MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2025, 21(5), em2632
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/16310

# Design of instruments for scientific creative thinking skills and creative thinking digital skills: Rasch models and confirmatory factor analysis

Firdaus Firdaus [1,2], Wiyanto Wiyanto [1*], Ngurah Made Darma Putra [1], Wiwi Isnaeni [1]

[1] Universitas Negeri Semarang, Semarang, INDONESIA
[2] Universitas Sains Al-Qur'an, Wonosobo, INDONESIA

Received 18 November 2024 ▪ Accepted 21 March 2025

**Abstract**

This study aims to design a valid and reliable instrument to measure scientific creative thinking skills (SCTS) and creative thinking digital skills (CTDS) for undergraduates majoring in physics education. The research employs the research and development approach, using an adapted Borg & Gall model. The instrument is designed based on the three-dimensional scientific structure creativity model for measuring SCTS and an adapted van Laar's model for CTDS. Validity and reliability testing is conducted through empirical testing using the Rasch model and confirmatory factor analysis (CFA). The study's results indicate that the developed instrument has met high standards of validity and reliability. The main finding of this study is the availability of an effective instrument to measure creative thinking skills in scientific and digital contexts. This research contributes to the development of instruments and encourages innovation in learning. The results can help educators provide targeted feedback and design learning that strengthens students' creative thinking skills.

**Keywords:** CFA, CTDS, instruments, Rasch models, SCTS

## INTRODUCTION

Students represent a young intellectual demographic that plays a pivotal role in society, the nation, and the state. Therefore, they must be equipped with the important skills required in the 21st century (Benbow et al., 2021; Valtonen et al., 2021), including hard and soft skills (Isnaeni et al., 2019). Several organizations, including the Partnership for 21st-Century skills (2015), Assessment and Teaching of 21st-Century Skills (Binkley et al., 2012), and the Organization for Economic Co-operation and Development (OECD, 2017), refer to these competencies as 21st century skills. Although each organization has its way of defining and categorizing these skills, they commonly emphasize four essential abilities: communication, collaboration, creativity, and critical thinking, often referred to as the 4Cs. The 4Cs are important skills for education and employment, vital for success in the present and the future (Herlinawati et al., 2024; Thornhill-Miller et al., 2023). The Partnership for 21st-Century Skills (2015) refers to the 4Cs as "super skills" because they are identified as necessary for individuals to contribute effectively to a nation's progress and prosperity.

Among the 4Cs, creative thinking skills are frequently researched. Creative thinking refers to original thinking that leads to new ideas or concepts that hold value within their context (Robson, 2014; Runco & Jaeger, 2012). Creativity is increasingly valued for its ability to foster innovation, solve complex problems, and generate unique solutions. Many researchers rank creative thinking skills among the top skills needed by today's and future workers (Marbach-Ad et al., 2019; Mark et al., 2018; Prinsley & Baranyai, 2015). This ranking reflects the importance of creative thinking in addressing ever-evolving challenges and the need for innovative thinking across various industries.

Enhancing and developing students' creative thinking skills is important. These skills have been regarded as a crucial foundation for students (Rodríguez et al., 2019; Srikongchan et al., 2021; Willemsen et al., 2023). Students who think scientifically must also be able to think creatively. Students with motivation, curiosity, and high imagination tend to be more creative (Mahama


✉ firdaus@unsiq.ac.id ✉ wiyanto@mail.unnes.ac.id **(*Correspondence)** ✉ ngurahmade.dp@mail.unnes.ac.id
✉ wiwiisna@mail.unnes.ac.id

**Contribution to the literature**

- This study contributes to developing creative thinking skills instruments in scientific and digital contexts.
- The empirical results of this study indicate that the creative thinking skills instrument has high validity and reliability.
- The creative thinking skills assessment results can help educators provide targeted feedback and design learning that strengthens students' creative thinking skills.

et al., 2023). They can express their creativity according to their talents and abilities if they possess strong creative thinking skills. As a result, they can solve problems and improve their future quality of life. Teaching creative thinking skills to students is vital in preparing them to excel in the workplace, personal life, and society (Marbach-Ad et al., 2019).

Creativity has been studied more extensively in the arts than in the science (Lehmann & Gaskins, 2019). It may be assumed that science requires more scientific activity than creative skills (Ploj Virtič, 2022). However, there is a significant relationship between creative skills and scientific activity (Yildiz & Guler Yildiz, 2021). Creativity plays an important role in scientific activity. Undergraduate students majoring in physics education must have good creativity to carry out various activities such as formulating hypotheses, seeking information from various sources, and solving problems (Daker et al., 2022). Scientific innovation and discovery require creativity to generate new and scientifically sound ideas. Thus, developing relevant measurements to evaluate creativity in the science domain, scientific creative thinking skills (SCTS), is key to understanding how scientific creativity can be enhanced through education. SCTS differs from general creativity because SCTS is concerned with creative science experiments, creative scientific discovery and problem-solving, and creative science activities (Hu & Adey, 2002). SCTS can be defined as the ability to use scientific knowledge and skills to produce specific, original products. The concepts of 'scientific' and 'creativity' are not separated but fully integrated to create a new term that accurately represents a student's creativity in and with science.

This modern era requires various scientific activities to be integrated with digital technology. These include online scientific discussions, searching for appropriate scientific references online, collaborating to complete projects through digital platforms, digital scientific presentations, and data management with digital technology (Haleem et al., 2022). Physics education students need creative skills in integrating technology into scientific activities. These skills are important for participation in the labor market and leverage (van Laar et al., 2020). van Laar et al. (2019) referred to the integration of creative thinking skills within a digital context as creative thinking digital skills (CTDS). CTDS can be considered part of digital literacy (Binkley et al., 2012). This skill demands particular attention as it is an integral component of digital competence (Ferrari, 2012). Digital skills have become important in the workforce (León-Pérez et al., 2020). In an increasingly competitive global environment, the digital skills required involve the ability to complete practical tasks and encompass broader competencies related to creating and sharing ideas or information within a digital environment.

Higher education institutions must specifically design programs to develop SCTS and CTDS (Cojocariu & Boghian, 2024; de Alencar et al., 2017; Georgiou et al., 2022). Creative thinking skills are the second most important skills that workers will need in the future (Forum, 2023). These two skills are interrelated and need to be provided to physics education students. Scientific activities are inseparable from integrating digital technology, such as collaboration to solve problems through specific digital platforms. Therefore, developing these two skills is very important, not only to create original scientific ideas but also to utilize technology in solving problems and collaborating effectively. The development of SCTS and CTDS needs to be carried out systematically and planned so that students are ready to face future job challenges.

Although the need to develop these skills is evident, measuring the creative thinking achievements of physics education students is challenging. Numerous assessment methods utilize validated, commercially available instruments to evaluate creative thinking skills (Said-Metwaly et al., 2017). Some commonly used creative thinking tests include the Wallach & Kogan creativity thinking test (Wallach & Kogan, 1965), Torrance test of creative thinking (Torrance, 2018), structure of intelligence (Guilford, 1957), creativity assessment packet (Williams, 1980), Gough's creative personality scale (Gough, 1979), and consensual assessment technique (Hennessey & Amabile, 2010). Each instrument has been used to measure creative thinking skills in different domains. However, this study will focus on creative thinking skills in the domain of scientific knowledge. Evidence suggests that if instruments are used to assess science students, the assessment context should relate to the field of science (Hong, 2013).

Meanwhile, creativity tests in the science domain include the scientific creativity test (SCT) (Hu & Adey, 2002), the divergent problem-solving ability test in science (DPAS) (Aschauer et al., 2022), and the creative scientific ability test (C-SAT) (Sak & Ayas, 2013). Each

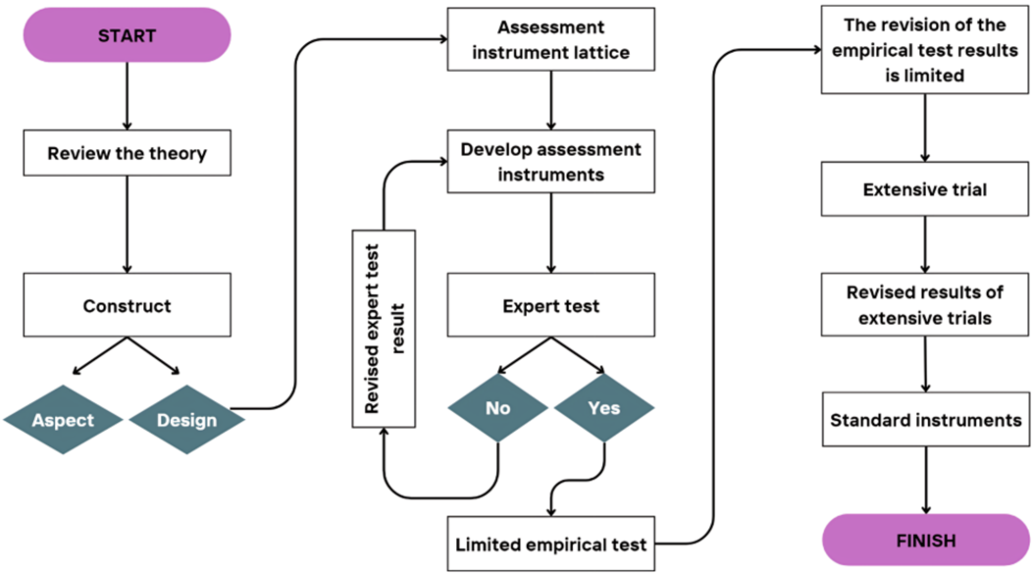



**Figure 1.** Research design (Adapted from Davis, 2012)

test is designed to measure different aspects of scientific creativity at different educational levels. All three tests are based on the concept of divergent thinking, which is the ability to generate multiple possible solutions to a problem or question (Guilford, 1957). However, the three also differ in several key aspects, including the theoretical framework for conceptualizing scientific creativity. Hu and Adey (2002) characterized the SCT as the ability to produce original products that are socially or personally valuable. Sak and Ayas (2013) designed the C-SAT, which consists of three science-related abilities: hypothesis generation, experimental design, and evaluation of evidence. The C-SAT has a specific domain focus, covering STEM topics such as biology, chemistry, physics, and ecology. Aschauer et al. (2022) designed the DPAS consisting of two subtests: divergent ideation in science tasks and divergent ideation in experimental tasks.

