

RESEARCH

Open Access



Development and validation of an instrument for measuring civic scientific literacy

Yuemei Liu¹, Jian Wang^{1*}, Zhimin Zhang², Jingchun Wang^{2*}, Tian Luo³, Shan Lin¹, Jiaxin Li⁴ and Sumeng Xu¹

Abstract

Civic scientific literacy (CSL) is a fundamental indicator of social development and national literacy and is very important for personal well-being and national competitiveness. To be in line with the technological advancements and demands of the current era, we focused on “science engagement”, which is now considered an important approach to enhance the CSL of the general public, and developed a 4-dimensional CSL assessment framework based on the crucial elements for citizens to engage in science. The 4 dimensions are scientific knowledge (SK), scientific method (SM), problem-solving (PS), and scientific thought and spirit of science (STSS). With this framework, a CSL measurement instrument including 50 items was developed. Various validity evidence, including evidence with reliability close to or exceeding 0.8 for all four dimensions, supports the use of the instrument to provide a valuation of CSL. Age, location type (urban or rural), educational background, and occupation are all significant factors associated with CSL.

Keywords Civic scientific literacy, Instrument development and validation, Rasch analysis

Introduction

Scientific literacy (SL) is a constantly evolving concept in both theory and practice (National Academies of Sciences, Engineering, and Medicine, 2016). Since its proposal in 1958, SL has been constantly redefined (Council of Canadian Academies, 2014; DeBoer, 2000; Miller, 1983; National Academies of Sciences, Engineering, and Medicine, 2016; Norris & Phillips, 2003; OECD, 2017; Rutherford & Ahlgren, 1991; Shen, 1975). Although SL

has received widespread attention and recognition as an important goal in education (especially in science education) (Council of Canadian Academies, 2014; National Research Council, 1996; OECD, 2017; Rutherford & Ahlgren, 1991), there is currently no consensus on a definition of SL (DeBoer, 2000).

To assist in conceptualizing the variety of SL definitions, Roberts proposed two competing visions of SL. Vision I points to the scientific inner field from the perspective of scientists, emphasizing the disciplinary nature of science itself, including its products, processes, and characteristics. Vision II places more emphasis on the role of science in human affairs, including scientific thinking and activities (Roberts, 2010). Two visions represent two types of cultivation objectives. From the perspective of personal and social development needs, what we need is more people with SL, rather than dedicated to cultivating everyone into scientists. There is an undeniable huge gap between the general public and scientists (X. Liu, 2009). Liu believes that bridging the gap between

*Correspondence:

Jian Wang

wangj@bnu.edu.cn

Jingchun Wang

wangjingchun@cstm.org.cn

¹ College of Life Sciences, Beijing Normal University, Beijing 100875, China

² China Research Institute for Science Popularization, Beijing 100081,

China

³ College of Teacher Education, Capital Normal University, Beijing 100048,

China

⁴ The Affiliated High School of South China Normal University,

Guangzhou 510630, China

the scientific society and the general public requires every citizen to participate in issues related to science (Liu, 2013). He suggests incorporating “science engagement” (SE) into the conceptual framework of SL, viewing it as vision III that transcends vision I and vision II. SE emphasizes the introduction of social, cultural, political, and environmental issues, aiming to cultivate the critical thinking, scientific communication, and consensus-building abilities of every citizen (Liu, 2013). The three visions are interdependent and mutually reinforcing, but vision III emphasizes active participation and seeking solutions to the urgent problems facing the world today (Liu, 2013). The three visions provide us with references for clarifying different SL definitions and evaluating SL with different goal orientations.

Due to the different conceptual connotations of SL, there are also differences in the measurement elements of SL (Fives et al., 2014; Gormally et al., 2012; Miller, 1998; Naganuma, 2017; OECD, 2017). Among them, the civic scientific literacy (CSL) assessment for Miller among adults is widely concerned and used worldwide. Based on the three kinds of science literacy (practical, civic, and cultural) proposed by Shen (Shen, 1975), from a practical perspective, Miller focuses on CSL and believes that CSL is a knowledge level that can support individuals in reading current scientific news, reports from different formal channels, or watch scientific television programs, instead of a scientific understanding of the field of science or engineering (Miller, 1983). He proposes three dimensions of CSL: scientific terminology and perspectives, scientific principles and methods, and the impact of science and society (Miller, 1983). Based on this, the developed CSL assessment tool (Miller, 1998) is widely used in over 40 countries and organizations worldwide. The SL assessment for school-age adolescents mainly includes two major international projects: Trends in International Mathematics and Science Study (TIMSS) and the Program for International Student Assessment (PISA). TIMSS does not specify the general concept of SL but emphasizes that students should be able to take action on a solid scientific foundation when facing various problems, to meet the requirements of the technological society for people’s abilities and higher levels of learning (Jones et al., 2015). TIMSS is guided by the school’s science curriculum and evaluates science content, science significance, and science practices (Jones et al., 2015). All projects are textbook questions without context. PISA has clarified and continuously iterated the definition and assessment framework of SL (OECD, 2019, 2017, 2023). According to PISA2015, SL refers to the reflective citizen’s capacity to engage in discussions about scientific issues, offer scientific insights, and further interpret it as three competencies: the competency

to scientifically explain phenomena scientifically, evaluate and design scientific inquiry, and interpret data and evidence scientifically (OECD, 2017). PISA is oriented towards social needs and develops testing items based on specific contexts. So far, PISA has drawn participation from an increasing number of nations and areas.

From the perspective of three visions, Miller’s CSL assessment, which mainly examines scientific knowledge, and TIMSS, which emphasizes the content of the scientific field, are consistent with vision I (Liu, 2013; Naganuma, 2017), while PISA, which focuses on different scales of contexts, is mainly vision II (Liu, 2013). The evaluation of SL is not centered around SE. In today’s era, with the rapid expansion of information, the quick development of technology, and the continuous prominence of socio-scientific issues such as global warming and genetically modified foods, higher demands have been placed on citizens’ SE. The newly included science identity in the recently released PISA2025 Science Framework (OECD, 2023) emphasizes a special focus on sustainable development and environmental education. It can be foreseen that in the future, there will be widespread attention to SL in the context of sustainability. Therefore, the conceptualization of CSL should be further developed. It is necessary to emphasize SE and sustainability in CSL.

Based on the above considerations, we adopt the definition of “to understand the necessary scientific knowledge, master the basic scientific method, establish scientific thought, advocate for the spirit of science, and have the ability to apply the above to practical problems and participate in public affairs” as defined by the State Council of PRC (The State Council of PRC, 2006). This definition focuses on SE and emphasizes sustainable development in scientific thought. The participation and resolution of issues is an important content and objective of SE. However, we realized that, in the process of SE, the scientific knowledge and methods widely concerned by existing SL assessments are necessary but insufficient, and it is also necessary to understand the nature of science. At the same time, the common good qualities in scientific research, such as curiosity and thirst for knowledge, pursuit of truth and innovation, play an important role in arousing public interest in SE and effective problem-solving. More significantly, more consideration needs to be given to whether the existing approaches to challenging issues involving the integration of several disciplines, including science, society, and the environment, satisfy the criteria of sustainable development. The meaning and efficacy of SE depend on whether it satisfies the demands of sustainable development. In the current complex socio-scientific context, all these elements that play an important role in SE mentioned above should be included in the framework of CSL.

However, only a portion of them (such as scientific knowledge and methods) has been given attention in former CSL assessments (Miller, 1998), while important content such as spirit of science and scientific thought of Scientific Outlook on Development are less present in existing CSL evaluation frameworks. Therefore, it is necessary to develop a CSL assessment instrument for adults based on vision III. This instrument fills in the key gaps left by the current assessments regarding effective SE by citizens. Further research is required in the development and design of instruments given the current emphasis on competency and situational literacy needs.

In research on the evaluation of CSL, researchers have extensively focused on the factors associated with CSL. Past research findings indicate that urbanization, gender, and education have significant effects on CSL (Miller, 1998; Ren et al., 2013). For instance, by examining data from a 2016 survey of CSL in the United States, Miller found that educational attainment and exposure to college science courses were the two best predictors of CSL. He also noted that men, as compared to women, are more likely to have higher literacy levels (Miller, 2016). According to data collected on CSL among Chinese citizens, education, economy, and urbanization can all strongly influence the level of CSL (Tian & Dai, 2018). The data also revealed considerable gender (Zhang et al., 2016) and geographic (Zhang et al., 2013) disparities in CSL. According to Ren et al. (Ren et al., 2013), the difference in education systems and curricula between the United States and China affects their citizens' CSL. An international comparative study revealed differences in CSL among individuals of different ages and occupations (Wu et al., 2018). Due to the differences in the content and dimensions of CSL assessment in the former research, it is necessary to further study the differences in CSL among different groups when developing measurement instrument based on the reconceptualized CSL.

The goal, then, is to reconceptualize CSL, put SE at the core of CSL assessment, and put forth a framework that covers the necessary and important parts of evaluating effective SE. Based on this framework, we developed and validated a CSL assessment instrument, and discussed the differences of groups with different demography characteristics in CSL based on the assessment results. Based on this research objective, the specific research questions of this study are determined as follows:

- (1) What evidence is available for the validity of the CSL measurement instrument?
- (2) Is there any significant difference in the CSL of citizens based on their demographic characteristics?

Conceptualizing of CSL

Conceptualizing the CSL and developing a framework for assessment are the fundamental aspects of instrument development. We have conducted an extensive review of the literature on SL and CSL, paying particular attention to how it is defined, the different elements that have been found, and the available tools for assessing it. The international comparative analysis lays the foundation for defining the connotation and extension of CSL and provides a reference for the construction of an assessment system and the development of an assessment instrument.

The new era requires citizens to participate in various issues and public affairs through SE. Therefore, we follow the State Council's definition of CSL (The State Council of PRC, 2006), emphasize SE, and point to SL's vision III. We developed CSL with the idea that CSL is the cornerstone of public health and well-being, the basis of social civilization and progress, and the essential assurance for the creation of a community of shared future and the long-term sustainability of human society as a whole. We combined the definition of CSL in this study, analyzed, compared and grouped the components of SL and CSL in relevant policy documents and extant literature, and identified the elements that are considered crucial to CSL. We believe that the public's understanding of scientific concepts and procedures, the demonstration of problem-solving abilities, and the spirit of science and thoughts demonstrated based on the aforementioned elements are all considered in the assessment of CSL. After experts' review and multiple rounds of revisions, a four-dimensional assessment framework for CSL was developed, which incorporates scientific knowledge (SK), scientific method (SM), problem-solving (PS), and scientific thought and spirit of science (STSS) (Table 1).

Table 1 Framework for the assessment of CSL

Dimension	Elements
Scientific knowledge	mathematics and information, matter and energy, life and health, earth and environment, engineering and technology
Scientific method	observation and experimentation, induction and deduction, model representation
Problem-solving	essential life and production skills, practical and reflective abilities, science engagement (SE)
Scientific thought and spirit of science	Scientific Outlook on Development, nature of science (NOS), exploration and persistence, rationality and questioning, empirical and innovation

From the process of citizens' SE, the foundation and premise parts are SK and SM. SK includes five elements: engineering and technology, life and health, earth and environment, mathematics and information, and matter and energy. SM contains three elements: model representation, induction and deduction, and observation and experimentation. The center part of CSL is PS, which also includes the following three elements: necessary life and production skills, practical and reflective abilities, and SE. STSS is an indispensable part of enabling the general public to think about scientific issues like scientists from a broader perspective of sustainable development, which includes five elements, namely Scientific Outlook on Development, nature of science (NOS), exploration and persistence, rationality and questioning, and empirical and innovation.

Overall, to reach effective SE, in addition to SK and SM, citizens should also possess critical aspects of the spirit of science and a wider range of scientific thought to better solve problems and participate in decision-making.

Scientific knowledge

SK appears in almost all SL visions (Carlson, 2008; Liu, 2013; Mc Eneaney, 2003). The SK discussed here is often of content knowledge, and is the foundation of SE. The first dimension in the framework, SK, is the fundamental dimension of CSL (Norris & Phillips, 2009). SK reflects the understanding of the main facts, concepts, and explanatory theories that form the basis of scientific knowledge, including the basic internal knowledge of mathematics and information, matter and energy, life and health, earth and environment, engineering and technology that the public should possess to meet the needs in life and production.

SK, which is the static component of science, is the end product of scientific investigation. SK consists of facts, concepts, laws, hypotheses, theories, and models. It can also be separated into earth and space science, material science, and life science, depending on the research subject (E. Liu, 2009). The understanding of SK is the most basic and important part of CSL. Students cannot further develop science-related skills, practice science, utilize science to solve issues, or take part in decision-making without the necessary knowledge base and a proper and clear understanding of science.

Scientific method

While the first dimension emphasizes the knowledge of nature and technology, this second dimension refers to the way to SK. Science is a body of knowledge based on rationality and criticism, and its essential feature is the method of constructing knowledge (Mönch & Markic, 2022). SM is a series of logical investigation processes

adopted by subjects engaged in scientific activities to raise and answer specific questions about nature (Mönch & Markic, 2022). Science can be recognized (Galloway, 1992) by the questions it asks and the way it answers them in an attempt to produce a knowledge system (Mönch & Markic, 2022) that is not influenced by individual beliefs, perceptions, values, attitudes, and emotions. To study things in nature and find patterns in nature or life phenomena, scientists need to explain or discover them through observation, experimentation, and analytical reasoning. The process by which scientists acquire knowledge, that is, the scientific process, is a set of methods to study and solve natural problems and is a dimension of scientific dynamics. For the layman, understanding how science works may be more useful than understanding its results (Galloway, 1992).

The basic methods and skills required to engage in scientific practice, such as the capacity to engage in inquiry, demonstration, scientific research, operation, etc., are included in the framework for SM. Understanding the methodologies and processes used in scientific research is a major component of SM assessment. The public should be able to apply scientific methods such as observation, experimentation, induction, deduction, modeling, and so on to assess and solve scientific problems. They should also be able to understand the relationships between these approaches.

Problem-solving

One important area of developing CSL is the ability to integrate science into daily life, particularly the application of science to address problems in daily life (DeBoer, 2000). The ability to develop social science decision-making and scientific problem-solving skills is more important than an understanding of basic content knowledge (Holbrook & Rannikmae, 2009). Problem-solving points to the final application elements of the improvement of CSL. Citizens can use the contents above to solve problems encountered in life and take part in social decision-making when they possess the necessary SK, the corresponding SM, and the ability to connect science with various disciplines and society. For example, they can use the above contents to select the best self-rescue method in the event of emergencies like earthquakes. They can also make decisions on socio-scientific issues considered to be an essential part of scientifically literate citizens (Bell & Lederman, 2003; Kolstø, 2001; Zeidler et al., 2002). For example, they can engage in debates and dialogues regarding some societal issues, such as global warming, and provide their persuasive views based on science.

Therefore, in this framework, the assessment of PS focuses primarily on the ability to analyze and resolve real-world problems as well as critical skills in life and

production. In daily life and production, citizens are expected to master necessary skills related to health, first aid, emergency avoidance, travel, and safety production, and be able to use common household electrical appliances safely. When analyzing and solving problems, scientifically literate individuals can utilize relevant SK and SM, constantly reflect and adjust their thinking, collect and analyze evidence, make judgments based on the evidence, and propose scientific and reasonable solutions. When required, a scientifically literate citizen should be able to work with others to address difficulties, communicate clearly, and make an effort to agree.

Scientific thought and spirit of science

STSS is a crucial part of citizen SE, although it is lacking in existing assessments. The dimensions of STSS mainly include an understanding of NOS, Scientific Outlook on Development, and spirit of science.

Science education researchers have long argued that a key component of CSL is NOS. NOS is the inherent values and assumptions of science (Lederman, 1992). Understanding NOS entails being aware of what it is, what qualities it possesses, and how to tell science from pseudoscience plays an important role in the development of CSL. When discussing CSL, it is important to address the creativity and potential consensus on the nature of science derived from scientific history and philosophy (Chakravartty, 2022), which can help citizens develop more favorable attitudes toward science. NOS here refers to:

- (1) Recognizing that science is a dynamic field that is continually evolving; imagination and creativity can have an impact on how science develops.
- (2) Recognizing how science, technology, society, and the environment are interconnected and influence one another.

Scientific Outlook on Development refers to using the concepts of ecological civilization and sustainable development to guide personal decision-making and emphasizes the thought of sustainability in a complex global environment (The State Council of PRC, 2008). It is an important prerequisite for citizens' SE.

Spirit of science is a group of clustering features from human values, personality traits, cognition preference, and habitual behavior that scientists demonstrate in their research efforts (Zhou et al., 2018). Its core attribute is the pursuit of truth. Specifically, possessing spirit of science means that an individual possesses the following excellent scientific research qualities (Educational Policy Commission, 1970; Zhou et al., 2018):

- (1) Exploration and persistence- having a curiosity and thirst for knowledge about scientific phenomena or things, and being able to correctly face difficulties and persevere in the exploration process.
- (2) Rationality and questioning- being able to analyze and solve problems based on logic, not superstitious about authority, respecting facts, daring to question one's own and others' viewpoints, and being willing to accept questioning from others.
- (3) Empirical and innovation- in the process of analyzing and solving problems, one can obtain evidence based on observations or experiments to test viewpoints and have the awareness of exploring new ideas and methods, striving for excellence.

The emphasis of STSS encourages an inquiry-based mindset, nurtures critical thinking, acknowledges and appreciates the positive impact of scientific research on social advancement, and emphasizes the fact that scientific findings can have both positive and negative effects.

Based on the above conceptualization of CSL, the next step is to develop, and validate the CSL assessment instrument, and explore the differences between people's CSL of various demographic parameters.

Methods

Item pool development

Context has played an important role in the assessment of SL in recent years (Naganuma, 2017; OECD, 2017, 2019). Developing tools for assessing abilities and literacy based on context has become a new trend. The vision III of CSL emphasizes the characteristics of the contexts with a scientific component (X. Liu, 2009, 2013). According to the cognitive perspective of situated cognition theory, item development aimed at evaluating people's cognitive level should be based on the specific context created. CSL (especially SM, PS, and STSS) has a robust internal recessiveness that needs to be activated in specific contexts to induce the explicit. In other words, CSL is appropriate for assessment based on contexts. To allow participants to concentrate and reflect on SK, SM, and STSS via PS in context, we suggest a context-based item development approach (Fig. 1). This approach places specific items in an authentic context. Researchers will develop corresponding research ideas based on scientific phenomena and problems, conduct research to provide results and explanations, and actively promote their applications. SK, SM, PS, and STSS are all required when proposing and carrying out research ideas, constructing conclusions and explanations, and promoting applications. There is a significant link between "doing science" and "using science," even if not everyone is a scientist or can pursue a scientific career. People choose to learn to "use science"

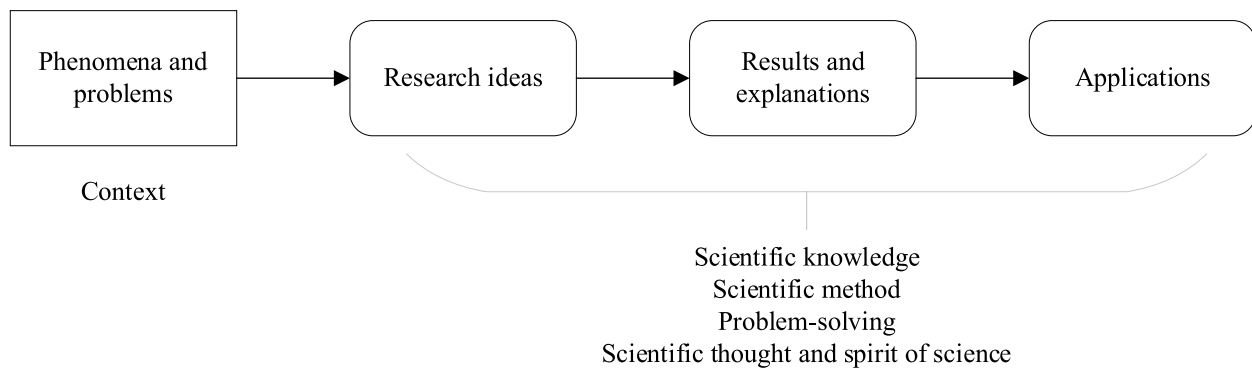


Fig. 1 Context-based item development approach

and think like scientists without having to “do science” by directly participating in scientists’ scientific research processes. This is an essential way to develop CSL. From this perspective, the development approach for items is feasible.

Context is an important carrier for achieving test objectives and a vital factor for arousing participants’ interest and effectively activating their cognition. The contexts of items are based on citizens’ production and lives, pointing to authentic scientific phenomena and problems familiar to citizens, being unbiased in terms of gender, ethnicity, race, etc., and having universality and fairness. The breadth of SK is an integral part of CSL. However, in specific contexts, there are often fewer categories of knowledge that can be examined. Based on this, we have developed three types of items:

- (1) Single items without contexts using a true or false question format, used to explore the breadth of SK.
- (2) Single items that are presented with contexts and no correlation, are used to examine SM and PS. They are all multiple-choice questions.
- (3) Item sets that are based on a specific context and arranged in a particular logical order (such as the course of scientific research). Item sets include multiple-choice questions and five-point Likert scale questions and be used to examine SM, PS, and STSS.

The first type of items without context all use a true or false question format, with scores assigned as 0 or 1. Given that the assessment of CSL with SK as the core has a long history and rich experience, we reviewed and learned from previous items on the SK assessment. To facilitate comparison with previous assessment data, some of the existing items are adopted in the instrument. For example, “Earth goes around the Sun once every year,”

“Light travels faster than sound,” and “The continents on which we live have been moving their locations for millions of years and will continue to move in the future,” which continue to appear in the CSL assessment in the United States and China (Miller, 2000). These items are of good quality, and the SK in them is still at the basic level of the SK system, which is central to measuring the public’s understanding of science (Miller, 2000; Miller, 2022; Wu et al., 2018). Therefore, these items are referenced and retained. Simultaneously, we also keep track of current scientific research developments and have chosen recent scientific findings of general interest, such as clean energy, 5th Generation mobile communication technology (5G), and chips, to be covered by the study. Examples include “Solar energy is a type of sustainable and clean energy” and “The 5th Generation Mobile Communication Technology (5G) transports signals through electrical waves”.

For items with context, we drew inspiration from the design of large-scale international testing projects (OECD, 2019). Contextual single items serve primarily as anchors for specific links in the scientific research process and construct tasks based on crucial context information and the dimensions to be assessed. These types of items are all multiple-choice items. For example, Item 24 (Fig. 2) provides the context of sugar block dissolution that citizens often encounter in their daily lives and aims to examine the “observation and experimentation” elements of the SM dimension. Based on the intention of the item, the task and option settings of the item have been clarified. The operations in the options are commonly seen or used by citizens during the observation and problem-solving process of the “sugar block dissolution” phenomenon. Only the first option is correct. Therefore, when assigning Rasch scores, assign the first option a value of 1 and the other options a value of 0. Although the incorrect options are interference ones, it’s

Which of the following methods will dissolve the sugar cubes in a glass of water the fastest?

- Stir the water continuously.
- Add ice cubes to the water.
- Pour the same temperature water into the glass.
- Do nothing.
- Don't know.

Fig. 2 Sample of a single item (Item 24) with context

also frequently occurring operations. It should be noted that we have always chosen contexts that are familiar, appropriate, and fair to citizens for item design.

Contextual item sets often involve a hot topic that is heavily debated in society, the subject of in-depth scientific study, and one that is well-known to the general public. While the latter provides a foundation for writing items that meet the level of citizens' cognition, the former offers rich materials and crystal-clear logical chains for organizing several items in item sets. According to the characteristics of the proposed evaluation dimension and context, the item set includes multiple-choice items and 5-point Likert scale items. A portion

of a 13-item item set is shown in Fig. 3. Global climate change and carbon dioxide emissions serve as the context for this item set. Global, regional, and personal concerns all revolve around the hot issue of global warming. It is anticipated that it will continue to be a subject of ongoing concern for many groups for a sizable period in the future, including researchers, regular people, and government decision-makers. Scientists have studied the topic of global climate change extensively over the last few decades, developing a variety of viewpoints and adaptive understandings. It may be said that the item set created based on this circumstance has some justice and stability, given the universality and consistency of the research

The increase in the amount of carbon dioxide in the atmosphere may have impact on the global climate. Countries around the world actively advocate low-carbon actions to reduce carbon dioxide emissions.

Q1.The following diagram shows changes in atmospheric carbon dioxide concentration and global average temperature since 1880. What kind of relationship do you think exists between the two figures?

Atmospheric carbon dioxide concentration (ppm) (Left Y-axis): 260, 280, 300, 320, 340, 360, 380, 400, 420
 Global average temperature (Celsius) (Right Y-axis): 56.5, 56.7, 56.9, 57.1, 57.3, 57.5, 57.7, 57.9, 58.1
 X-axis: Year (1880, 1900, 1920, 1940, 1960, 1980, 2000, 2020)

- Causal relationships
- Correlation relationships
- No relationships
- Don't know

Q2. Some people think that "the gas produced by the volcanic eruption enters the atmosphere, which can block the sun's rays and alleviate global warming", but others hold the opposite opinion that "the volcanic eruption releases a lot of heat and will accelerate global warming". For these two different viewpoints, please choose how much do you agree with the following statements.

Question	Strongly disagree	Disagree	Uncertain	Agree	Strongly agree
Q2-1.Neither view provides sufficient evidence, so both are not credible.	○	○	○	○	○
Q2-2.I would like to discuss with others, if possible, to reach a consensus on this issue.	○	○	○	○	○

Context

The assessment element of ItemQ1 is the "induction and deduction" of the "scientific method" dimension.

The assessment element of ItemQ2-1 is "rationality and questioning" in the dimension of "scientific thought and spirit of science".

The assessment element of ItemQ2-2 is "science engagement" in the dimension of "problem-solving".

Fig. 3 Part of the sample item set

and conversation on this topic. Furthermore, this item set can serve as effective items for assessing CSL for a long time. A graph of the changes in the average global temperature and atmospheric carbon dioxide concentration since 1880 is shown in Item Q1 in Fig. 3. The task is to examine the relationship between the two variables. Only the second option is correct. Therefore, when assigning Rasch scores, assign the second option a value of 1 and the other options a value of 0. To respond to this question, participants must summarize and evaluate the relationship between the two curves based on the graph's evidence of their shifting trends. Moreover, SM's "induction and deduction" elements are examined in this item. Two positions on global warming were presented in Item Q2, and participants could indicate whether they agreed or disagreed with the two positions. Item Q2-1 requires participants to express their opinions on the credibility of the two different viewpoints and evaluates the "rationality and questioning" elements of the STSS dimension. Item Q2-2 requires participants to describe their desire to participate in scientific dialogue between two perspectives and evaluates the "science engagement" elements of the PS dimension. For Q2-1 and Q2-2, during Rasch scoring, "Agree" and "Strongly agree" were reassigned as 1, while others were assigned as 0.

We closely adhered to the evaluation indicator framework during the item development process to maintain consistency between the item and each dimension. All of the items underwent three rounds of discussion and improvement after being drafted, resulting in an item pool that included all dimensions and elements.

Pilot study

Fifty people aged between 18 and 70 were randomly chosen for the pilot test, to preliminarily explore the distribution of various options in the test questions and the required duration of the test (approximately 30 minutes). Then, 10 participants were selected for additional cognitive interviews to further explore their understanding of the items.

Data source

Data from a random sample were gathered via online questionnaires. Online surveys are a feasible method, as evidenced by the 74.4% Internet penetration rate (CNNIC, 2022) and the current practice of network collection-based nationwide CSL surveys in China. The data was collected from April to May 2022 using a popular survey app in China that generated and randomly distributed 603 anonymous survey questionnaires, in which 578 valid responses were kept after data cleaning. The demographic details of the valid samples are shown in Table 2.

362 females and 216 males made up the sample. Their ages ranged from 18 to 70.

Data analysis

Combining item difficulty and person ability on one unit (logit), the Rasch model offers a reliable method for assessing latent human qualities (Boone et al., 2014). Rasch's modeling method predicts that for each item, a person with high ability should be more likely to provide the correct answer than a person with poor ability and that for any individual, their performance on easy items should always be better than that on difficult ones. In human research and education, Rasch analysis is frequently employed (Boone et al., 2014; Lamar, 2012). Since 2007, China has been in line with international standards and started using the IRT method for calculating the CSL score, which is an effective way to gauge CSL (He, 2019; Miller, 1998). Rasch analysis in IRT was thus a logical choice for our study.

ConQuest combines a variety of item response models into a single computer program that can fit item response and potential regression models. The multidimensional item response model can be used to analyze items intended to generate measures on up to 10 potential dimensions (Wu et al., 2007). Two multidimensional

Table 2 Demographic information of the valid sample

Characteristics		N	Percent
Gender	Female	362	62.63%
	Male	216	37.37%
Age	<20 years	27	4.67%
	20–29 years	135	23.36%
	30–39 years	155	26.82%
	40–49 years	206	35.64%
	50–59 years	38	6.57%
	60–70 years	17	2.94%
Location Type	Urban	465	80.45%
	Rural	113	19.55%
Educational Background	Primary school and below	8	1.38%
	Junior middle school	20	3.46%
	Senior high school (secondary junior college, technical school)	84	14.53%
	College degree	122	21.11%
	Bachelor's degree	316	54.67%
	Master's degree or above	28	4.84%
Occupation	Have a job	326	56.40%
	No job	117	20.24%
	Students waiting for further study	124	21.45%
	Retired personnel	7	1.21%
	Inability to work	4	0.69%

Rasch measurement designs are named between-item design and within-item design. Unlike the latter, where each item can be used as an indicator of numerous potential dimensions, each item in the former can only be used as an indicator of one potential dimension (Wu et al., 2007). The two designs mentioned above can be subjected to a fitting validation examination by ConQuest. Each item on the multidimensional between-item design instrument for this study only aims to evaluate one potential dimension. Based on our conceptualization of CSL, we used the Rasch partial credit model to verify the instrument. In particular, the fit statistics and reliability of the instrument were closely examined. We learned about the quality of each item through Winstep to better support instrument validation because Winstep can provide more information.

We optimized the items based on experts' reviews and cognitive interviews and ultimately selected 54 items for the test. We first calculated the ability values of the participants in four dimensions using ConQuest to compare the differences in CSL among people of different genders, ages, location types (urban or rural), educational backgrounds, and occupations. We then used SPSS22.0 for analysis, confirmed that the variables did not conform to Normal distribution, and proceeded to use a nonparametric test for further analysis.

Validation of the instrument

According to the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), five different types of evidence may support a validity argument based on test content, response process, internal structure, relations to other variables, and consequences of testing. We argue that data generated through the instrument for measuring CSL can be used to valid inferences about their CSL. Under this premise, we focused on the validity evidence based on test content, response process, and internal structure of the instrument for measuring CSL.

Evidence based on test content

To improve content validity, we conducted two rounds of expert reviews on the framework and one expert discussion on the tool.

The first expert review is to collect experts' opinions on the description of important components of the connotation and extension of CSL. 28 experts from China, the United States, Malaysia, Pakistan, Turkey, and other countries in the fields of science research, science education, science popularization, and science and technology were invited to participate in a questionnaire survey

on the importance of determining the meaning of CSL and its four dimensions of elements and behaviors. The questionnaire consists of 51 items using the 5-point Likert scale, with a score of 5 representing the highest degree of importance. The Cronbach's alpha coefficient is 0.824 and indicates the questionnaire has high reliability. Based on the judgment basis of experts' responses to the questionnaire (including practical experience, theoretical knowledge, domestic and foreign peer research, and intuitive perception), as well as their familiarity with CSL, the authoritative coefficient (Cr) of experts is determined, mainly distributed in the range of 0.802 to 0.901 (all greater than 0.8) (Wu et al., 2022), indicating that experts have a relatively high level of authority and have a high level of credibility in their responses to the questionnaire. All items have a high average importance score, and experts believe that the vast majority of items are very important. The coefficient variation (Cv) of 45 items out of 51 items is less than 0.2 (Wu et al., 2022), indicating that experts have a relatively consistent level of agreement with the specific descriptions of the four dimensions. Based on this, a consensus has been reached on the four important components of CSL and their descriptions.

Thirty experts in the fields of science education, measurement and evaluation, scientific literacy research, scientists, and engineering technicians were invited to evaluate the structure and completeness of the assessment framework. The questionnaire consists of 13 items, using a 5-point Likert scale, with 5 representing the highest level of recognition. Among them, one item is used to investigate the overall structural rationality of the CSL evaluation framework, with each dimension containing three items, respectively exploring the completeness, accuracy, and testability of the description of the dimensions. The survey results show that experts have a high level of recognition for the overall and various dimensions of the CSL framework.

To ensure the content validity of the instrument, we invited 25 experts from 12 countries including China, Israel, Singapore, Australia, etc. to review the items in the form of online meetings. They proposed improvement suggestions for the effectiveness of the items, including the scientificity, fairness, and consistency between the items with the corresponding indicators. Based on the experts' opinions, we revised some items, including further refinements have been made to the wording of some items, converting relevant content and expressions into a more understandable way for the general public, and revising or deleting items that may be influenced by public ideology during the investigation.

Evidence based on response processes

The response process refers to the cognitive process of survey participants, that is whether the interpretation of the items by participants aligns with the intended interpretation of items of items by test developers (American Educational Research Association et al., 2014). The cognitive interview is a widely used qualitative survey development method for questionnaire design. Its purpose is to gain insights into whether the respondent’s understanding of the survey items aligns with the intentions of instrument developers (Willis, 2005) and can provide evidence of instrument development validity based on reaction processes. To ensure that the wording of the items is understandable to the target person and conforms to the intention of instrument development, we selected 3 people aged 18–30, 4 people aged 31–50, and 4 people aged 51–70 in the pilot study, and conducted three focus group cognitive interviews to determine any problems related to the items in terms of language understanding, appropriateness, etc., and revised any complex and incomprehensible words or phrases in the items marked by the interviewees. According to the interviews, all of items the were understandable and their answers reflected their real opinions.

Evidence based on internal structure

The internal structure is the extent to which the instrument conforms to the constructs and covers the instrument’s dimensionality and item functioning (American Educational Research Association et al., 2014). Item functioning refers to internal consistency, item-total and interitem correlations, etc. Here, we mainly present validity evidence based on internal structure by providing fit statistics, Wright Map, reliability, and correlation between SCL dimensions (Rios & Wells, 2014).

Fit statistics

In Rasch analysis, item functioning is usually checked through fit statistics. The standardized residual, which is the difference between the observed response and the expected response under the model, is the main input in Rasch’s goodness of fit analysis (Wright & Masters, 1982). Both unweighted and weighted MNSQ are important indicators of fit statistics, with an acceptable range of 0.6 to 1.4 (Bond & Fox, 2015). The former is more sensitive to outliers, whereas the latter is more sensitive to aberrant response patterns near item difficulty or person ability. Through ConQuest, the item fit was examined to check whether the items met the criteria. According to Table 3, unweighted MNSQ greatly surpasses 1.4, and the vast majority of T-values far exceed 2, which indicates poor fitting of these four

Table 3 Fit statistics of items with poor fit

Item	Dimension	Unweighted		Weighted	
		MNSQ	T	MNSQ	T
DEL1	SM	1.82	8.0	1.31	4.5
DEL2	PS	1.85	8.3	1.29	3.6
DEL3	STSS	1.84	8.2	1.18	1.8
DEL4	STSS	1.56	5.8	1.42	7.1

items (DEL1-DEL4). The change in Cronbach’s alphas if an item was deleted, the description and dimensions of the items, and the information gathered from interviews were all taken into account when deciding how to handle (whether to retain or delete) these items. In the end, these four items were removed to improve model fit and reliability.

Winstep was used to examine each item, and Fig. 4 displays the Item Characteristic Curve (ICC) of Item 45. The horizontal axis represents the difference between the person’s ability and item difficulty. Theoretically, the larger the difference, the more likely it is to receive a higher score. The probability that a score will fall within that range is shown on the vertical axis. When two curves overlap, it indicates that there is an equal chance that reactive participants will receive two scores. Each rating curve should have an identifiable ‘peak,’ showing the most likely score for participants within that ability range. The peaks of the curves should ideally appear in numerical sequence, and each curve should suitably span a specific ability area. From Fig. 4, the “Level 1” curve in Fig. A is in reverse order to the intersection of “0” and “2” and lacks notable peaks. Both the Infit and Outfit MNSQ at level 1 exceed the acceptable range of 1.4 (Bond & Fox, 2015), indicating a problem with horizontal division. We then combined the original level 1 with level 0 and adjusted level 2 to level 1 based on responses and the item descriptions. After horizontal division adjustment, Fig. B demonstrates that the ICC and fit statistics perform well, suggesting that the revision is appropriate.

The final instrument consists of 50 items in total after the five items have undergone the aforementioned adjustments, of which SK, SM, PS, and STSS each include 23, 7, 8, and 12 items. The multidimensional Rasch model in ConQuest is then used to analyze four-dimensional data. According to the findings (Table 4), both the weighted and unweighted MNSQ of each item falls within the permitted range of 0.6 to 1.4 (Bond & