Jurnal Pendidikan Siber Nusantara (JPSN)
E-ISSN : 2964-7517
P-ISSN : 2964-7525
+62 811 7404 445 https://siberpublisher.org/ siberpublisher.info@gmail.com
**DOI:** https://doi.org/10.38035/jpsn.v4i3.683


# Developing SDG 13 Based Instruments for Critical Thinking and Sustainability Consciousness on Climate Change: A Rasch Model Approach

**Najwa Syafaatul Fadhilah[1], Widi Purwianingsih[2], Amprasto Amprasto[3]**
[1]Universitas Pendidikan Indonesia, Indonesia, najwasyafaatulf@upi.edu
[2]Universitas Pendidikan Indonesia, Indonesia, widipurwianingsih@upi.edu
[3]Universitas Pendidikan Indonesia, Indonesia.

Corresponding Author: widipurwianingsih@upi.edu

***Abstract:*** *The escalation of the global climate crisis emphasizes the urgency of strengthening Sustainability Consciousness (SC), Critical Thinking Dispositions (CTD), and Critical Thinking Skills (CTS) through education, as emphasized in SDG 13.3. However, integrated measurement instruments for all dimensions in the context of climate change are still limited. This study developed three SDG 13-based instruments (SC, CTD, CTS) for climate change material in 10th grade Biology using the Rasch model approach. Content validation was tested by two biology experts, while empirical trials involved 115 students, with 75 respondents analyzed after misfit elimination. Winsteps analysis showed that all three instruments met unidimensionality (raw variance 38.6%-62.5%; unexplained first variance contrast 6.8%-7.4%) with high reliability (cronbach alpha 0.94-0.97). After item elimination based on fit (valid items) and local independence, 20 SC items, 14 CTD items, and 9 CTS items were obtained with an ideal distribution of difficulty. This instrument is valid, reliable, and supports integrated assessment for ESD climate action learning.*

***Keywords:*** *Sustainability Consciousness, Critical Thinking (Skill-Disposition), Climate Change, SDG 13, Rasch Model.*

## INTRODUCTION

The worsening global climate crisis is a crucial phenomenon that cannot be ignored. 2024 was recorded as the hottest year on record, with global temperatures rising 1.55°C above pre-industrial levels, surpassing the Paris Agreement's 1.5°C threshold for the first time (WMO, 2025; UN SDGs, 2025). Meanwhile, Indonesia ranks 6th as the largest contributor of greenhouse gases in the world with an increase in emissions of 5.0% compared to 2023 (European Commission, 2025). Through SDG target 13.3, the United Nations explicitly requires strengthening education and awareness in climate change mitigation and adaptation (UN, 2015). However, a 2023 study revealed that 69% of high school curricula did not include any reference to climate change, and 66% still ignored sustainability aspects altogether (UN

SDGs, 2024). In line with this, critical thinking competencies are positioned as a primary foundation within the Partnership for 21st Century Skills framework (Thornhill-Miller et al., 2023). Furthermore, Minister of Elementary and Secondary Education Regulation Number 10 of 2025 establishes critical reasoning as a key dimension of the graduate profile, as well as Phase E Biology Learning Outcomes for Grade 10 in the Independent Curriculum, which makes climate change an essential content (Kemendikbudristek, 2022).

Field conditions underscore this urgency, pre-student research revealed fundamental misconceptions regarding climate change and climate action. These findings indicate that students' critical thinking and sustainability awareness in the context of climate change have not yet developed sufficiently. Critical thinking is rational and reflective thinking that focuses on beliefs and decisions to be made (Ennis, 1996). More comprehensively Facione (1990) defines critical thinking as “purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based”. Critical thinking is not merely a skill process, but rather encompasses two complementary components, namely cognitive skill and dispositions (Facione, 1990).

Based on a study of 22 high school students in Indonesia, students' critical thinking skills, especially in Biology, showed suboptimal results, with 71% of students in the invisible or underdeveloped category (Aji & Nugraheni, 2023; Mahanal et al., 2018). In line with scientific literacy which is closely related to critical thinking, the score *Program for International Students Assessment* (PISA), Indonesia continued to decline until it reached its lowest point with a score of 383 in 2022, 102 points below the global average score (OECD, 2023). Critical thinking enables individuals to evaluate information multidimensionally and dissect the complexity of cause and effect in environmental problems (Voulvoulis et al., 2022), so that it transforms from being just a logical tool into the main foundation for the formation of sustainability consciousness (Setianingsih et al., 2024). UNESCO (2020) established critical thinking as one of the core competencies of sustainability transformation that moves individuals from technical literacy to real action.

Sustainability consciousness (SC) was introduced and developed by Gericke et al (2018) as an individual's experience and awareness of sustainable development, operationalized through three main dimensions: knowledge, attitudes, and behavior, within three pillars: environmental, social, and economic. Findings indicate that attitudes toward sustainability have a more significant influence on behavior than mere knowledge acquisition (Mohamad Saleh et al., 2022; Kollmuss & Agyeman, 2002). Therefore, SC needs to be instilled more intensively, to encourage changes in behavior and lifestyles that are more environmentally friendly, increase social awareness and real action for sustainability, and acquire relevant knowledge and competencies for the future (Purwianingsih et al., 2022).