Although there have been various studies on creative thinking skills in the science domain, there is still a significant gap in the literature regarding developing instruments specifically designed to measure SCTS and CTDS in physics education students. The development of SCT by Hu and Adey (2002) is for the secondary school level, while SCTS will be developed for the college level. C-SAT (Sak & Ayas, 2013) was developed with STEM topics, namely biology, chemistry, physics, and ecology, while the SCTS that was developed focused on physics. DPAS is assessed based on fluency and flexibility, while SCTS is assessed based on aspects of fluency, flexibility, and originality. SCTS is conceptualized in a theoretical framework, namely the three-dimensional scientific structure creativity model (SSCM), for physics education students (Hu & Adey, 2002). The SCT developed by Hu and Adey (2002) is the beginning of research on scientific creativity as a science domain-specific creativity (Wiyanto & Hidayah, 2021).

Moreover, the CTDS instrument has not been studied much. The measurement of CTDS is still limited, although the importance of this skill in the modern workforce and digital society has been recognized (León-Pérez et al., 2020). A study on digital creativity was developed by van Laar et al. (2018) and was aimed at industrial workers. Existing research has not provided a comprehensive solution for measuring CTDS effectively, especially for physics education students who are expected to innovate using information and communication technology. CTDS was developed by referring to the instruments of van Laar et al. (2019) and Guilford (1957), as test questions for physics education students. Because of the importance of SCTS and CTDS in an increasingly digitally connected world, research related to the development of these instruments needs to continue to be carried out in order to be able to measure creative skills more accurately and relevantly according to the demands of the times.

## METHODS

### Research Design

This study used research and development methodology. This study developed two instruments: an instrument to measure SCTS and an instrument to measure CTDS. The design of this study followed the modified Borg & Gall model (Davis, 2012), as depicted in **Figure 1**.

**Figure 1** is a research design that starts with a review of the theory, developing assessment instruments and expert tests, and then moves to empirical tests to obtain valid and reliable standard instruments. The first stage is *reviewing the theory*, the terms SCTS and CTDS, and their characteristics, aspects, and standards for developing test instruments, which were examined in this stage. The *construction* stage involved determining the design and format of the SCTS and CTDS instruments. The *assessment instrument lattice* stage involves compiling an instrument lattice that describes the relationship between SCTS and CTDS indicators as measured by tests submitted to students. The next stage is the *development of assessment instruments*, where the instruments arranged in a *lattice* are developed into test instruments. The instruments are used to measure SCTS and CTDS. The *expert test* stage assessed the instrument's material, construction, and language aspects, yielding quantitative and qualitative data. Based on expert validation, the *revised expert test result* stage involved refining the instruments according to expert feedback. In the *limited empirical test* stage, the revised instruments were tested on a small group of students to obtain further feedback on the instrument's feasibility. If deficiencies are found, the instrument is revised at the *revision of the empirical test results limited stage*. Furthermore, the instruments were tested on a larger sample in the *extensive trial* stage, which aims to assess the reliability and validity of the instrument under more varied conditions. The results of these extensive trials were used to improve the revised results of the *extensive trials* stage so that the instruments are ready for widespread use. Finally, the instruments that have gone through all these stages, considering high validity and reliability, are declared to meet the standards as *standard instruments* that can be used to measure SCTS and CTDS effectively and accurately.

### Sample and Data Collection

Samples were selected using a purposive sampling technique. The first criterion was undergraduate students majoring in physics education. These students were selected based on their characteristics that met the criteria for measuring scientific creativity. The second criterion was undergraduate students taking fundamental physics courses. The material was fluid mechanics. This material was chosen because of its relevance and complexity in teaching fundamental physics. Applying fluid mechanics concepts in scientific activities allows students to develop creative ideas related to the innovation of scientific tools or experiments while assessing SCTS and CTDS. Based on these criteria, a sample of 88 undergraduate students was obtained. The sample is sufficient and representative to test its validity and reliability using Rasch model analysis and confirmatory factor analysis (CFA). This is based on the minimum number of samples, which is 50, for analysis using the Rasch model (Linacre, 1994) and CFA (Comrey & Lee, 2013; Marsh et al., 1998). The instruments used for data collection included validation sheets and the SCTS and CTDS instruments. Qualitative data from the instrument development process, such as suggestions and feedback, were used for revisions, while quantitative data from the limited and extensive empirical tests contained test results.

### Analyzing the Data

Data were analyzed through instrument validity testing using content validity with Aiken's (1985) V formula:

$$V = \frac{\sum(r-l_o)}{n(c-1)}, \quad (1)$$

where $r$ is the score given by the evaluator, $l_o$ is the lowest validity score, $n$ is the number of evaluators, and $c$ is the highest validity score.

Empirical test results were examined utilizing the product-moment correlation coefficient ($r_{xy}$) to assess validity and $r_{11}$ to evaluate the instrument's reliability (Hair et al., 2019). $r_{11}$ is a coefficient used to measure item reliability. In the Rasch model, $r_{11}$ refers to the reliability coefficient that measures the internal consistency of items in the tested instrument. The $r_{11}$ value is used to assess the extent to which the instrument measuring the SCTS and CTDS can provide consistent results among the items used in the test. A high $r_{11}$ indicates that each item in the instrument works well to measure the relevant dimensions of the SCTS and CTDS without any inconsistency. The empirical test data were evaluated utilizing the Rasch model, including item and person measures. Rasch's analysis assesses the quality of items and respondents by examining the effectiveness of items in measuring the target variables and the respondents' capacity to respond based on item difficulty (Boone et al., 2014). Item and person measures were employed to identify incongruent items (outliers or misfits) according to the following criteria: acceptable outfit mean square (MNSQ) values are defined as 0.5 < MNSQ < 1.5. Acceptable outfit z-standard (ZSTD) values range from -2.0 to +2.0, while acceptable point measure correlation (Pt Mean Corr) values range from 0.4 to 0.85 (Boone et al., 2014).

In addition to using the Rasch model, this study also used CFA. CFA was employed to evaluate the latent factor structure of the observed SCTS and CTDS variables (Mueller & Hancock, 2015). This statistical approach assesses the model's quality (goodness of fit) using Chi-square, RMSEA, GFI, SRMR, CFI, and NFI metrics. A good model fit follows the criteria: p-value (> 0.05), RMSEA (< 0.08), GFI (≥ 0.90), SRMR (≤ 0.10), CFI (≥ 0.90), and NFI (≥ 0.90) (Hooper et al., 2008). CFA guarantees that the measurement model, consisting of factors and indicators, corresponds with the foundational theory, thus improving the SCTS and CTDS instrument's construct validity.

The analysis in this study combines Rasch analysis and CFA. Both have different approaches and objectives but complement each other to evaluate the instrument comprehensively. The Rasch model evaluates the quality of items and measurements in the SCTS and CTDS instruments. Rasch measures items and individual responses more precisely by considering variability between participants. Meanwhile, CFA was used to verify whether the factor structure measured in the SCTS and CTDS instruments was by the theory underlying its measurement (Lin & Pakpour, 2017; Lin et al., 2019). The Rasch model was used to evaluate the suitability of items to student responses by looking at fit statistics, such as MNSQ, ZSTD, and Pt Mean Corr. This model helps identify items that are misfits or do not provide optimal contributions to measuring SCTS and CTDS. The Rasch model also analyzes person and item reliability, indicating the extent to which the instrument can consistently differentiate students' SCTS and CTDS levels. CFA analyzes the relationships between items and determines whether the hypothesized factor model fits the empirical data using indicators such as GFI, CFI, RMSEA, NFI, Chi-square, and SRMR. This study can ensure that the indicators in each dimension (SCTS and CTDS) truly reflect the measured constructs. The results of the Rasch model show that all items in the SCTS and CTDS are highly reliable and fit the model. CFA shows that the factor structure of both instruments is based on the empirical data, with model fit values that meet the good fit criteria.

## FINDINGS

### Review the Theory

SCTS is recognized as having the capacity to produce innovative and valuable ideas (Shi et al., 2020), engage in divergent thinking (Sun et al., 2020), identify new problems (Sternberg & Lubart, 2014), encompass novelty and value (Gruys et al., 2011). Furthermore, SCTS is grounded in theory through the three-dimensional SSCM (Hu & Adey, 2002), as in **Figure 2**.

Twenty-four cells are components of SCTS; each cell combines three dimensions: process, trait, and product. The components of the SCTS described by SSCM are shown in seven test items with mapping as in **Figure 2**. Each item can cover more than one SSCM cell, but some cells from the model are not represented. For example, combining imagination with knowledge is difficult (Hu & Adey, 2002) because imagination and knowledge are opposites. Imagination provides an opportunity to experience thinking about something that is not real or even impossible, while knowledge tends to reject it (Stoetzler & Yuval-Davis, 2002).

CTDS is the use of technology to create new or previously unknown ideas and process existing ideas with a different approach. The CTDS framework is built from indicators of generating innovative ideas, seeking information, performing tasks creatively, demonstrating the originality of ideas or work, following trends in producing innovative work and disseminating ideas (van Laar et al., 2020). Each indicator contains three

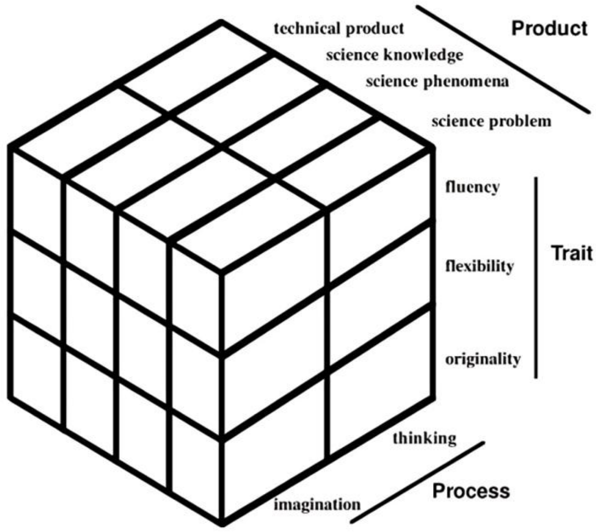



**Figure 2.** The SSCM (Adapted from Hu & Adey, 2002)

aspects of creativity: fluency, flexibility, and originality (Guilford, 1957). These three aspects are used as a reference for assessing each indicator. For example, the first indicator is generating innovative ideas.

1. **Fluency aspect:** Students use the Internet to generate ideas/concepts and are assessed by counting all answers given by students, regardless of their quality.
2. **Flexibility aspect:** Students use the Internet to generate ideas/concepts and are assessed by counting the number of approaches or areas used in the answers.
3. **Originality aspect:** Students use the Internet to generate ideas/concepts and are assessed from the frequency tabulation of all answers obtained. The lower the frequency, the more original the idea.

### Construct

The SCTS instrument covers seven aspects: unusual uses, problem finding, product improvement, scientific imagination, problem-solving, creative experimental, and creative science product design. Each aspect comprises three SSCM dimensions: process, trait, and product. The CTDS consists of six indicators adapted from van Laar et al. (2020), but it also includes a review of fluency, flexibility, and originality. Both SCTS and CTDS instruments are arranged in the form of test instruments. The types of questions for the SCTS and CTDS instruments are semi-open questions. Through this question, students are asked to answer questions with precision and structure (Glerum et al., 2014). The questions presented are not entirely open-ended but are limited to several contexts, making it easier to analyze student answers.

### Assessment Instrument Lattices

The SCTS instrument lattices are shown in **Table 1**. The SCTS instrument lattices are shown in **Table 2**.

**Table 1.** Lattices in SCTS instruments

| No | Creativity aspects | Measured dimensions | | | QN |
|---|---|---|---|---|---|
| | | Trait | Product | Process | |
| 1 | Unusual uses | Fluency, flexibility, originality | Science knowledge | Thinking | 1 |
| 2 | Problem finding | Fluency, flexibility, originality | Science problems | Thinking, imagination | 2 |
| 3 | Product improvement | Fluency, flexibility, originality | Technical product | Thinking, imagination | 3 |
| 4 | Scientific imagination | Fluency, flexibility, originality | Science phenomena | Imagination | 4 |
| 5 | Problem-solving | Flexibility, originality | Science problem | Thinking, imagination | 5 |
| 6 | Creative experimental | Flexibility, originality | Science phenomena | Thinking | 6 |
| 7 | Creative science product design | Flexibility, originality | Technical product | Thinking, imagination | 7 |

Note. QN: Question number

**Table 2.** Lattices in CTDS instruments

| No | Creativity aspects | Indicators | QN |
|---|---|---|---|
| 1 | Fluency, flexibility, originality | Using the Internet to generate ideas/concepts | 1 |
| 2 | | Finding reference sources via the Internet | 2, 3 |
| 3 | | Using the Internet to carry out or complete tasks | 4 |
| 4 | | Following trends on the Internet to produce products or works | 5 |
| 5 | | Using the Internet to evaluate your ideas or works | 6 |
| 6 | | Showing the originality of your work using the Internet | 7 |

Note. QN: Question number


INSTRUMENT FOR ASSESSING STUDENTS' SCIENTIFIC CREATIVE THINKING SKILLS (SCTS)

Information

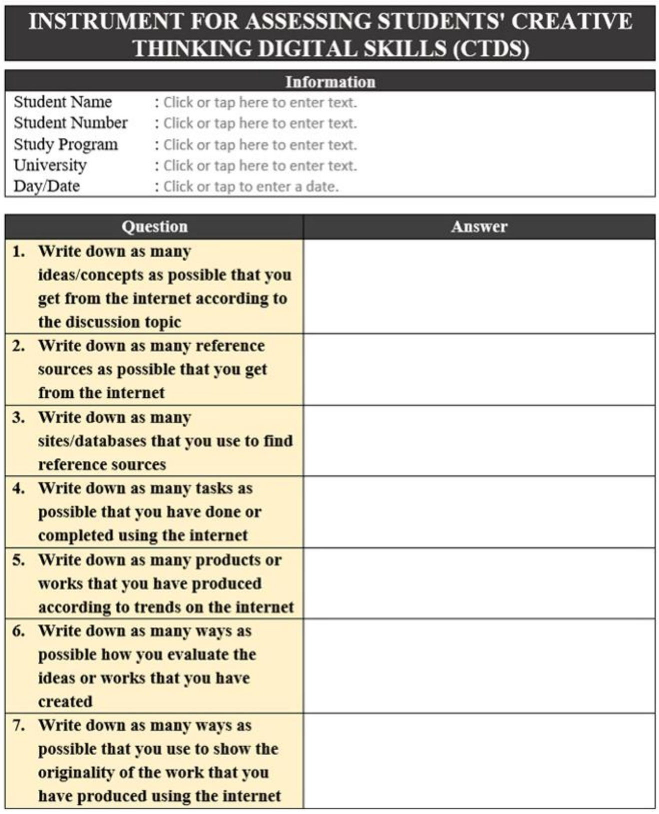
Student Name :
Student Number :
Study Program :
University :
Day/Date :

1. Write as many scientific uses of images/graphs/diagrams in learning the physics of fluid materials.
2. If you could dive deeply, what scientific questions would you want to research? Please list as many as you can.
3. The Torricelli experiment was conducted with the tool below.

Write as many improvements as possible that could be made to the tool so that the experimental activity is more interesting, more accessible, easier to understand, and the results are more valid.
4. Suppose each object has the same density; explain what phenomena will occur?
5. An object is inserted into a liquid in a floating state, as in the picture below.

Determine as many ways as possible for the object to float.
6. There are two types of liquids. Write as many ways as possible to test the viscosity of the liquid.
7. Design a carika plant watering tool. Make a picture and mention the name and function of each part.

**Figure 3.** SCTS instrument design (Source: Authors' own elaboration)


INSTRUMENT FOR ASSESSING STUDENTS' CREATIVE THINKING DIGITAL SKILLS (CTDS)

Information

Student Name : Click or tap here to enter text.
Student Number : Click or tap here to enter text.
Study Program : Click or tap here to enter text.
University : Click or tap here to enter text.
Day/Date : Click or tap to enter a date.

| Question | Answer |
|---|---|
| 1. Write down as many ideas/concepts as possible that you get from the internet according to the discussion topic | |
| 2. Write down as many reference sources as possible that you get from the internet | |
| 3. Write down as many sites/databases that you use to find reference sources | |
| 4. Write down as many tasks as possible that you have done or completed using the internet | |
| 5. Write down as many products or works that you have produced according to trends on the internet | |
| 6. Write down as many ways as possible how you evaluate the ideas or works that you have created | |
| 7. Write down as many ways as possible that you use to show the originality of the work that you have produced using the internet | |

**Figure 4.** CTDS instrument design (Source: Authors' own elaboration)

## Development of Assessment Instruments

The results of the instrument development are in the form of SCTS and CTDS test questions, which are equipped with assessment guidelines. The SCTS instrument consists of seven questions, while the CTDS consists of six questions. The SCTS instrument design is shown in **Figure 3**. The CTDS instrument is shown in **Figure 4**.

For example, question number 3 on the SCTS instrument is an SSCM cell consisting of six cells: technical product x £ fluency, flexibility, originality x thinking, and imagination.

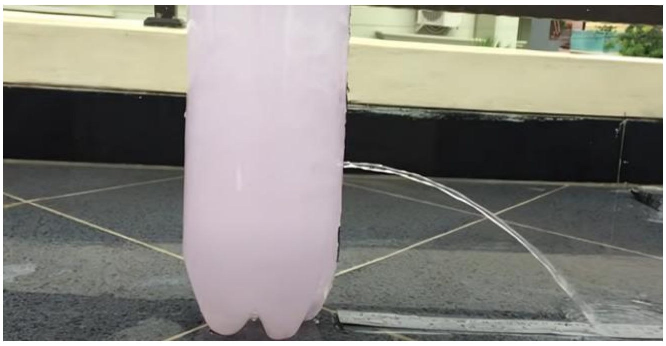


**Figure 5.** Apparatus for Torricelli's experiment (Source: Authors' own elaboration)

Torricelli's experiment was carried out using the apparatus in **Figure 5**.

> Write as many improvements as possible that could be made to the tool so that the experimental activity is more interesting, more accessible, easier to understand, and the results are more valid. For example, reducing the leak hole so that the tool meets the requirements of Torricelli's law, namely the existence of a large enough difference between the upper and lower holes.

Product improvement is vital to fostering creativity within science in the context of SSCM. Question 3, which aims to evaluate students' ability to enhance a technical product, holds significant importance. The product is the Torricelli experiment, an essential tool in fundamental physics courses. This item is evaluated based on fluency, flexibility, and originality, making it a key element in the overall assessment process.

CTDS assessment is based on the sum of fluency scores (A), flexibility (B), and originality (C).

A. Fluency assessment is seen from the number of student answers. The assessment method is to count all answers written by students. Each answer has a value of 1.

B. Flexibility assessment is seen from the number of categories/types/groups of student answers. The assessment method arranges student answers based on specific categories/types/groups and then calculates the number of categories. Each category/type/group has a value of 1.

C. Originality assessment is seen from the size of the rarity of student answers. The assessment method groups student answers, and then the answers are tabulated into a frequency distribution table. Students will receive a score of 2 if the probability of their answer is less than 5%, a score of 1 if the probability of their answer is between 5% and 10%, and a score of 0 if the probability of their answer is greater than 10% (Hu & Adey, 2002). Each question on the CTDS instrument has the same difficulty level and uses the same assessment method.

SCTS assessment refers to the assessment described by Hu and Adey (2002). The questions on the SCTS instrument have different objectives and levels of difficulty, so some questions have different assessment methods. Questions 1, 2, 3, and 4 are assessed based on fluency, flexibility, and originality, while questions 5, 6, and 7 are assessed based on flexibility and originality. *The scores of questions 1, 2, 3, and 4* are the sum of fluency, flexibility, and originality. The score calculation is as explained previously in the CTDS assessment. *The score of question 5* is the sum of flexibility and originality. The flexibility score is obtained by counting the number of categories/types/groups used in the answers. Each category/type/group has a value of 1. The originality score will receive a score of 3 if the probability of a student's answer is less than 5%, a score of 2 if the probability of their answer is between 5% and 10%, and a score of 1 if the probability of their answer is greater than 10%. *The score of question 6* is the sum of flexibility and originality. The flexibility score has a maximum value for one method of 9 (instrument: 3, principle: 3, and procedure: 3). The originality score will receive a score of 4 if the probability of a student's answer is less than 5%, a score of 2 if the probability of their answer is between 5% and 10%, and a score of 0 if the probability of their answer is greater than 10%. *The score of question 7* is the sum of flexibility and originality. The calculation of the flexibility score is a value of 3 for each function/design component. The originality score will receive a score of 5 if the probability of a student's answer is less than 5%, a score of 3 if the probability of their answer is between 5% and 10%, and a score of 1 if the probability of their answer is greater than 10%.

### Expert Test

At this stage, the SCTS and CTDS instruments were validated by five experts with minimum requirements: having a doctorate in a science field and having research related to creative thinking skills. The instruments were validated using a validation sheet consisting of three aspects, namely material, construct, and language. The total statements in the validation sheet were 12 items. The assessment results were analyzed using the Aiken' V formula. The average Aiken $V_{count}$ value calculation for the SCTS instrument was 0.90, and for the CTDS instrument, it was 0.88. The Aiken $V_{table}$ value with five raters and a scale with a range of 5 scales is 0.80. When compared between $V_{count}$ and $V_{table}$, $V_{count}$ is greater than $V_{table}$. We can conclude that the SCTS and CTDS instruments developed are valid.

### Revised Expert Test Result

The SCTS and CTDS instruments were revised based on expert input at this stage. The first revision was the addition of images to question number 3 of the SCTS instrument. The addition of images aims to clarify information so students can understand the teaching

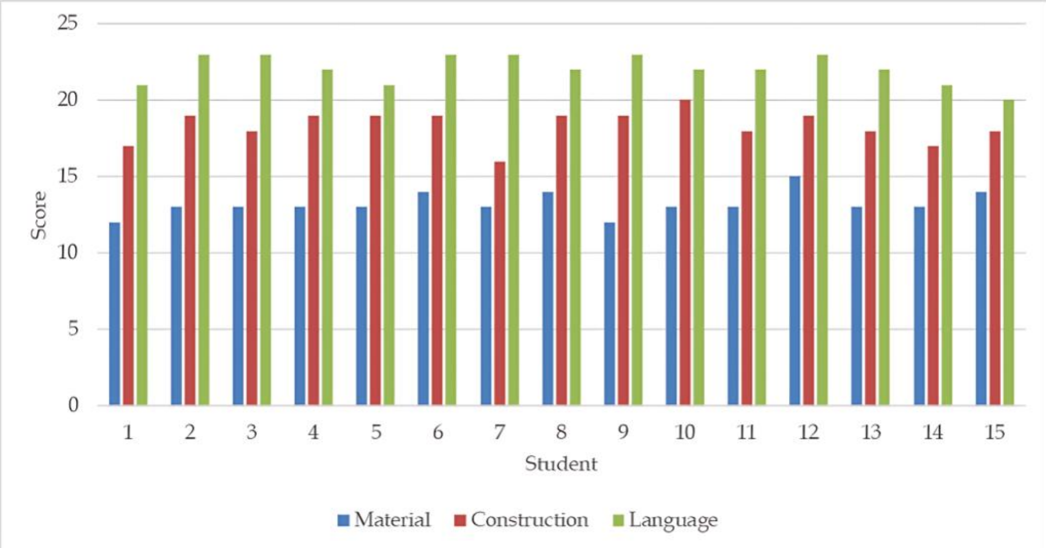



**Figure 6.** Student assessment of the SCTS and CTDS instruments (Source: Authors' own elaboration)

aids. The second revision was a change in the sentence structure of question number 2 of the SCTS instrument. The revision aims to make the questions more explicit and not ambiguous. The third revision improved the sentence structure of the CTDS instrument. The revision aims to clarify the question so that students better understand the intent of the question. The fourth revision was an improvement in the CTDS instrument assessment guidelines. The assessment score is based on three aspects, namely fluency, flexibility, and originality.

### Limited Empirical Test

The instrument assesses SCTS and CTDS in important physics learning at this stage. Students are asked to work on test questions using the SCTS and CTDS instruments. Then, students are asked to provide responses or assessments of the SCTS and CTDS instruments. The average score for the material aspect is 13.2, with a maximum score of 15. The average score for the construction aspect is 18.3, with a maximum score of 20. The average score for the language aspect is 22.1, with a maximum score of 25. The results of student assessments of the SCTS and CTDS instruments are presented in **Figure 6**.

### The Revision of the Empirical Test Results Limited

The limited empirical trial stage produced good quantitative values, and there was no input or suggestions in qualitative form. No revisions were made to the SCTS and CTDS instruments at this stage.

### Extensive Trial

The extensive test stage was conducted on 88 students taking the Basic Physics I course. The lectures were held over four sessions using a blended learning approach, which integrates traditional in-class instruction with online learning. After the lessons, students were tasked with answering questions using the SCTS instrument and completing a questionnaire based on the CTDS instrument.

**Table 3.** Statistical item reported

| Number item | Measure | Outfit MNSQ | Outfit ZSTD | Pt Mean Corr |
|---|---|---|---|---|
| SCTS instruments | | | | |
| SCTS2 | 1.61 | 1.19 | 1.13 | 0.65 |
| SCTS1 | 0. 31 | 0.98 | -0.08 | 0.78 |
| SCTS3 | 0.15 | 1.14 | 0.87 | 0.79 |
| SCTS5 | 0.15 | 0.81 | -1.16 | 0.84 |
| SCTS6 | 0.07 | 0.93 | -0.38 | 0.81 |
| SCTS4 | -1.03 | 1.29 | 1.75 | 0.72 |
| SCTS7 | -1.26 | 0.73 | -1.81 | 0.84 |
| CTDS instruments | | | | |
| CTDS1 | 0.72 | 1.26 | 1.54 | 0.75 |
| CTDS4 | 0.57 | 0.90 | -0.60 | 0.76 |
| CTDS2 | 0.33 | 0.79 | -1.35 | 0.84 |
| CTDS3 | -0.25 | 0.83 | -1.04 | 0.82 |
| CTDS5 | -0.38 | 1.01 | 0.09 | 0.83 |
| CTDS7 | -0.40 | 0.96 | -0.22 | 0.79 |
| CTDS6 | -0.58 | 1.21 | 1.27 | 0.81 |

**Table 3** shows the measurement results of each MNSQ outfit item, ZSTD outfit, and Pt Mean Corr, which are in the range indicating good fit. MNSQ scores within the acceptable range indicate that the item fits the Rasch model and measures the skill consistently as expected. Item fit indicates that the instrument is reliable in measuring the intended construct. ZSTD is used to check whether the item has problems related to response patterns that are inconsistent with the model. If the ZSTD score falls outside the acceptable range, the item must be revised or removed to improve the instrument's validity. A high Pt Mean Corr score indicates that the item is relevant and consistently measures the desired skill. Conversely, a low score indicates that the item does not fit or is ineffective in measuring what the instrument is intended to measure.

**Table 4.** Summary of measured items

| | Total | | Model | | Infit | | Outfit | |
|---|---|---|---|---|---|---|---|---|
| | Score | Count | Measure | SE | MNSQ | ZSTD | MNSQ | ZSTD |
| SCTS instruments | | | | | | | | |
| Mean | 341.6 | 75.0 | 0.00 | 0.13 | 0.97 | -0.19 | 1.01 | 0.05 |
| P.SD | 55.0 | 0.0 | 0.88 | 0.01 | 0.19 | 1.16 | 0.19 | 1.19 |
| S.SD | 59.4 | 0.0 | 0.95 | 0.01 | 0.20 | 1.25 | 0.20 | 1.28 |
| Real RMSE | 0.13 | | Separation | | 6.65 | Reliability | | 0.98 |
| Mode RMSE | 0.13 | | Separation | | 6.65 | Reliability | | 0.98 |
| CTDS instruments | | | | | | | | |
| Mean | 305.3 | 75.0 | 0.00 | 0.15 | 1.01 | .03 | 0.99 | -0.04 |
| P.SD | 22.5 | 0.0 | 0.20 | 0.00 | 0.18 | 1.06 | 0.17 | 1.02 |
| S.SD | 24.3 | 0.0 | 0.49 | 0.00 | 0.19 | 1.14 | 0.18 | 1.10 |
| Real RMSE | 0.15 | | Separation | | 3.02 | Reliability | | 0.90 |
| Mode RMSE | 0.15 | | Separation | | 3.15 | Reliability | | 0.91 |

All fit indices on the SCTS instrument are within the accepted range. The item fit limit for the MNSQ outfit score is 0.73 < MNSQ < 1.29, which indicates that all items are in the acceptable range. The MNSQ value of 0.73 indicates a variation of 23% less than the model, while the value of 1.29 indicates a variation of 29% more than expected by the model. Since the tolerance limit for model fit is 50% (Bond & Fox, 2015), this indicates that the MNSQ measurement results in this study have an item fit with the model. ZSTD Outfit scores are in the range of -1.81 < ZSTD < 1.75. This indicates that all items based on ZSTD are fit because the acceptance limit is -2.0 < ZSTD < 2.0 (Bond & Fox, 2015). SCTS Pt Mean Corr score is in the range of 0.65 < Pt Mean Corr < 0.84. All items meet the suitability criteria because the acceptance limit is 0.4 < Pt Mean Corr < 0.85 (Bond & Fox, 2015). In the CTDS instrument, the item fit limit for the MNSQ outfit score is 0.79 < MNSQ < 1.26, which shows that all items are in the acceptable range. ZSTD Outfit scores are in the range of -1.35 < ZSTD < 1.54. This shows that all items based on ZSTD are fit. The Pt Mean Corr score is in the range of 0.75 < Pt Mean Corr < 0.84. This shows that all items meet the item suitability criteria.

Item quality can be based on item reliability scores. Item reliability scores can be seen in **Table 4**.

The item reliability in **Table 4** shows a figure of 0.98 for the SCTS Instrument and 0.91 for the CTDS Instrument. The separation is 6.65 for the SCTS Instrument and 3.15 for the CTDS instrument. This indicates that the instrument items demonstrate significant stability and consistency, with questions distributed across the sample measurement on a linear interval scale. Reliability is quantified on a scale from -1.00 to +1.00, where a higher coefficient signifies greater reliability (Hair et al., 2019). The reliability scores in **Table 4**, which are 0.98 and 0.91, indicate that the instrument used exhibits a high level of consistency. This suggests that the SCTS and CTDS instruments can consistently measure research subjects and yield similar results when applied repeatedly to the same or similar subjects. High reliability correlates with minimal measurement error in obtaining results. The greater the instrument's reliability, the smaller the measurement error. (Bartlett & Frost, 2008).

Person reliability represents a score indicating the consistency of student responses and the interaction between question items and their responses (Boone, 2016). **Table 5** presents the person reliability for the SCTS and CTDS instruments.

**Table 5.** Summary of measured person

| | Total | | Model | | Infit | | Outfit | |
|---|---|---|---|---|---|---|---|---|
| | Score | Count | Measure | SE | MNSQ | ZSTD | MNSQ | ZSTD |
| SCTS instruments | | | | | | | | |
| Mean | 31.9 | 7.0 | -0.25 | 0.42 | 0.99 | -0.33 | 1.01 | -0.30 |
| P.SD | 8.0 | 0.0 | 1.42 | 0.04 | 0.91 | 1.64 | 0.92 | 1.66 |
| S.SD | 8.0 | 0.0 | 1.43 | 0.04 | 0.92 | 1.65 | 0.92 | 1.67 |
| Real RMSE | 46.0 | | Separation | | 2.70 | Reliability | | 0.88 |
| Mode RMSE | 16.0 | | Separation | | 3.22 | Reliability | | 0.91 |
| Cronbach's alpha (KR-20) person raw score "test" reliability = 0.89 | | | | | | | | |
| CTDS instruments | | | | | | | | |
| Mean | 28.5 | 7.0 | 0.17 | 0.48 | 0.99 | -0.18 | 1.01 | -0.18 |
| P.SD | 7.2 | 0.0 | 1.61 | 0.03 | 0.78 | 1.34 | 0.19 | 1.34 |
| S.SD | 7.2 | 0.0 | 1.62 | 0.03 | 0.78 | 1.35 | 0.20 | 1.35 |
| Real RMSE | 0.55 | | Separation | | 2.75 | Reliability | | 0.88 |
| Mode RMSE | 0.48 | | Separation | | 3.17 | Reliability | | 0.91 |
| Cronbach's alpha (KR-20) person raw score "test" reliability = 0.90 | | | | | | | | |

**Table 6.** DIF

| Summary DIF Chi-squared | df | Probility | Item | |
|---|---|---|---|---|
| | | | Number | Name |
| SCTS instrument | | | | |
| 0.4430 | 2 | 0.8015 | 1 | SCTS1 |
| 1.1961 | 2 | 0.5470 | 2 | SCTS2 |
| 0.7608 | 2 | 0.6821 | 3 | SCTS3 |
| 0.8955 | 2 | 0.6370 | 4 | SCTS4 |
| 0.3587 | 2 | 0.8365 | 5 | SCTS5 |
| 1.8665 | 2 | 0.3896 | 6 | SCTS6 |
| 2.4315 | 2 | 0.2930 | 7 | SCTS7 |
| CTDS instrument | | | | |
| 6.3134 | 2 | 0.0616 | 1 | CTDS1 |
| 1.5548 | 2 | 0.4561 | 2 | CTDS2 |
| 0.6899 | 2 | 0.7070 | 3 | CTDS3 |
| 0.5892 | 2 | 0.7441 | 4 | CTDS4 |
| 0.2467 | 2 | 0.8854 | 5 | CTDS5 |
| 2.4928 | 2 | 0.2841 | 6 | CTDS6 |
| 2.6401 | 2 | 0.2637 | 7 | CTDS7 |

**Table 5** shows the reliability of person in 0.91 for the SCTS and CTDS instruments. This indicates that student response consistency is relatively high. **Table 5** demonstrates the interaction between person and item with Cronbach's alpha scores of 0.89 for the SCTS instrument and 0.90 for the CTDS instrument. These scores are strong, falling within the acceptable range of 0.70 to 0.95 (DeVellis & Thorpe, 2003). The high interaction between individuals and question items suggests that students show a high level of consistency, and the items effectively measure individuals across a range of abilities from low to high.

Instrument validity can be based on differential item functional (DIF). DIF provides information about items in the instrument that favor certain groups (Andrich & Marais, 2019). The probability score becomes the standard DIF on this instrument, which is presented in **Table 6**. The quality of the test items can be seen by detecting bias in the test items. In the Rasch model, the DIF score functions to detect test item bias. An item is considered biased when it gives an advantage to a particular group based on gender, regional origin, or other demographic characteristics (Martinková & Drabinová, 2019 Martinková et al., 2017). The minimum standard for biased test items is based on the probability of the test item being more significant than 0.05.

**Table 6** shows that all items from the SCTS and CTDS instruments have probabilities above 0.05. This shows that no bias was detected in the test items. All items can measure participants' abilities from various groups without favoring or benefiting certain groups.

CFA is used to examine measurement invariance, which is the ability of an instrument to produce consistent results across groups or conditions (Milfont & Fischer, 2010). CFA also helps in confirming the latent factor structure of the measured items. The validity criteria in this calculation use factor loading. An item is considered valid if its factor loading value exceeds 0.30 (Hair et al., 2019). Based on the calculation shown in **Figure 7**, all factor loadings are more than 0.30, indicating that all items in the model have met the established validity criteria. The CFA results indicate that all items in this instrument are valid based on the factor loading criteria, and this model is suitable for use in empirical measurement.

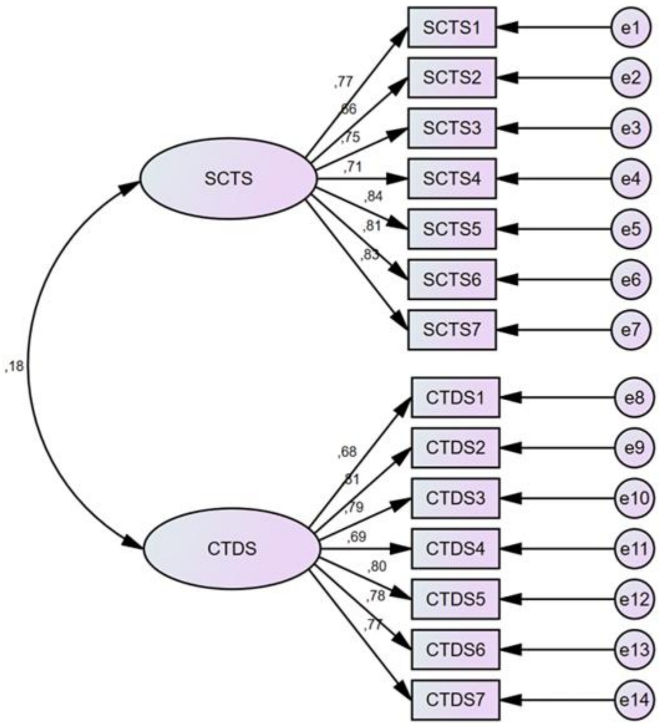



**Figure 7.** Factor structure and standardized factor loadings (Source: Authors' own elaboration)

Based on the results of data analysis in **Table 7**, the statistical values and model fit criteria indicate that this model is following the empirical data. A chi-square statistic with a p-value of 0.12 indicates that the model is not significantly different from the perfect model. RMSEA of 0.03 and SRMR of 0.06 indicate that the model has a good level of approximation. In addition, GFI of 0.98, CFI of 0.98, and NFI of 0.95 indicate that the model is a good fit for the data. These fit criteria indicate that the theoretical model tested through CFA is a good fit with the empirical data. This indicates that the model can accurately represent the relationship between variables, according to the theory hypothesized in the study (Hooper et al., 2008).

## Revised Results of Extensive Trials

Our comprehensive trial has confirmed that the developed SCTS and CTDS instruments accurately measure creative thinking skills and do not require revision. This indicates that the instruments have met the expected validity and reliability criteria, providing a solid foundation for their use.

**Table 7.** Model measurements

| Criteria | p-value | RMSEA | GFI | SRMR | CFI | NFI |
|---|---|---|---|---|---|---|
| Value | 0.12 | 0.03 | 0.98 | 0.06 | 0.98 | 0.95 |
| Criteria "fit" | > 0.05 | < 0.08 | ≥ 0.90 | ≤ 0.10 | ≥ 0.90 | ≥ 0.90 |
| Result | Fit | Fit | Fit | Fit | Fit | Fit |

### Standard Instruments

Each SCTS and CTDS instrument consists of 7 question items, each with an assessment rubric. The assessment rubric includes three aspects: fluency, flexibility, and originality. The instrument has been declared valid by experts (covering aspects of material, construction, and language) and through testing the validity of the test items at the limited empirical test stage and wider trials. In addition, this instrument also shows high reliability based on the trial results. Based on the Rasch model test results, the SCTS and CTDS instruments have question item suitability according to the expected model. Based on the CFA results, the SCTS and CTDS instruments have data that aligns with the expected theoretical construction. The SCTS and CTDS instruments can be categorized as standard instruments widely used to assess creative thinking skills in various contexts, both in the scientific and digital realms.

## DISCUSSION

This study examines the SCTS and CTDS instruments analyzed using the Rasch and CFA models (Lin et al., 2019; Lin & Pakpour, 2017). This instrument was developed to measure students' creative thinking skills, where they are free to choose, prepare, and present their ideas in the form of words. The main advantage of this instrument is its ability to explore students' creative abilities because the essay format allows them to express more complex and original ideas (Kubiszyn & Borich, 2013). Through the Rasch model, the validity and reliability of the instrument can be analyzed in more detail, ensuring that each question item can measure the desired aspects consistently (Boone et al., 2014). CFA was used to ensure the suitability of the theoretical model with empirical data, strengthening the instrument's construct validity (Hooper et al., 2008). Combining these two approaches helps ensure that the SCTS and CTDS instruments can effectively measure creative thinking skills in scientific and digital fields.

The characteristics of a good instrument must meet the validity aspect, which ensures that the instrument measures what it is supposed to measure (Berkowitz et al., 2012; Gable & Wolf, 1993). The SCTS and CTDS instruments have gone through a validation process by experts. Experts assess the instrument from the material, construct, and language aspects to ensure that each instrument item follows the desired measurement objectives and can be well understood by respondents (Sullivan, 2011). This validation process was carried out using V-Aiken's formula calculation, a statistical method for measuring item validity based on the assessment of a panel of experts (Aiken, 1985). The analysis results show that the SCTS and CTDS instruments developed have achieved adequate validity. This instrument is a valid measuring tool to assess students' creative thinking skills effectively and accurately.

Instruments are also analyzed based on empirical data to ensure the reliability and accuracy of their measurements (Falcão et al., 2023; O'Leary-Kelly & Vokurka, 1998). The SCTS and CTDS instruments were analyzed using the Rasch model. The instrument's reliability was measured by indicators such as MNSQ, Z-STD, and Pt Mean Corr (Boone et al., 2014). The values of the three indicators are within the expected range, so it can be concluded that the instrument has good quality in measuring creative skills empirically. In addition, the SCTS and CTDS instruments show a high level of reliability, both in terms of items and persons. Item reliability shows how consistently the test items measure the same concept, while person reliability measures the consistency of results among different participants (Bartlett & Frost, 2008). This is very important in measuring creative skills because data consistency shows that the instrument can objectively assess creative thinking skills across groups of participants. The Cronbach's alpha scores obtained are within the acceptable range. This indicates a strong interaction between person and item, with high consistency of student answers (Taber, 2018). The instrument's question items can accurately measure participants' abilities, both for those with low and high abilities. DIF analysis was also conducted to detect potential bias, which refers to any systematic difference in the performance of different groups of participants in the SCTS and CTDS instruments (Andrich & Marais, 2019). The analysis showed that no bias was detected, meaning that all question items could measure participants' abilities without favoring certain groups based on characteristics such as gender, regional origin, or other demographic factors (Martinková et al., 2017). With a probability above 0.05, this instrument is proven to be fair and valid for use in a broad assessment context.

In addition to being analyzed using the Rasch model, the SCTS and CTDS instruments were also analyzed using the CFA method. The results of the CFA showed that each item in the two instruments was closely related, and the dimensions of the SCTS and CTDS were consistently measured. This CFA is important because it ensures that the instrument measures specific aspects of creativity without interference from other irrelevant factors. In the SCTS instrument, the ability to think creatively in a scientific context is the main focus of measurement. At the same time, CTDS is designed to assess creativity in using digital technology. With the CFA results describing a strong relationship between items, this instrument can be relied on to provide valid and accurate assessment results in comprehensively evaluating creative thinking skills in various scientific and digital contexts. This entire process ensures that the SCTS and CTDS instruments effectively measure creative skills, cover various levels of participant ability, and provide in-depth insight into the dimensions of creative thinking.

The main characteristic of the SCTS instrument is its ability to measure divergent thinking, namely the ability to generate various ideas or solutions to a particular problem. Divergent thinking is one of the key indicators of creativity, where someone can see a problem from various perspectives and find innovative and original solutions (Dollinger et al., 2004). In addition, this instrument also identifies the emergence of new ideas that have never been thought of before by students, reflecting a higher level of innovation (Jones & Townley, 2016). Thus, the SCTS instrument effectively evaluates creative thinking skills in a scientific context, where students must think outside the box to generate new ideas.

On the other hand, the CTDS instrument is specifically designed to measure creativity in a digital context. This instrument assesses the extent to which participants can utilize technology to create new ideas or adapt existing ideas in more innovative and creative ways (van Laar et al., 2020). In today's digital era, thinking creatively in a digital context is very important, especially when facing ever-evolving changes and challenges. Both instruments, SCTS and CTDS, provide important insights into assessing creative skills highly relevant to modern science and technology development. Through this instrument, participants' ability to think creatively in various fields can be measured comprehensively and accurately.

## CONCLUSION

Developing instruments to measure creative thinking skills in scientific and digital contexts is important. The developed instruments are the SCTS instrument and the CTDS instrument. Each instrument consists of 7 questions. Experts have verified the validity of this instrument through the assessment of material, construct, and language aspects, as well as through testing the validity of the questions at the limited empirical test stage and extensive trials. This instrument also shows high reliability based on the trial results. The Rasch model analysis shows that the SCTS and CTDS instruments follow the expected model, reviewed from the MNSQ, ZSTD, and Pt Mean Corr indicators. In addition, the results of the DIF analysis indicate that there is no bias in both instruments. Through CFA, the SCTS and CTDS instruments also show alignment of empirical data with the expected theoretical constructs. Thus, the SCTS and CTDS instruments can be categorized as standard instruments suitable for widespread use to measure creative thinking skills in various scientific and digital contexts.

The empirical results of this study indicate that the SCTS and CTDS instruments have high validity and reliability in measuring creative thinking skills. Therefore, these instruments have various practical applications in education, especially in measuring creative thinking skills. Educators can use the SCTS instrument to elicit students' creative thinking skills in subjects involving scientific experiments and problem-solving, while the CTDS instrument can assess how students utilize digital technology to generate creative ideas. These instruments can be applied in formative assessments to reflect students' creative skills development throughout the semester or in summative assessments to produce final learning outcomes. Educators can use the results of these instruments to provide more targeted feedback and design more in-depth learning to improve students' creative thinking skills. The results of this study contribute to the development of instruments and majorly contribute to more innovative teaching and learning practices in the classroom.

Although the SCTS and CTDS instruments have been tested on physics education students, this study has limitations in generalizing the research results to other disciplines. However, because this context is limited to physics education students, the results of this study do not fully reflect the conditions in other disciplines, such as social sciences or humanities. Therefore, further trials with various disciplines are needed to improve the external validity of these instruments, as well as expand their application to broader educational fields. Further research in various disciplines will provide a more comprehensive picture of how the SCTS and CTDS instruments measure student creativity. The validity and reliability testing of this study's SCTS and CTDS instruments was conducted on a relatively limited sample. This causes limitations in the generalization of the results because the instruments need to reflect the diversity of creative thinking abilities in the broader population. Further research is needed using larger and more heterogeneous samples regarding age, gender, and educational background to improve external validity. Diversifying participants will enable a more accurate assessment of creative thinking abilities in various contexts and increase the instrument's reliability in measuring SCTS and CTDS. Research with a larger sample can help detect potential biases that may not be visible in a limited sample.

There are concerns about the application of this instrument in various countries due to cultural differences. Differences in values, norms, and perceptions between countries can affect how participants respond to the instruments we develop. Therefore, we encourage researchers in other countries to adopt this instrument according to the local cultural context to reduce the potential for bias caused by cultural differences. This adaptation can be done by adjusting the terminology or approach used in the instrument without changing the essence of the measurement. The contribution of further research is expected to provide deeper insights and enrich the literature related to SCTS and CTDS measurements.

**Author contributions: F:** concept and design, data analysis, drafting manuscript, and writing; **W:** data acquisition, supervision, critical revision of manuscript, securing funding, and final approval; **NMDP:** critical revision of manuscript, and supervision; & **WI:** critical revision of manuscript, statistical analysis, and supervision. All authors agreed with the results and conclusions.

**Funding:** No funding source is reported for this study.

**Acknowledgments:** The authors would like to thank all the experts and participants who contributed to the validation and testing of the SCTS and CTDS instruments. The authors would also like to thank the educational institutions and students who participated in the empirical trials, as well as the research team members for their support throughout the study.

**Ethical statement:** The authors stated that the study was conducted according to the highest ethical guidelines and principles of research involving human participants. Informed consent was obtained from all participants prior to their participation in the study. The authors further stated that the study was designed without personal identifiers or invasion of privacy.

**Declaration of interest:** No conflict of interest is declared by the authors.

**Data sharing statement:** Data supporting the findings and conclusions are available upon request from the corresponding author.

## REFERENCES

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

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement: Measuring in the educational, social, and health sciences*. Springer. https://doi.org/10.1007/978-981-13-7496-8

Aschauer, W., Haim, K., & Weber, C. (2022). A contribution to scientific creativity: A validation study measuring divergent problem solving ability. *Creativity Research Journal, 34*(2), 195-212. https://doi.org/10.1080/10400419.2021.1968656

Bartlett, J. W., & Frost, C. (2008). Reliability, repeatability and reproducibility: Analysis of measurement errors in continuous variables. *Ultrasound in Obstetrics and Gynecology, 31*(4), 466-475. https://doi.org/10.1002/uog.5256

Benbow, R. J., Lee, C., & Hora, M. T. (2021). Exploring college faculty development in 21st-century skill instruction: An analysis of teaching-focused personal networks. *Journal of Further and Higher Education, 45*(6), 818-835. https://doi.org/10.1080/0309877X.2020.1826032

Berkowitz, D., Caner, M., & Fang, Y. (2012). The validity of instruments revisited. *Journal of Econometrics, 166*(2), 255-266. https://doi.org/10.1016/j.jeconom.2011.09.038

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17-66). Springer. https://doi.org/10.1007/978-94-007-2324-5_2

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge. https://doi.org/10.4324/9781315814698

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

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4

Cojocariu, V. M., & Boghian, I. (2024). A literature review on digital creativity in higher education–Toward a conceptual model. *Education Sciences, 14*(11), Article 1189. https://doi.org/10.3390/educsci14111189

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press. https://doi.org/10.4324/9781315827506

Daker, R. J., Colaizzi, G. A., Mastrogiannis, A. M., Sherr, M., Lyons, I. M., & Green, A. E. (2022). Predictive effects of creative abilities and attitudes on performance in university-level computer science courses. *Translational Issues in Psychological Science, 8*(1), 104-124. https://doi.org/10.1037/tps0000293

Davis, G. A. (2012). *Gifted children and freedom education*. PT Index.

de Alencar, E. M. L. S., Fleith, D. D. S., & Pereira, N. (2017). Creativity in higher education: Challenges and facilitating factors. *Trends in Psychology, 25*(61), 553-561. https://doi.org/10.9788/TP2017.2-09

DeVellis, R. F., & Thorpe, C. T. (2003). *Scale development theory and applications*. SAGE.

Dollinger, S. J., Urban, K. K., & James, T. A. (2004). Creativity and openness: Further validation of two creative product measures. *Creativity Research Journal, 16*(1), 35-47. https://doi.org/10.1207/s15326934crj1601_4

Falcão, F., Jalali, C., & Costa, P. (2023). Empirical validation study and psychometric evaluation of the properties of the populist attitudes scale for the

Portuguese population. *BMC Psychology, 11*, Article 93. https://doi.org/10.1186/s40359-023-01118-1

Ferrari, A. (2012). *Digital competence in practice: An analysis of frameworks*. Publications Office of the European Union. https://doi.org/10.2791/82116

Forum, W. E. (2023). *Future of jobs report 2023*. World Economic Forum. https://doi.org/10.1142/11458

Gable, R. K., & Wolf, M. B. (1993). The validity of effective instruments. In *Instrument development in the affective domain. Evaluation in education and human services, vol 36* (pp. 95-200). Springer. https://doi.org/10.1007/978-94-011-1400-4_4

Georgiou, H., Turney, A., Matruglio, E., Jones, P., Gardiner, P., & Edwards-Groves, C. (2022). Creativity in higher education: A qualitative analysis of experts' views in three disciplines. *Education Sciences, 12*(3), Article 154. https://doi.org/10.3390/educsci12030154

Glerum, A., Atasoy, B., & Bierlaire, M. (2014). Using semi-open questions to integrate perceptions in choice models. *Journal of Choice Modelling, 10*(1), 11-33. https://doi.org/10.1016/j.jocm.2013.12.001

Gough, H. G. (1979). A creative personality scale for the adjective check list. *Journal of Personality and Social Psychology, 37*(8), 1398-1405. https://doi.org/10.1037/0022-3514.37.8.1398

Gruys, M. L., Munshi, N. V., & Dewett, T. C. (2011). When antecedents diverge: Exploring novelty and value as dimensions of creativity. *Thinking Skills and Creativity, 6*(2), 132-137. https://doi.org/10.1016/j.tsc.2011.01.005

Guilford, J. P. (1957). Creative abilities in the arts. *Psychological Review, 64*(2), 110-118. https://doi.org/10.1037/h0048280

Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis*. Cengage Learning. https://doi.org/10.3390/polym12123016

Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the role of digital technologies in education: A review. *Sustainable Operations and Computers, 3*, 275-285. https://doi.org/10.1016/j.susoc.2022.05.004

Hennessey, B. A., & Amabile, T. M. (2010). Creativity. *Annual Review of Psychology, 61*, 569-598. https://doi.org/10.1146/annurev.psych.093008.100416

Herlinawati, H., Marwa, M., Ismail, N., Junaidi, Liza, L. O., & Situmorang, D. D. B. (2024). The integration of 21st century skills in the curriculum of education. *Heliyon, 10*(15), Article e35148. https://doi.org/10.1016/j.heliyon.2024.e35148

Hong, E. (2013). Creative thinking abilities: Measures for various domains. In H. O'Neil, R. Perez, & E. Baker (Eds.), *Teaching and measuring cognitive readiness* (pp. 201-222). Springer. https://doi.org/10.1007/978-1-4614-7579-8_11

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60. https://doi.org/10.21427/D7CF7R

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education, 24*(4), 389-403. https://doi.org/10.1080/09500690110098912

Isnaeni, W., Budi Prasetyo, A. P., Ristianafi, D. P., & Setiawan, H. (2019). Analyzing students' critical thinking, communication, collaboration, and creativity skills during animal physiology courses. *Unnes Science Education Journal, 8*(2), 168-175. https://doi.org/10.15294/usej.v8i2.34047

Jones, C., & Townley, B. (2016). Misfits, mavericks and mainstreams: Drivers of innovation in the creative industries. *Organization Studies, 37*(6), 751-768. https://doi.org/10.1177/0170840616647671

Kubiszyn, T., & Borich, G. D. (2013). *Educational testing and measurement: Classroom application and practice*. Wiley Global Education.

