in Table 1. For instance, Figure 3 shows the translated test about carbonized beverage, whereas Figure 4 depicts the real test for the fertilizer context. Here, the text and figure are set as stimulation and followed by gradual questions from numbers 1 to 4. The first question is intended to recall or memorize which particles have been found during reading stimulation. Students can repeat reading and searching. The second question demands higher thinking than the first question. Students should think beyond the text and figure and have a mental model to imagine the involved particles in the systems. Here, students are asked to write the chemical reactions in the correct symbols. While the fourth one asks the reason for the given answer, the third question reveals numerical skills which need more analytical thinking compared to the second question. As such, the third question in Figure 3 depicts the concentration of either H^+ ions or OH^- ions in the solutions, yet the intended answer is true or false based on both displayed concentrations. Students must analyze the system's loop or cause and effect within the carbonized beverage. According to the fourth question, students are asked to make a chain of concept understanding that they hold from the first question to the third one. Their understanding would be categorized into holistic or analytical thinker, or fragmented thinker, or neither.

This study produces twenty test packages that are divided into A and B packages. As depicted in Figure 3, the novel breakthrough of our instruments is that the questions are arranged from understanding to analytical skills. Yet, the scoring for each question is designed equally. The score for each question for a correct answer is 1, whereas an incorrect answer is 0. Once their answer shows an inconsistency pattern like correct, incorrect, incorrect and correct for the first to fourth questions, those are meant to show that students hold fragmented knowledge. Table 4 shows the rubrics that stimulate gradual analytical thinking ability using the designed test.

Table 4: *The Rubrics of the Created Test Design that Integrate the Systems Thinking Approach and the Cognitive Taxonomy by Bloom*

	The designed code				Categories
	ST-B 1	ST-B 2	ST-B 3	ST-B 4	
Pattern	C	C	C	C	holistic and analytical thinker
	C	C	C	I	less analytical thinker
	C	C	I	I	lack analytical thinker
	C	I	I	I	merely memorizing
	I	C	C	C	fragmented thinker
	I	I	C	C	fragmented thinker
	I	I	I	C	fragmented thinker
	I	I	I	I	no ability

Figure 3: *The Displayed Test in the A Package Stimulates Students to Think About (a) the Carbonized Beverage Systems and (b) the Fertilizer Systems*



Menganalisis konsentrasi ion OH^- yang terdapat dalam larutan

2 **Banyak orang menyukai sensasi gelembung pecah di mulut saat meminum minuman bersoda. Gelembung ini terjadi karena karbonasi, dimana karbonasi terjadi ketika karbon dioksida $CO_{2(g)}$ larut dalam air $H_2O_{(l)}$ atau larutan encer dan berair. Di dalam kaleng bersoda, CO_2 ada dalam dua bentuk. Beberapa CO_2 larut dalam air dan sebagian CO_2 berada dalam bentuk gas di antara bagian atas botol atau kaleng dan cairan.**

Soal:

a. Tuliskan partikel kimia yang terdapat pada gelembung minuman bersoda seperti pada bacaan di atas
b. Ketika gas karbon dioksida dilarutkan dalam air, air dan gas karbon dioksida akan bereaksi membentuk larutan encer berupa asam karbonat. Tentukan persamaan reaksi yang terjadi berdasarkan informasi tersebut
c. Reaksi ionisasi yang terjadi pada asam karbonat adalah $H_2CO_3(aq)$ $H^+(aq) + HCO_3^-(aq)$. Dalam suatu larutan asam karbonat, diketahui konsentrasi ion H^+ adalah 2.0×10^{-4} M. Apabila konsentrasi ion OH^- dalam larutan tersebut dilaporkan sebesar 5.0×10^{-10} M, artinya pernyataan tersebut (Benar/Salah)? (Asumsikan suhu larutan adalah 25°C, di mana $Kw=1.0 \times 10^{-14}$).
d. Tuliskan alasan Anda mengapa memilih jawaban tersebut!