Although various critical thinking assessment instruments have been developed in the context of science education, there is a significant gap in the specific integration of SDG 13 into these instruments. Most instruments only measure the skill aspects of critical thinking, without explicitly linking them to environmental sustainability issues such as climate change and global warming (Abidin & Fatimatuzzaro, 2025; Dar et al., 2025). In addition, based on the Education for Sustainable Development (ESD) framework, critical thinking essentially goes beyond the skill aspect alone, but this competency demands an affective tendency (disposition) that is able to move students to take real action in overcoming the environmental crisis (Kioupi & Voulvoulis, 2019). Kusumawati, Supahar, and Pebriana (2024) stated that although students have a basic understanding of global warming, students still have difficulty analyzing the relationship between environmental, social, and economic aspects in the context of sustainable development. This indicates that the available assessment instruments are not yet able to comprehensively measure the dimensions of Sustainability Consciousness (SC) within the scope of sustainability knowledge, attitudes, and behavior as conceptualized by Gericke et al. (2018) which includes environmental, economic, and social dimensions. This strengthens the argument that there is no instrument that is able to integratedly measure critical thinking skills in the skill and dispositional dimensions as well as sustainability consciousness in the context of climate change that is relevant to learning Biology class X. Yulisharyasti et al (2023) emphasized that instruments that do not integrate a comprehensive measurement approach risk failing to accurately detect student abilities. Therefore, developing an instrument that integrates critical thinking assessment and sustainability awareness in the context of SDG 13 for 10th-grade Biology learning is an urgent, unmet need.

The Rasch Model approach was chosen as the primary validation method due to its comprehensive psychometric parameter testing capabilities, thus demonstrating empirical validity and reliability of the resulting assessment instrument. Unlike the Classical Test Theory (CTT) approach, which relies on descriptive statistics based on raw scores and is sample-dependent, the Rasch Model produces estimates of participant ability (person ability) that are independent of item characteristics, and conversely, estimates of item difficulty are independent of the ability of the sample used (Sumintono & Widhiarso, 2014; Yulisharyasti et al., 2023). This advantage is crucial in the development of new instruments, as it allows researchers to evaluate the quality of each item more objectively and accurately. The study conducted Avinç and Doğan (2024) states that the Rasch Model can analyze instrument unidimensionality, item bias through Differential Item Function (DIF), and verify data-model suitability through item fit and person fit analysis, all of which cannot be done comprehensively through CTT. In addition, the Rasch Model can convert ordinal data from a Likert scale into interval data with logit units, so that measurements become more precise and meet the fundamental requirements of quantitative measurement (Sumintono & Widhiarso, 2015). In the context of an instrument that integrates critical thinking and sustainability consciousness with constructs that are measured in a complex and multidimensional manner, the superiority of the Rasch Model in mapping the relationship between student abilities and item difficulty levels through a calibration map becomes very relevant (Avinç & Doğan, 2024; Yulisharyasti et al., 2023). Based on the description above, this study aims to develop and validate three integrated instruments based on SDG 13 which simultaneously measure critical thinking skill, dispositional critical thinking and sustainability consciousness in the context of climate change in Biology learning for class X using the Rasch Model approach.

## METHOD

This research is an instrument development research with a quantitative approach based on Rasch model psychometric analysis, which was chosen because of its superiority in producing estimates of person ability and item difficulty levels independently on the same logit scale (Bond & Fox, 2015). The development of the instrument began with mapping the indicators for the three instruments including: (1) a critical thinking skill essay test (14 items) adapted from the California Critical Thinking Skills Test (CCTST) based on the APA Delphi Report (Facione, 1990) which was taken from five indicators including interpretation, analysis, evaluation, inference, and explanation; (2) a critical thinking disposition questionnaire (21 items, 4 point Likert scale) adapted from the California Critical Thinking Dispositions Inventory (CCTDI) based on seven indicators Facione (1992), which includes truth-seeking, open-mindedness, analyticity, systematicity, self-confidence, inquisitiveness, and cognitive maturity; and (3) sustainability consciousness questionnaire (21 items, 4-point Likert scale) based on SCQ Gericke et al (2018) which includes dimensions of knowingness, attitudes, and behavior on three pillars of environment, social, and economy. The three instruments are contextualized in the climate change material for class X in Biology learning oriented towards SDG 13. The details of the indicators and items in each instrument are as follows.

**Table 1. Grid of sustainability consciousness instrument**

| Indicator | | Item details |
|---|---|---|
| **Psychological constructs** | **Dimensions** | |
| Sustainability Knowingness | Environmental | S1, S2, S3 |
| | Social | S4, S5 |
| | Economic | S6, S7 |
| Sustainability Attitude | Environmental | S8, S9, S10 |
| | Social | S11, S12 |
| | Economic | S13, S14 |
| Sustainability Behaviors | Environmental | S15, S16, S17 |
| | Social | S18, S19 |
| | Economic | S20, S21 |

**Table 2. Grid of critical thinking disposition instrument**

| Indicator | Item details |
|---|---|
| Truth-Seeking | D1, D2, D3 |
| Open-mindedness | D4, D5, D6 |
| Analyticity | D7, D8, D9 |
| Systematicity | D10, D11, D12 |
| Inquisitiveness | D13, D14, D15 |
| Critical Thinking Self-confidence | D16, D17, D18 |
| Cognitive Maturity | D19, D20, D21 |

**Table 3. Grid of critical thinking skill instrument**

| Indicator | Item details |
|---|---|
| Interpretation | K1, K2, K3 |
| Analysis | K4, K5, K6 |
| Evaluation | K7, K8 |
| Inference | K9, K10, K11 |
| Explanation | K12, K13, K14 |

Before being tested, the three instruments were validated by two expert lecturers: a Biology material expert and an Education Evaluation expert. Validation was conducted using a structured validation sheet that included item relevance to indicators, language readability, and suitability to the learning outcomes of the Biology climate change material for grade 10 of the Independent Curriculum. The empirical trial involved 115 grade 10 students selected through purposive sampling, with the criteria being that students had studied climate change material in Biology lessons. Prior to item analysis, a person fit analysis was conducted to identify respondents who provided inconsistent response patterns, such as answering randomly or guessing, which could distort item parameter estimates. Respondents were considered misfit if their scores of the Outfit and Infit MNSQ were outside the 0.5-1.5 range (Linacre, 2002). Based on these criteria, 40 respondents were identified as misfit and excluded from further analysis, resulting in a final data analysis of 75 respondents.

Empirical analysis was conducted using Winsteps version 5.10.3.0 to test the unidimensionality local independence, level of difficulty, validity, and reliability of the instrument. Unidimensionality was tested through Principal Component Analysis of Residuals (PCAR), with the criteria of a minimum measurement dimension variance of >20% and a residual first component variance of <15% (Linacre, 2010); local independence was analyzed through standardized residual correlation between items with a criterion of ≤0.30 (Fisher, 2007). The fit of each item to the Rasch model was analyzed through the Infit and Outfit MNSQ with a range of 0.5–1.5 which is considered productive for measurement and shows that the item contributes meaningfully without causing excessive distortion (Boone et al., 2014; Sumintono & Widhiarso, 2015), items whose values are outside this range are categorized as misfit and considered for elimination. The level of item difficulty is analyzed from the sign of the item measure logit value (+ or -), while item and person reliability are tested with the criteria of values > 0.70 as explained in Linacre (2002) as an indicator of measurement consistency.

## RESULT AND DISCUSSION

Prior to the empirical trial, all three instruments were validated by biology experts, and measurements were analyzed using a structured validation sheet covering the relevance of items to indicators, language readability, and suitability to the learning outcomes and learning objectives for climate change material in biology for grade X phase E of the independent curriculum. The instruments had a systematic structure, but there were two points of improvement. First, the sentence wording was simplified to make it less academic for grade X students, in line with Abidin and Fatimatuzzaro (2025) which emphasizes the importance of readability as a validating factor so that responses reflect students' actual abilities. Second, a remapping was conducted so that all items explicitly integrate the four elements of climate change: causes, mitigation, impacts, and adaptation (Kemendikbudristek, 2024).

An empirical instrument trial was conducted to determine the feasibility of the sustainability consciousness, critical thinking disposition, and critical thinking skill instruments. This trial involved 115 respondents, but only 75 respondents were used for data analysis because 40 respondents were identified as misfits. The instrument trial included dimensionality, local independence, item difficulty, validity, and reliability tests, all of which were analyzed using the Rasch model test application using the Winsteps application version 5.10.3.0.

### Unidimensionality

The results of the dimensionality test of the Sustainability Consciousness (SC), Critical Thinking Disposition (CTD), and Critical Thinking Skill (CTS) can be seen in Table 4 below.

**Table 4. Results of the dimensionality test of SC, CTD, and CTS**

| Aspect | Observed (%) | | |
|---|---|---|---|
| | ***Sustainability Consciousness* (SC)** | **Critical Thinking Disposition (CTD)** | **Critical Thinking Skill (CTS)** |
| *Raw Variance Explained by Measures* | 38.6% | 48.6% | 62.5% |
| *Unexplained Variance in 1st Contrast* | 6.8% | 7.4% | 7.2% |

The results of the dimensionality test show that the Raw variance explained by measures value is 38.6% in the SC questionnaire, 48.6% on the CTD questionnaire, and 62.5% on the CTS questions, indicating that the instrument has a strong ability to explain the measured variables. The Unexplained variance in 1st contrast value shows a result of 6.8% on the SC questionnaire, 7.4% on the CTD questionnaire, and 7.2% on the CTS questions. According to Linacre (2010), an instrument is categorized as unidimensional if the raw variance explained by measures is at least ≥20%. A range of 20%-40% is categorized as "acceptable," 40%-60% as "good," and ≥60% as "excellent". Meanwhile, the unexplained variance in first contrast is <15%. Based on these criteria, the SC instrument has the ability to explain variables in the "acceptable" category (38.6%), CTS in the "excellent" category (62.5%), and CTD in the "good" category (48.6%). The SC raw variance value, which is in the acceptable category, is understandable given the high complexity of the SC construct, which inherently integrates nine sub-dimensions from a combination of three psychological constructs and three pillars of sustainability (Gericke et al., 2019). Nevertheless, this value still exceeds the minimum threshold of 20% set by Linacre (2010), thus the SC instrument still meets unidimensionality. The unexplained variance in first contrast values for all three instruments were well below the maximum threshold of 15% (6.8%-7.4%), indicating that the residual variance not explained by the measurement dimensions was very small. Thus, it can be concluded that all three instruments met the unidimensionality criteria, meaning that all items in each instrument consistently measure one dimension of the intended construct without any significant interference from other dimensions.

## Validity

The results of the validity test of the Sustainability Consciousness (SC), Critical Thinking Disposition (CTD), and Critical Thinking Skill (CTS) can be seen in Table 5 below.

**Table 5. Validity test results of SC, CTD, and CTS**

| ***Sustainability Consciousness* (SC)** | | | | **Critical Thinking Disposition (CTD)** | | | | **Critical Thinking Skill (CTS)** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Item | Infit MNSQ | MNSQ Outfit | Category | Item | Infit MNSQ | MNSQ Outfit | Category | Item | Infit MNSQ | MNSQ Outfit | Category |
| S6 | 1.58 | 1.63 | Invalid | D20 | 1.66 | 1.46 | Invalid | K9 | 1.84 | 1.67 | Invalid |
| S18 | 1.36 | 1.40 | Valid | D14 | 1.54 | 1.22 | Invalid | K1 | 1.56 | 1.43 | Invalid |
| S7 | 1.35 | 1.36 | Valid | D3 | 1.42 | 1.43 | Valid | K4 | 1.33 | 1.22 | Valid |
| S2 | 1.32 | 1.09 | Valid | D21 | 1.18 | 1.36 | Valid | K2 | 1.07 | 1.08 | Valid |
| S1 | 1.24 | 1.21 | Valid | D5 | 1.18 | 1.32 | Valid | K3 | 1.03 | 1.04 | Valid |
| S21 | 1.17 | 1.24 | Valid | D2 | 1.02 | 1.16 | Valid | K13 | 1.03 | 0.98 | Valid |
| S20 | 1.02 | 1.11 | Valid | D12 | 1.11 | 0.97 | Valid | K10 | 0.94 | 0.90 | Valid |
| S19 | 1.05 | 0.93 | Valid | D17 | 1.10 | 0.88 | Valid | K12 | 0.88 | 0.86 | Valid |
| S3 | 1.04 | 0.90 | Valid | D6 | 0.92 | 1.08 | Valid | K7 | 0.84 | 0.83 | Valid |
| S12 | 1.02 | 0.88 | Valid | D19 | 0.94 | 1.01 | Valid | K8 | 0.77 | 0.83 | Valid |
| S4 | 1.01 | 0.92 | Valid | D4 | 0.98 | 0.85 | Valid | K6 | 0.64 | 0.77 | Valid |
| S8 | 0.91 | 1.00 | Valid | D15 | 0.98 | 0.86 | Valid | K11 | 0.71 | 0.70 | Valid |
| S5 | 0.98 | 0.84 | Valid | D13 | 0.90 | 0.79 | Valid | K14 | 0.70 | 0.69 | Valid |
| S15 | 0.97 | 0.91 | Valid | D9 | 0.82 | 0.61 | Valid | K5 | 0.68 | 0.69 | Valid |
| S9 | 0.65 | 0.89 | Valid | D16 | 0.82 | 0.68 | Valid | | | | |
| S14 | 0.87 | 0.84 | Valid | D8 | 0.76 | 0.61 | Valid | | | | |
| S11 | 0.85 | 0.81 | Valid | D18 | 0.75 | 0.63 | Valid | | | | |
| S16 | 0.76 | 0.70 | Valid | D1 | 0.69 | 0.67 | Valid | | | | |
| S13 | 0.74 | 0.70 | Valid | D11 | 0.63 | 0.52 | Valid | | | | |
| S10 | 0.64 | 0.61 | Valid | D10 | 0.59 | 0.42 | Invalid | | | | |
| S17 | 0.62 | 0.61 | Valid | D7 | 0.53 | 0.57 | Valid | | | | |

The validity of the item is evaluated using the Infit and Outfit Mean Square (MNSQ) values, with the criteria that the item is categorized as valid (fit) if the value is in the range $0.5 < \text{MNSQ} < 1.5$ (Wright & Linacre, 1994). Items with MNSQ values above 1.5 indicate underfit (items are too noisy/unpredictable), while values below 0.5 indicate overfit (items are too predictable) (Boone et al., 2013; Sumintono & Widhiarso, 2015). The results of the SC instrument analysis showed that most items functioned well, but one item, namely S6, was indicated as unfit with an Infit MNSQ value of 1.58 and an Outfit MNSQ value of 1.63, both exceeding the upper limit of 1.50. Item S6 was declared invalid (misfit) and eliminated, while the other 20 items were within the acceptable range and were declared valid. The results of the CTD instrument analysis showed that most items functioned well, but three items, namely D20 with Infit MNSQ 1.66, D14 with Infit MNSQ 1.54, and D10 with MNSQ Outfit 0.42 indicated unfit items. D20, D14, and D10 were declared invalid (misfit) and eliminated, while the other 18 items were within the acceptable range and declared valid. The results of the CTS instrument analysis show that most of the items function well, but two items, namely K9 with an Infit MNSQ value 1.84 and MNSQ Outfit 1.67, and K1 with *Infit MNSQ* 1.56 indicated unfit. Items K9 and K1 were declared invalid (misfit) because they exceeded the upper limit of 1.50 and therefore needed to be eliminated, while the other 12 items were within the acceptable range and were declared valid.

## Local independence

The results of the local independence test of the Sustainability Consciousness (SC) questionnaire, Critical Thinking Disposition (CTD), and Critical Thinking Skill (CTS) questions can be seen in Table 6 below.

**Table 6. Results of local independence tests of SC, CTD, and CTS**

| ***Sustainability Consciousness* (SC)** | | | **Critical Thinking Disposition (CTD)** | | | **Critical Thinking Skill (CTS)** | | |
|---|---|---|---|---|---|---|---|---|
| **Correlation** | **Item 1** | **Item 1** | **Correlation** | **Item 1** | **Item 1** | **Correlation** | **Item 1** | **Item 1** |
| 0.28 | S4 | S10 | 0.51 | D4 | D19 | 0.42 | K12 | K14 |
| 0.27 | S1 | S2 | 0.43 | D15 | D16 | 0.37 | K12 | K13 |
| 0.26 | S11 | S13 | 0.36 | D1 | D7 | 0.36 | K13 | K14 |
| 0.26 | S17 | S21 | 0.35 | D8 | D19 | 0.31 | K6 | K8 |
| 0.26 | S6 | S19 | 0.33 | D4 | D8 | -0.44 | K3 | K14 |
| 0.23 | S2 | S3 | 0.30 | D2 | D16 | -0.44 | K4 | K13 |
| -0.45 | S1 | S18 | -0.40 | D5 | D15 | -0.34 | K1 | K8 |
| -0.42 | S3 | S6 | -0.39 | D14 | D19 | -0.32 | K5 | K12 |
| -0.41 | S12 | S19 | -0.38 | D16 | D19 | -0.32 | K3 | K10 |
| -0.35 | S11 | S21 | -0.38 | D4 | D7 | -0.31 | K3 | K12 |
| -0.35 | S5 | S21 | -0.37 | D2 | D21 | -0.30 | K4 | K14 |
| -0.34 | S13 | S16 | -0.36 | D2 | D17 | -0.27 | K9 | K12 |
| -0.30 | S7 | S20 | -0.36 | D10 | D18 | -0.27 | K4 | K10 |
| -0.28 | S2 | S19 | -0.35 | D16 | D21 | -0.26 | K2 | K14 |
| -0.27 | S3 | S19 | -0.33 | D5 | D16 | -0.26 | K5 | K13 |
| -0.26 | S4 | S20 | -0.31 | D3 | D20 | -0.26 | K2 | K12 |
| -0.25 | S4 | S15 | -0.31 | D1 | D15 | -0.25 | K1 | K6 |
| -0.24 | S4 | S8 | -0.31 | D4 | D15 | -0.25 | K7 | K8 |
| -0.24 | S15 | S16 | -0.30 | D15 | D20 | -0.23 | K3 | K13 |
| -0.23 | S5 | S6 | -0.29 | D16 | D20 | -0.23 | K1 | K10 |

The results of the local independence test are used to identify whether there are pairs of items with high standardized residual correlations, which may indicate local dependency or additional dimensions not measured by the primary dimensions. The criterion used is a residual correlation ≤ 0.30 (Fisher, 2007); item pairs with correlation values above 0.30 potentially indicate local dependency. In the SC instrument, the analysis results show that all item pairs have residual correlation values below 0.30 (the highest value is 0.28), so it can be concluded that the SC instrument fulfills the local independence assumption well. In the CTD instrument, several item pairs were found that exceeded the 0.30 threshold, namely D4-D19 (0.51), D15-D16 (0.43), D1-D7 (0.36), D8-D19 (0.35), and D4-D8 (0.33). The high correlation in these pairs indicates the possibility of overlapping content between items, especially in item D19 which appears in pairs with two other items (D4 and D8). For each pair that is declared twins, one of the items must be eliminated. The items that were eliminated from each pair include: D15, D1, D19 and D4. In the CTS instrument, item pairs K12-K14 (0.42), K12-K13 (0.37), K13-K14 (0.36), and K6-K8 (0.31) were found to exceed the threshold. This pattern indicates that items K12, K13, and K14 are substantially related, indicating the possibility that the three items measure similar aspects of the same indicator. For each pair declared twins, one of the items must be eliminated. Some of the items eliminated from each pair include: K6, K13, and K14.

The decision to eliminate items in pairs identified as twins is prioritized based on two main considerations: first, items with infit and outfit MNSQ values that are further than 1.0 are prioritized for elimination (Wright & Masters, 1982; Linacre, 2002); Second, the eliminated items are selected with the condition that each indicator remains represented by at least one item (De Vellis, 2017). High residual correlations across a given item pair indicate substantial content overlap, with items within the same indicator likely measuring constructs that are so close together that one of them is redundant (Christensen et al., 2017). In the critical thinking skill instrument, there is one indicator whose three items tend to overlap, namely the explanation indicator, so one item among the three must be maintained to represent this indicator, namely K12.

## Item difficulty

The results of the difficulty test of the Sustainability Consciousness (SC), Critical Thinking Disposition (CTD), and Critical Thinking Skill (CTS) items can be seen in Table 7 below.

**Table 7. Item difficulty test results of SC, CTD, and CTS**

| ***Sustainability Consciousness* (SC)** | | | **Critical Thinking Disposition (CTD)** | | | **Critical Thinking Skill (CTS)** | | |
|---|---|---|---|---|---|---|---|---|
| **Item** | **JMLE Measure (Logit)** | **Category** | **Item** | **JMLE Measure (Logit)** | **Category** | **Item** | **JMLE Measure (Logit)** | **Category** |
| S6 | 1.02 | Difficult | D1 | 1.41 | Difficult | K9 | -1.48 | Easy |
| S18 | 0.63 | Difficult | D2 | -1.87 | Easy | K1 | -0.76 | Easy |
| S7 | -0.19 | Easy | D3 | -0.19 | Easy | K4 | -0.02 | Easy |
| S2 | -0.78 | Easy | D4 | -1.09 | Easy | K2 | 0.17 | Difficult |
| S1 | 1.02 | Difficult | D5 | -1.60 | Easy | K3 | 0.45 | Difficult |
| S21 | -0.03 | Easy | D6 | -0.85 | Easy | K13 | 0.28 | Difficult |
| S20 | 0.23 | Difficult | D7 | 1.84 | Difficult | K10 | -0.44 | Easy |
| S19 | -0.54 | Easy | D8 | -0.97 | Easy | K12 | 0.06 | Difficult |
| S3 | -0.72 | Easy | D9 | 0.12 | Difficult | K7 | 0.59 | Difficult |
| S12 | -0.84 | Easy | D10 | -0.73 | Easy | K8 | 0.34 | Difficult |
| S4 | -0.19 | Easy | D11 | 0.02 | Difficult | K6 | 0.17 | Difficult |
| S8 | 0.28 | Difficult | D12 | 0.22 | Difficult | K11 | 0.37 | Difficult |
| S5 | -0.91 | Easy | D13 | 0.88 | Difficult | K14 | -0.13 | Easy |
| S15 | 0.23 | Difficult | D14 | -0.19 | Easy | K5 | 0.40 | Difficult |
| S9 | 0.23 | Difficult | D15 | 1.84 | Difficult | | | |
| S14 | 0.54 | Difficult | D16 | 1.24 | Difficult | | | |
| S11 | 0.28 | Difficult | D17 | 0.41 | Difficult | | | |
| S16 | -0.66 | Easy | D18 | 1.06 | Difficult | | | |
| S13 | -0.14 | Easy | D19 | -1.47 | Easy | | | |
| S10 | -0.31 | Easy | D20 | 0.12 | Difficult | | | |
| S17 | 0.83 | Difficult | D21 | -0.19 | Easy | | | |

Item difficulty was analyzed using the JMLE Measure (logit) value, which reflects how difficult a statement/question is for respondents to agree with or answer correctly. A positive logit value indicates a relatively more difficult item, while a negative value indicates an easier item. In the SC instrument, the logit value ranges from -0.91 (S5, easiest) to +1.02 (S6 and S1, most difficult). In the CTD instrument, the logit value ranges from -1.87 (D2, easiest) to +1.84 (D7 and D15, most difficult), indicating a more diverse distribution of difficulty levels compared to SC. In the CTS instrument, the logit value ranges from -1.48 (K9, easiest) to +0.59 (K7, most difficult). According to Bond and Fox (2015), a good distribution of item logit values ranges from -2.00 to +2.00. Thus, all items in the three instruments are within the acceptable range, indicating that the level of item difficulty varies but is still within reasonable limits for measuring student abilities at various levels.

In the SC instrument, the items with the highest level of difficulty tend to come from the sustainability knowing and attitude dimensions, which require students to understand the complex interrelationships between environmental, social, and economic aspects of sustainable development. Conversely, the items in the behavior dimension are mostly in the easy category, indicating that students find it relatively easier to express their willingness to behave in an environmentally friendly manner than to conceptually understand sustainability. In the CTD instrument, the concentration of difficult items is on the truth-seeking, systematicity, and CT self-confidence indicators which are theoretical, reflective and self-regulatory, this requires a higher level of thinking maturity compared to motivational dispositions such as inquisitiveness (Facione, 1990). In the CTS instrument, the concentration of difficult items is on the evaluation and analysis indicators which require students not only to understand information but also to assess the validity of arguments and analyze cause-and-effect relationships in the context of climate change which include high-level skill abilities, while the interpretation and inference indicators tend to be easier, because they generally only require literal understanding.

## Reliability Test

The results of the reliability test of the Sustainability Consciousness (SC), Critical Thinking Disposition (CTD), and Critical Thinking Skill (CTS) items can be seen in Table 8 below.

**Table 8. Reliability test results of SC, CTD, and CTS**

| Reliability Index | Sustainability Consciousness (SC) | Critical Thinking Disposition (CTD) | Critical Thinking Skill (CTS) |
|---|---|---|---|
| *Person Reliability* | 0.84 | 0.93 | 0.92 |
| *Cronbach Alpha* (KR-20) | 0.94 | 0.97 | 0.96 |
| *Item Reliability* | 0.81 | 0.90 | 0.72 |

The reliability of the instrument is evaluated through three indexes, namely Person Reliability, Item Reliability, and Cronbach Alpha (KR-20), with value criteria in the range of 0.70-0.99 as an indicator of acceptable measurement consistency (Bond & Fox, 2007). The analysis results show a Cronbach's Alpha (KR-20) value of 0.94 on the SC questionnaire, 0.97 on the CTD questionnaire, and 0.96 on the CTS questions, these values indicate that the overall interaction between respondents and the test items is in the very good category. Person Reliability values of 0.84 (SC), 0.93 (CTD), and 0.92 (CTS), and Item Reliability values of 0.81 (SC), 0.90 (CTD), and 0.72 (CTS), all of which are in the high and very high categories. The results indicate that the consistency of students' answers and the functional quality of the test items developed have high reliability.

Based on the results of empirical instrument trials, it can be concluded that several items were eliminated from each instrument. The details of the items eliminated from each instrument are as follows.

**Table 9. Final item recapitulation of SC, CTD, and CTS**

| Instrument | Initial items | Elimination of validity | Eliminate local independence | Final item |
|---|---|---|---|---|
| SC | 21 | 1 (S6) | - | 20 |
| CTD | 21 | 3 (D20, D14, D10) | 4 (D15, D1, D19, D4) | 14 |
| CTS | 14 | 2 (K9, K1) | 3 (K6, K13, K14) | 9 |

Based on the analysis results, in the SC instrument, there was only one item eliminated, namely the sustainability knowingness (economic) indicator. In the CTD instrument, almost every indicator had an item eliminated except for the analyticity and critical thinking self-confidence indicators, while the indicator with the most eliminated items was the inquisitiveness and cognitive maturity indicators. Similarly, in the CTS instrument, almost every indicator had an item eliminated except for the evaluation indicator, while the indicator with the most eliminated items was the explanation indicator. The novelty of this research lies in the development of three integrated instruments within one SDG 13 framework that is adapted to the ESD climate action learning objectives that measure critical thinking holistically, both skill and dispositional dimensions as well as sustainability consciousness in the context of climate change in class X Biology. This integration is relevant to the views of Facione (1990) and Dwyer et al (2014) that critical thinking cannot be separated from disposition, because measurements that only cover one dimension will produce an incomplete picture. In line with this, the ESD framework emphasizes that sustainability competencies must encompass knowledge, attitudes, and behavior simultaneously (Gericke et al., 2019; Kioupi & Voulvoulis, 2019). While UNESCO (2017) explicitly states that achieving SDG 13 requires the development of competencies that integrate these three dimensions, the SCQ has been proven valid in measuring SC (Ariza et al., 2021). However, the instrument has a broad scope that represents all 17 SDGs. SC instruments that focus on one specific SDG 13 goal and combine skill CT and dispositional CT measurements in a specific climate change context in 10th grade Biology learning have not been widely developed (Dar et al., 2025), so that this research fills this gap substantively.

## CONCLUSION

This study produced three integrated instruments for SDG 13, namely sustainability consciousness (SC), critical thinking disposition (CTD), and critical thinking skill (CTS) on climate change material for grade X Biology which have been tested valid and reliable through Rasch model analysis. All instruments have met the requirements of unidimensionality with a raw variance value range of 38.6%-62.5%, unexplained variance first contrast with a value range of 6.8%-7.4%, and a very high reliability value (Cronbach Alpha) with a range of 0.94-0.97. Based on the results of the fit and local independence analysis, the final items were obtained as many as 20 SC items, 14 CTD items, and 9 CTS items with an ideal difficulty range (logit) to map student abilities. This instrument provides a novel contribution in the form of a comprehensive measuring tool that integrates skill, dispositional, and sustainability awareness dimensions in one specific context to support the evaluation of ESD climate action competencies.

## REFERENCE


Abidin, Z., & Fatimatuzzaro, F. (2025). Development of a critical thinking skills assessment instrument for students on diffusion and osmosis subject. Mangifera Edu Journal, 9(2), 86–95.

Aji, GP, & Nugraheni, BI (2023). Critical thinking skills of high school and vocational school students in Indonesia: A literature review. International Journal of Education and Social Science Research, 6(3), 353–367.

Ariza, M.R., Boeve-de Pauw, J., Olsson, D., Van Petegem, P., Parra, G., & Gericke, N. (2021). Promoting environmental citizenship in education: The potential of the Sustainability Consciousness Questionnaire to measure the impact of interventions. Sustainability, 13(20), 11420.

Avinç, E., & Doğan, F. (2024). Digital literacy scale: Validity and reliability study with the Rasch model. Education and Information Technologies, 29, 22895–22941.

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

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. Springer.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model. Educational and Psychological Measurement, 77(3), 518–531.

Dar, SR, Emiliannur, Festiyed, Riyasni, S., & Winarno, N. (2025). The importance of integrating Education for Sustainable Development in global warming assessments to improve students' critical thinking skills. Journal of Physics Learning Research (JPPF), 11(2), 223–234.

DeVellis, R.F. (2017). Scale development: Theory and applications (4th ed.). SAGE Publications.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. Thinking Skills and Creativity, 12, 43–52.

Ennis, R.H. (1996). Critical thinking. Prentice Hall.

European Commission, Joint Research Center (JRC). Emissions Database for Global Atmospheric Research (EDGAR). 2025.

Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). California Academic Press.

Facione, P. A., & Facione, N. C. (1992). The California Critical Thinking Disposition Inventory (CCTDI) and the CCTDI test manual (The Delphi Report). California Academic Press.

Fisher, WP, Jr. (2007). Rating scale instrument quality criteria. Rasch Measurement Transactions, 21(1), 1095.

Gericke, N., Boeve-de Pauw, J., Berglund, T., & Olsson, D. (2018). The sustainability consciousness questionnaire: The theoretical development and empirical validation of an evaluation instrument for stakeholders working with sustainable development. Sustainable Development, 26(6), 613–627.

Kioupi, V., & Voulvoulis, N. (2019). Education for Sustainable Development: A systemic framework for connecting the SDGs to educational outcomes. Sustainability, 11(21), 6104.

Kollmuss, A., & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? Environmental Education Research, 8(3), 239–260.

Kusumawati, AE, Supahar, S., & Pebriana, IN (2024). Development of a global warming theme assessment instrument to measure high school students' science process skills and collaboration skills. Journal of Science Education, 6(2), 105–113.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. Journal of Applied Measurement, 3(1), 85–106.

Linacre, JM (2010). Winsteps® Rasch measurement computer program user's guide. Winsteps.com.

Mahanal, S., Tendrita, M., Ramadhan, F., Ismirawati, N., & Zubaidah, S. (2018). The analysis of students' critical thinking skills on biology subject. Anatolian Journal of Instruction, 3(2), 21–39.

Kemendikbudristek. (2022). Learning Outcomes of Biology Subject Phase E for Senior High Schools (SMA/MA). Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia.

Mohamad Saleh, MS, Mehellou, A., Huang, M., & Briandana, R. (2022). The influence of sustainability knowledge and attitude on sustainable intention and behavior of Malaysian and Indonesian undergraduate students. Research in Comparative and International Education, 17(4).

OECD. 2023. PISA 2022 Assessment and Analytical Framework. OECD Publishing.

Purwianingsih, W., Novidsa, I., & Riandi, R. (2022). Efforts to enhance sustainable consciousness and critical thinking in high school students through learning projects.

Indonesian Journal of Teaching in Science, 2(2). https://ejournal.upi.edu/index.php/IJoTis/article/view/55447

Setianingsih, ES, Purwianingsih, W., & Riandi. (2024). Critical thinking as a foundation for sustainable consciousness in biology education. Indonesian Journal of Biological Education, 8(1), 45–58.

Sumintono, B., & Widhiarso, W. (2014). Application of the Rasch model to social science research (revised edition). Trim Komunikata Publishing House.

Sumintono, B., & Widhiarso, W. (2015). Application of Rasch modeling in educational assessment. Trim Komunikata Publishing House.

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. Journal of Intelligence, 11(3), 54.

UNESCO. (2020). Education for sustainable development: A roadmap (ESD for 2030). UNESCO Publishing.

United Nations Sustainable Development Goals [UN SDGs]. (2024). Goal 13: Take urgent action to combat climate change and its impacts. https://sdgs.un.org/goals/goal13

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development (A/RES/70/1). https://sustainabledevelopment.un.org/post2015/transformingourworld

Voulvoulis, N., Burgman, M. A., & Alquier, A. M. (2022). Critical thinking for sustainability: Moving beyond disciplinary boundaries. One Earth, 5(3), 225–228.

World Meteorological Organization. (2025). State of the global climate 2024 (WMO-No. 1368). https://wmo.int/publication-series/state-of-global-climate/state-of-global-climate-2024

Wright, B.D., & Linacre, J.M. (1994). Reasonable mean-square fit values. Rasch Measurement Transactions, 8(3), 370.

Wright, B.D., & Masters, G.N. (1982). Rating scale analysis. MESA Press.

Yulisharyasti, L., Nurdin, A., Aulia, N., Arfa, FAH, & Fadjryani. (2023). Analyzing the quality of measurement instruments of multiple choice questions on Class XI Economics material in Public High School 3 Gorontalo through Classical Test Theory and Rasch Models. Parameters: Journal of Statistics, 3(1), 28–38.