Lehmann, J., & Gaskins, B. (2019). Learning scientific creativity from the arts. *Palgrave Communications, 5*(1), 4-8. https://doi.org/10.1057/s41599-019-0308-8

León-Pérez, F., Bas, M.-C., & Escudero-Nahón, A. (2020). Self-perception about emerging digital skills in higher education students. *Comunicar, 28*(62), 91-101. https://doi.org/10.3916/C62-2020-08

Lin, C. Y., Hwang, J. S., Wang, W. C., Lai, W. W., Su, W. C., Wu, T. Y., Yao, G., & Wang, J. Der. (2019). Psychometric evaluation of the WHOQOL-BREF, Taiwan version, across five kinds of Taiwanese cancer survivors: Rasch analysis and confirmatory factor analysis. *Journal of the Formosan Medical Association, 118*(1 Part 2), 215-222. https://doi.org/10.1016/j.jfma.2018.03.018

Lin, C.-Y., & Pakpour, A. H. (2017). Using hospital anxiety and depression scale (HADS) on patients with epilepsy: Confirmatory factor analysis and Rasch models. *Seizure, 45*, 42-46. https://doi.org/10.1016/j.seizure.2016.11.019

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), Article 328.

Mahama, I., Yusuf Dramanu, B., & Asamoah-Gyimah, K. (2023). Predictive abilities of curiosity, creativity, and motivation on academic performance of high school students in Ghana. *Education Research International*. https://doi.org/10.1155/2023/4626945

Marbach-Ad, G., Hunt, C., & Thompson, K. V. (2019). Exploring the values undergraduate students attribute to cross-disciplinary skills needed for the workplace: An analysis of five stem disciplines. *Journal of Science Education and Technology, 28*(5), 452-469. https://doi.org/10.1007/s10956-019-09778-8

Mark, K. P., So, J. C. H., Chan, V. C. W., Luk, G. W. T., & Ho, W. T. (2018). Surviving in the gig economy: Change of stem students' perceptions on the generic skills for the workplace. In *Proceedings of the 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering* (pp. 1085-1090). IEEE. https://doi.org/10.1109/TALE.2018.8615414

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181-220. https://doi.org/10.1207/s15327906mbr3302_1

Martinková, P., & Drabinová, A. (2019). ShinyItemAnalysis for teaching psychometrics and to enforce routine analysis of educational tests. *The R Journal, 10*(2), 503-515. https://doi.org/10.32614/RJ-2018-074

Martinková, P., Drabinová, A., Liaw, Y. L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE–Life Sciences Education, 16*(2). https://doi.org/10.1187/cbe.16-10-0307

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*(1), 111-130. https://doi.org/10.21500/20112084.857

Mueller, R. O., & Hancock, G. R. (2015). Factor analysis and latent structure analysis: Confirmatory factor analysis. In N. J. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 686-690). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.25009-5

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management, 16*(4), 387-405. https://doi.org/10.1016/s0272-6963(98)00020-5

OECD. (2017). *Pisa 2015 assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving*. OECD Publishing. https://doi.org/10.1787/9789264281820-en

Partnership for 21st-Century Skills. (2015). Partnership for 21st-century skills-core content integration. *battelle for kids*. https://www.battelleforkids.org/insights/p21-resources/

Ploj Virtič, M. (2022). Teaching science & technology: Components of scientific literacy and insight into the steps of research. *International Journal of Science Education, 44*(12), 1916-1931. https://doi.org/10.1080/09500693.2022.2105414

Prinsley, R., & Baranyai, K. (2015). STEM skills in the workforce: What do employers want? *Occasional Paper Series*, Number 9. https://doi.org/10.13140/RG.2.2.12120.60167

Robson, S. (2014). The analyzing children's creative thinking framework: Development of an observation-led approach to identifying and Analyzing young children's creative thinking. *British Educational Research Journal, 40*(1), 121-134. https://doi.org/10.1002/berj.3033

Rodríguez, G., Díez, J., Pérez, N., Baños, J. E., & Carrió, M. (2019). Flipped classroom: Fostering creative skills in undergraduate students of health sciences. *Thinking Skills and Creativity, 33*, Article 100575. https://doi.org/10.1016/j.tsc.2019.100575

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92-96. https://doi.org/10.1080/10400419.2012.650092

Said-Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: A systematic literature review. *Creativity. Theories–Research-Applications, 4*(2), 238-275. https://doi.org/10.1515/ctra-2017-0013

Sak, U., & Ayas, M. B. (2013). Creative scientific ability test (C-SAT): A new measure of scientific creativity. *Psychological Test and Assessment Modeling, 55*(3), 316-329. https://doi.org/10.1037/t72324-000

Shi, Y., Yuan, T., Bell, R., & Wang, J. (2020). Investigating the relationship between creativity and entrepreneurial intention: The moderating role of creativity in the theory of planned behavior. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.01209

Srikongchan, W., Kaewkuekool, S., & Mejaleurn, S. (2021). Backward instructional design based learning activities to developing students' creative thinking with lateral thinking technique. *International Journal of Instruction, 14*(2), 233-252. https://doi.org/10.29333/iji.2021.14214a

Sternberg, R. J., & Lubart, T. I. (2014). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 3-15). Cambridge University Press. https://doi.org/10.1017/CBO9780511807916

Stoetzler, M., & Yuval-Davis, N. (2002). Standpoint theory, situated knowledge and the situated imagination. *Feminist Theory, 3*(3), 315-333. https://doi.org/10.1177/146470002762492024

Sullivan, G. M. (2011). A Primer on the validity of assessment instruments. *Journal of Graduate Medical Education, 3*(2), 119-120. https://doi.org/10.4300/jgme-d-11-00075.1

Sun, M., Wang, M., & Wegerif, R. (2020). Effects of divergent thinking training on students' scientific creativity: The impact of individual creative potential and domain knowledge. *Thinking Skills and Creativity, 37*, Article 100682. https://doi.org/10.1016/j.tsc.2020.100682

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

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), Article 54. https://doi.org/10.3390/jintelligence11030054

Torrance, E. P. (2018). *Torrance tests of creative thinking*. APA PsycTests. https://doi.org/10.1037/t05532-000

Valtonen, T., Hoang, N., Sointu, E., Näykki, P., Virtanen, A., Pöysä-Tarhonen, J., Häkkinen, P., Järvelä, S., Mäkitalo, K., & Kukkonen, J. (2021). How pre-service teachers perceive their 21st-century skills and dispositions: A longitudinal perspective. *Computers in Human Behavior, 116*, Article 106643. https://doi.org/10.1016/j.chb.2020.106643

van Laar, E., van Deursen, A. J. A. M., Dijk, J. A. G. M., & de Haan, J. (2019). The sequential and conditional nature of 21st-century digital skills. *International Journal of Communication, 13*(409), 3462-3487. https://doi.org/10.1037/t73734-000

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2018). 21st-century digital skills instrument aimed at working professionals: Conceptual development and empirical validation. *Telematics and Informatics, 35*(8), 2184-2200. https://doi.org/10.1016/j.tele.2018.08.006

van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2020). Determinants of 21st-century skills and 21st-century digital skills for workers: A systematic literature review. *SAGE Open, 10*(1). https://doi.org/10.1177/2158244019900176

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction*. Rinehart & Winston.

Willemsen, R. H., de Vink, I. C., Kroesbergen, E. H., & Lazonder, A. W. (2023). The role of creative thinking in children's scientific reasoning. *Thinking Skills and Creativity, 49*, Article 101375. https://doi.org/10.1016/j.tsc.2023.101375

Williams, F. (1980). *Creativity assessment packet*. DOK.

Wiyanto, W., & Hidayah, I. (2021). Review of a scientific creativity test of the three-dimensional model. *Journal of Physics: Conference Series, 1918*, Article 052088. https://doi.org/10.1088/1742-6596/1918/5/052088

Yildiz, C., & Guler Yildiz, T. (2021). Exploring the relationship between creative thinking and scientific process skills of preschool children. *Thinking Skills and Creativity, 39*, Article 100795. https://doi.org/10.1016/j.tsc.2021.100795