Sumber: https://www.static-src.com/wcsstore/Indraprastha/images/catalog/full//89/MTA-2564071/coca-cola_coca-cola-can-minuman-bersoda--330-ml-_full02.jpg

Menganalisis konsentrasi ion OH^- yang terdapat dalam larutan

2 **Pupuk NPK merupakan salah satu jenis pupuk buatan yang berbentuk cair atau padatan padat yang mengandung unsur hara esensial bagi tanaman. Komposisi dalam pupuk NPK biasanya diwakili oleh serangkaian tiga angka, nilai NPK, yang menunjukan presentase nitrogen (N), fosfor (P) (sebagai fosfor pentaoksida P_2O_5), dan kalium (K) (sebagai kalium oksida K_2O). Rasio ini dapat disesuaikan dengan kebutuhan tanaman, misalnya pupuk NPK '10-20-10' terdiri dari: 10% nitrogen (N) (biasanya dalam bentuk ammonium sulfat $(NH_4)_2SO_{4(s)}$, urea $CH_4N_2O_{(s)}$ atau ammonium nitrat $NH_4NO_{3(s)}$), 20% fosfor (P) (biasanya dalam bentuk superfosfat ($Ca(H_2PO_4)_2$) atau monoamonium fosfat $NH_4H_2PO_{4(s)}$), dan 10% kalium (K) (dalam bentuk kalium klorida $KCl_{(s)}$ atau kalium sulfat $K_2SO_{4(s)}$**

Soal:

a. Identifikasi dan tuliskan semua partikel komponen kimia yang ditemukan dalam bacaan di atas
b. Amonium sulfat $(NH_4)_2SO_{4(s)}$ merupakan salah satu komponen senyawa dalam pupuk NPK. Tentukanlah reaksi dalam pembuatan senyawa tersebut dalam bentuk persamaan reaksi?
c. Asam sulfat merupakan salah satu komponen pembentuk ammonium sulfat, diketahui konsentarsi ion H^+ dalam larutan asam sulfat adalah 0,050 M. Analisislah konsentrasi ion OH^- yang terlibat.
d. Tuliskan alasan Anda mengapa memilih jawaban tersebut!

Sumber: https://cdns.klimg.com/merdeka.com/i/w/news/2020/06/09/1185315/content_images/670x335/20200609102509-1-macam-pupuk-kimia-001-khulafa-pinta-winastya.png

(a) (b)

Besides the design of the test framework and rubrics, we also collected the evidence-based argument from the created indicators. Here, quantitative description calculation is limited in assessing content validity from Aiken's V. Table 4 details the calculation of the Likert scale in the questionnaire, which was conducted by two experts. After designing and validating the content and construct related to cognitive achievement to think on the system, the piloting project was conducted with the implementation of test items to 102 students. The pilot result before analyzing the psychometric is depicted in Figure 6 for the A package and Figure 7 for the B package.

Table 4: *Aiken's V distribution*

Experts	S1	S2	Ss	V

Test items	1	2				n(c-1)	
1	2,9	3	1,9	2	3,9	4	0,97
2	2,8	2,8	1,8	1,8	3,6	4	0,89
3	2,8	3	1,8	2	3,8	4	0,94
4	2,8	3	1,8	2	3,8	4	0,94
5	3	2,9	2	1,9	3,9	4	0,97
6	2,9	2,9	1,9	1,9	3,8	4	0,94
7	3	2,8	2	1,8	3,8	4	0,94
8	2,9	2,8	1,9	1,8	3,7	4	0,92
9	2,9	2,6	1,9	1,6	3,4	4	0,86
10	2,8	2,9	1,8	1,9	3,7	4	0,92
11	3	2,6	2	1,6	3,6	4	0,89
12	3	2,6	2	1,6	3,6	4	0,89
13	3	2,3	2	1,3	3,3	4	0,83
14	2,9	2,4	1,9	1,4	3,3	4	0,83
15	3	2,6	2	1,6	3,6	4	0,89
16	2,9	2,6	1,9	1,6	3,4	4	0,86
17	2,9	3	1,9	2	3,9	4	0,97
18	2,9	3	1,9	2	3,9	4	0,97
19	2,9	3	1,9	2	3,9	4	0,97
20	2,8	2,6	1,8	1,6	3,3	4	0,83

This study focuses on the content validity to support the designed instruments. Aiken's V coefficient is aimed to measure the content validity, which ranges in value from 0 to 1 (Aiken, 1985). The percentage of similar

V values is depicted in Figure 5. Despite the expert suggestion, this study also collects data from teachers and students.

Figure 5: *Aiken's V distribution in Percentages*

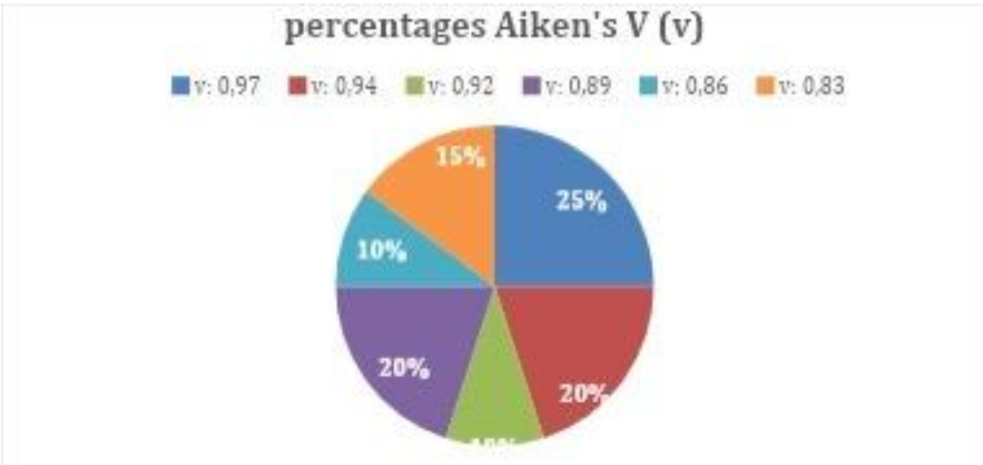


Figure 6: *The Pattern of Students' Correct Answers in the Percentage that Shows Their Ability to Show Consistency of Analytical Thinking from the A Package*

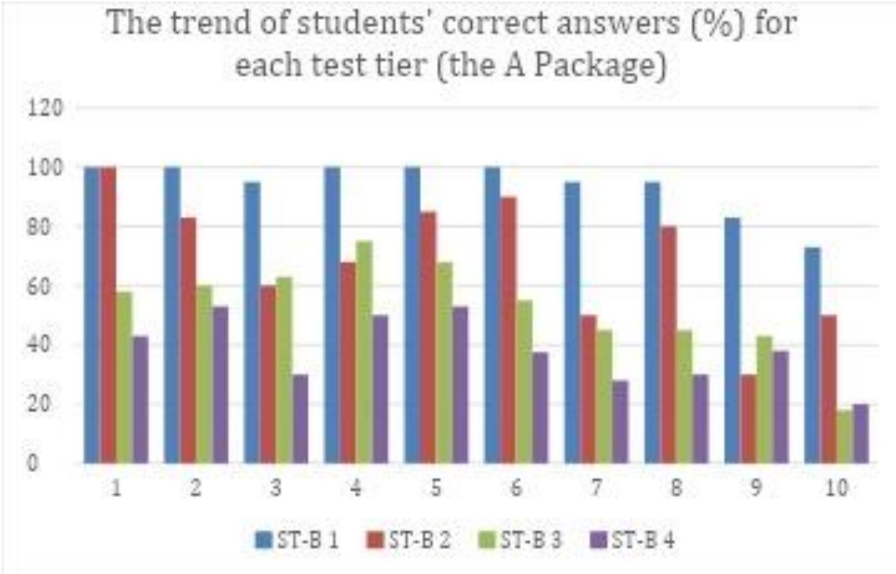
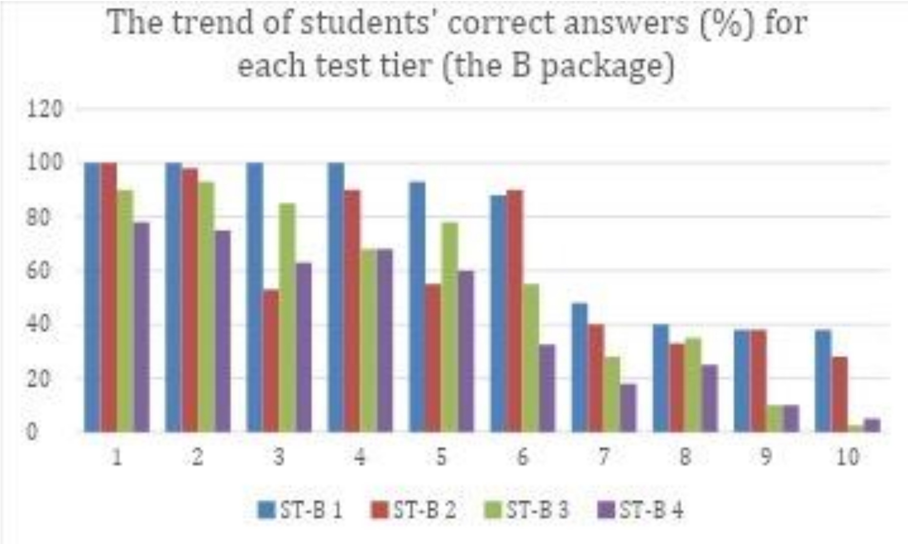


Figure 7: *The Pattern of Students' Correct Answers in the Percentage that Shows Their Ability to Show Consistency of Analytical Thinking from the B Package*



B. Discussion

Assessment for the systems thinking approach in the chemistry education division remains limited. This study promotes the novel design of an instrument which involves the cognitive domain. Regarding the cognitive domain, some well-known models or taxonomies describe how the human brain works. The most frequently used cognitive taxonomy in Indonesia's education system is Bloom's taxonomy. When test developers refer to this framework, test items are typically constructed from levels such as memorizing, understanding, applying, analyzing, evaluating and creating (Bloom, 1986). Bloom further revised the taxonomy. Regarding how cognition works, these taxonomies could potentially be connected to the systems thinking approach for chemistry education. This study primarily explores ways to motivate students to think analytically about systems, especially when they are engaged with well-known environmental issues. Through the systems thinking approach, students are expected to develop analytical thinking towards the given issues. Therefore, with the help of the sequence of cognitive abilities outlined in Bloom's taxonomy, the study addresses the following two research questions.

Research Question 1: The construct map for analytical thinking via systems thinking sequences

The correlation between systems thinking and analytical skills is significant, as evidenced by several studies that highlight their interrelatedness in education and the job environment, as mentioned in the research background. Systems thinking in chemistry education fosters a holistic understanding of complex systems, enabling students to analyze the system's composition, structure, behavior and effects (Talanquer & Szozda, 2024). These analytical skills would allow students to discern relationships and patterns within data. However, few test items were designed to promote practical use of the cognitive taxonomy. This paper offers the bridging of systems' structure as part of the systems thinking attribute with the frequently used Bloom's taxonomy. This relationship is connected to Bloom's taxonomy, which stimulates the higher-order thinking skills in its higher taxonomy. While Table 1 presents a conceptual chemistry design, Table 5 depicts the combination of systems thinking skills and Bloom's taxonomy. However, in the pilot testing, the top of analytical ability in the fourth question seems difficult to achieve for either the A or the B packages. It proves that students alike have a good ability to recall certain concepts, yet, once they are asked to connect their concept to another concept, they stay with fragmented knowledge. Here, thinking in systems should promote students to avoid reductionist perspectives (Orgill *et al.*, 2019).

Table 5: *The Developed Construct Map for Integrating the Systems Thinking Attribute (Talanquer & Szozda, 2024) and Bloom's Taxonomy*

No.	Attribute of systems thinking	The designed competencies with adaptation from CheMIST and Indonesian curriculum	Bloom's taxonomy
1	System composition	Identify system components, characterizing the properties in the molecular level.	Understanding and comprehension
2	System structure	Explore and identify interaction between system components—limited in the chemical reactions	Application and analysis
3	System behavior	Infer the interaction between components from the calculating the involved particles in the system	Application and analysis
4	System effects	Identify interaction of the system with its environment—through the mechanistic reasoning	Application and analysis

The design of aligning systems thinking with Bloom's taxonomy is applied to a particular concept, that is, the acid-base concept. The main reason for choosing this concept was this concept has numerous contextual properties to foster students to think about systems. Here, systems thinking encourages a shift from a linear to a systemic approach, as demonstrated by analytical thinking. Research found that thinking in systems through the Chemistry Systemic Learning Approach (CSLA) improved students' analytical skills (Fitriyana *et al.*, 2019). Another study also showed that integrating systems thinking into project-based learning allows students to engage in hands-on experiences, promoting systems thinking and analytical skills simultaneously (Shekh-Abed & Barakat, 2022). Regarding the assessment settings of this study, the analytical skills are assessed through the stimulus text to achieve the following indicators:

a) Analyzing the H^+ ions in the solutions,

b) Analyzing the OH^- ions in the solutions,

c) Analyzing the H^+ ions and the OH^- ions in the solutions,

d) Analyzing the properties of substances from their pH level,

e) Correlating the pKb or weak base chemical equilibrium constant with pKw or the equilibrium constant for autoionization of water,

f) Proving the base constant (Kb) from specific chemical solutions,

g) Correlating volume and concentration of solution in the dissolving process

h) Predicting the involved concentration of the solution from the titration data,

i) Summary of the pH values from the titration data, and

j) Analyzing the ionization degree for the given sample.

The ten mentioned indicators have different stimulus texts. Those competencies were articulated into a ten-stimulus test completed with four-tier questions referring to systems thinking grade. The given stimulus text aims to promote students to think analytically. Through the real chemicals used in daily life, testing refers to the systems thinking sequences, emphasizing the chemical solutions.

The test was developed in two versions, each containing ten test items, for a total of twenty test items. Each test item has four questions that refer to the systems thinking indicator in Table 6. The four questions are prompts designed to start by identifying the particles involved in the stimulation text, recognizing the interaction between particles in relation to the intended chemical reactions, calculating the number of particles associated with the indicators mentioned earlier, and explaining the reasoning behind analytical thinking.

Table 6: *Designed Competencies and the General Test Design*

The designed competencies with adaptation from CheMIST and Indonesian curriculum (York & Orgill, 2020)	**The designed general stems for every stimulation text**
Identify system components, characterizing the properties in the molecular level. *(competency number 1/C-1)*	Please identify and write the involved particles based on the stimulation text! *(question 1)*
Explore and identify interaction between system components—limitted in the chemical reactions. *(competency number 2/ C-2)*	Predict the chemical reactions based on the text, then analyze the interaction between particles from your mentioned reactions and write the net ionic reactions! *(question 2)*
Infer the interaction between components from the calculating the involved particles in the system. *(competency number 3/ C-3)*	If one particle has some mass, then calculating concentration another particle involved in the chemical reaction! *(question 3)*

Identify interaction of the system with its environment—through the mechanistic reasoning *(competency number 4/ C-4)*	Please explain the reasoning of your three previous mentioned answer above! *(question 4)*

Research Question 2: The resulting Aiken's V for content validity from experts', teachers' and students' perspectives

The subsequent stages following Wilson's four building blocks are the item test. The combination of Tables 1, 7 and 8 was translated into an item test. The intended test item design was conducted by breaking down the framework into questions. Still, following the intended indicators of measurement, which are structured according to systems thinking sequences. Subsequently, the two equal test types were produced by every type having ten packages. The whole test package represents the framework, but a review is needed to ensure the content and construct validity. Here, an evidence-based argument is significant to produce a good quality instrument. Indeed, the respondent, measurer and test developer gain similar perspectives.

Content validity is one prerequisite process in the development of an achievement test. This study is only limited by the Likert scale, which is fulfilled by experts, teachers and students. The resulting data type is an ordinal scale, which is allowed to be calculated by Aiken's V or validity based on Aiken's formulas. A variety of situations where judgments of the content validity of items or questionnaires are made on ordinal rating scales (Aiken, 1980). Subsequently, tables 3 and 4 depict the number of Aikens' V based on experts' and teachers' responses. According to the dominant Aiken's V is around 0,97, the content validity based on the academia's perspective indicates a very high level. Additionally, teacher participants also grant the high values of V, with the average of two participants reaching 0,95. Both values embrace three different aspects of measurement validation adapted from the Indonesian Centre of Curriculum. Those are the content of chemistry, test construction, and the language aspect. The students' responses show a different pattern compared to the experts' and teachers' responses. According to the moderate result, students predict that the time or duration for answering a question doesn't fit with the setting. Consequently, they give a moderate score for the indicator; the time provided is according to the number of test items (questions) given. Interestingly, they give an excellent score for the layout, including a font that is appropriate and comfortable to read.

Conclusion

A systems thinking approach for chemistry education has been explored through the development of a measurement framework. This study is part of a comprehensive measurement scaffolding, which is reported only for the initial steps, as outlined in the four building blocks by Wilson. The results of this study suggest that systems thinking sequences could potentially be stimulated by a set of subsystem phenomena. Furthermore, students are expected to examine how the subsystems change. Through our framework and its rubrics, chemistry teachers could diagnose students' analytical thinking about certain real systems. Further, students scaffold their analytical thinking similar to high-order thinking skills. The four tiers of questions are aimed at adjusting the systems thinking sequences. Twenty packages of questionnaires were developed based on the expected indicators. This measurement instrument was assessed by participants and obtained a valid scale ranging from 0,83 to 0,97, Aiken's V index, or construct validity calculated by Aiken's formula. This scale range was derived from the responses of experts and teachers. In contrast, students perceive that research should consider the duration needed to answer or solve the questionnaire.

This study is part of the whole measurement design that is divided into two papers. In this regard, this paper is set up to describe how we construct a framework or measurement map and only calculates content validity using the classical method of content validity assessment. Future work will complete this research with modern calculation methods and scientific definitions. Moreover, this study is still underway in fieldwork by the first author's team and will be extended to incorporate the next of Wilson's four building blocks. A limitation of this paper is the practical issues in diagnosing the latent cognitive ability beyond their answers. It requires modifying the form of the assessment test.

References

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, *45*(1), 131–142. https://doi.org/10.1177/0013164485451012

An, J., Loppnow, G. R., & Holme, T. A. (2021). Measuring the impact of incorporating systems thinking into general chemistry on affective components of student learning. *Canadian Journal of Chemistry*, *99*(8), 698–705. https://doi.org/10.1139/cjc-2020-0218

An Nabil, N. R., Wulandari, I., Yamtinah, S., Ariani, S. R. D., & Ulfa, M. (2022). Analisis Indeks Aiken untuk mengetahui validitas isi instrumen asesmen kompetensi minimum berbasis konteks sains kimia. *PAEDAGOGIA*, *25*(2), 184. https://doi.org/10.20961/paedagogia.v25i2.64566

Arnold, R. D., & Wade, J. P. (2015). A definition of systems thinking: A systems approach. *Procedia Computer Science*, *44*, 669–678. https://doi.org/10.1016/j.procs.2015.03.050

Aubrecht, K. B., Dori, Y. J., Holme, T. A., Lavi, R., Matlin, S. A., Orgill, M., & Skaza-Acosta, H. (2019). Graphical tools for conceptualizing systems thinking in chemistry education. *Journal of Chemical Education*, *96*(12), 2888–2900. https://doi.org/10.1021/acs.jchemed.9b00314

Beck, C. T., & Gable, R. K. (2001). Ensuring content validity: An illustration of the process. *Journal of Nursing Measurement*, *9*(2), 201–215. https://doi.org/10.1891/1061-3749.9.2.201

Bloom, Benjamin S. (1986). *Taxonomy of educational objectives. 1: Cognitive domain* (29. print). Longman.

Chen, A. Z., Peeks, M. D., & Kyne, S. H. (2025). Design, implementation, and evaluation of authentic learning activities to introduce chemistry students to systems thinking through green chemistry. *Journal of Chemical Education*, *102*(6), 2283–2293. https://doi.org/10.1021/ acs.jchemed.4c01326

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed). SAGE Publications.

Crocker, Linda. (2015). In L. Crocker, *International Encyclopedia of the Social & Behavioral Sciences* (pp. 774–777). Elsevier. https://doi.org/10.1016/b978-0-08-097086-8.44011-0

Delaney, S., Fehervari, A., Moon, V., & Schultz, M. (2022). Building more sustainably with concrete: A guided inquiry investigation. *Journal of Chemical Education*, *99*(12), 4169–4174. https://doi.org/10.1021/acs.jchemed.2c00344

D'eon, J., & Silverman, J. R. (2023). Using systems thinking to connect green principles and United Nations Sustainable Development Goals in a reaction stoichiometry module. *Green Chemistry Letters* *and Reviews*, *16*(1), 2185109. https://doi.org/10.1080/17518253.2023.2185109

Eskin, Mehmet. (2013). In M. Eskin, *Problem solving therapy in the clinical practice* (pp. 17–27). Elsevier. https://doi.org/10.1016/b978-0-12-398455-5.00003-6

Fitriyana, N., Marfuatun, M., & Priyambodo, E. (2019). The profile of students' analytical thinking skills on chemistry systemic learning approach. *Scientiae Educatia*, *8*(2), 207. https://doi.org/10.24235/sc.educatia.v8i2.5272

Johnstone, A. H. (2000). Teaching of chemistry: Logical or psychological? *Chem. Educ. Res. Pract.*, *1*(1), 9–15. https://doi.org/10.1039/A9RP90001B

Jumilah, J., & Wasis, W. (2023). Development of four-tier diagnostic test instrument to introduce misconceptions and identify causes of student misconceptions in the sub-topic of Bernoulli's principle. *Jurnal Penelitian Pendidikan IPA*, *9*(7), 5773–5781. https://doi.org/10.29303/jppipa.v9i7.4588

MacDonald, R., Elgersma, A., Holme, T., Snyder, J., Reynders, M., & Mahaffy, P. (2025). SOCKit: An online tool for systems thinking. *Journal of Chemical Education*, acs.jchemed.5c00310. https://doi.org/10.1021/acs.jchemed.5c00310

MacDonald, R. P., Pattison, A. N., Cornell, S. E., Elgersma, A. K., Greidanus, S. N., Visser, S. N., Hoffman, M., & Mahaffy, P. G. (2022). An interactive planetary boundaries systems thinking learning tool to integrate sustainability into the chemistry curriculum. *Journal of Chemical Education*, *99*(10), 3530–3539. https://doi.org/10.1021/acs.jchemed.2c00659

Mahaffy, P. G. (n.d.). *Systems thinking in chemistry for sustainability: Toward 2030 and beyond (STCS 2030+)*. Retrieved https://iupac.org/project/2020-014-3-050/

Mahaffy, P. G., Matlin, S. A., Holme, T. A., & MacKellar, J. (2019). Systems thinking for education about the molecular basis of sustainability. *Nature Sustainability*, *2*(5), 362–370. https://doi.org/10.1038/s41893-019-0285-3

Meadows, D. H. (2008). *Thinking in Systems*. Earthscan

Mulyani, S., Krismonita, M., & Yamtinah, S. (2022). Analisis butir soal dan kecukupan HOTS soal ujian akhir semester mata pelajaran kimia SMK Kelas X. *PAEDAGOGIA*, *25*(2), 162. https://doi.org/10.20961/paedagogia.v25i2.60913

Orgill, M., York, S., & MacKellar, J. (2019). Introduction to systems thinking for the chemistry education community. *Journal of Chemical Education*, *96*(12), 2720–2729. https://doi.org/10.1021/acs.jchemed.9b00169

Prodjosantoso, Anti Kolonial, Hertina, A. M., Department of Chemistry Education, Yogyakarta State University, Yogyakarta, Indonesia, Irwanto, I., & M.Pd., Department of Chemistry Education, Yogyakarta State University, Yogyakarta, Indonesia. (2019). The misconception diagnosis on ionic and covalent bonds concepts with three tier diagnostic test. *International Journal of Instruction*, *12*(1), 1477–1488. https://doi.org/10.29333/iji.2019.12194a

Resnik, D. B. (2024). *The ethics of research with human subjects: Protecting people, advancing science, promoting trust* (Vol. 111). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-82757-0

Reynders, M., Pilcher, L. A., & Potgieter, M. (2023). Teaching and assessing systems thinking in first-year chemistry. *Journal of Chemical*

Education, acs.jchemed.2c00891. https://doi.org/10.1021/acs.jchemed.2c00891

Reynders, M., Pilcher, L., & Potgieter, M. (2025). Development of systems thinking in a large first-year chemistry course using a group activity on detergents. *Journal of Chemical Education*, *102*(4), 1352–1366. https://doi.org/10.1021/acs.jchemed.4c01048

Sarita, M. R., & Wisudawati, A. W. (2024). Enhancing 21st Century Skills in Students with Special Needs Through STEM Learning. *JTK (Jurnal Tadris Kimiya)*, *9*(2), 209–219. https://doi.org/10.15575/jtk.v9i2.38883

Schultz, M., Lai, J., Ferguson, J. P., & Delaney, S. (2021). Topics amenable to a systems thinking approach: Secondary and tertiary perspectives. *Journal of Chemical Education*, *98*(10), 3100–3109. https://doi.org/10.1021/acs.jchemed.1c00203

Shekh-Abed, A., & Barakat, N. (2022). Exploring the correlation between systems thinking and soft skills for improved effectiveness of project based learning. *2022 IEEE Frontiers in Education Conference (FIE)*, 1–4. https://doi.org/10.1109/FIE56618.2022.9962414

Szozda, A. R., Mahaffy, P. G., & Flynn, A. B. (2023). Identifying chemistry students' baseline systems thinking skills when constructing system maps for a topic on climate change. *Journal of Chemical Education*, *100*(5), 1763–1776. https://doi.org/10.1021/acs.jchemed.2c00955

Talanquer, V., & Szozda, A. R. (2024). An educational framework for teaching chemistry using a systems thinking approach. *Journal of Chemical Education*, *101*(5), 1785–1792. https://doi.org/10.1021/acs.jchemed.4c00216

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wilson, M. (2023). *Constructing measures: An item response modeling approach* (2nd ed.). Routledge. https://doi.org/10.4324/9781003286929

Wisudawati, A. W., & Barke, H.-D. (2024). Systems thinking approach to understand Indonesia's ocean acidification. *Sustainable Chemistry and Pharmacy*, *37*, 101384. https://doi.org/10.1016/j.scp.2023.101384

York, S., Lavi, R., Dori, Y. J., & Orgill, M. K. (2019). Applications of Systems Thinking in STEM Education. *Journal of Chemical Education*, *96*(12), 2742–2751. https://doi.org/10.1021/acs.jchemed.9b00261

York, S., & Orgill, M. (2020). ChEMIST Table: A Tool for Designing or Modifying Instruction for a Systems Thinking Approach in Chemistry Education. *Journal of Chemical Education*, *97*(8), 2114–2129. https://doi.org/10.1021/acs.jchemed.0c00382

Yu, L., Wong, G. K. W., Zhang, B., & Wang, F. (2026). Educational Measurement with Emerging Technologies: A Systematic Review Through Evidentiary Lens on Granularity and Constructing Measures Theory. *Education Sciences*, *16*(4), 661. https://doi.org/10.3390/educsci16040661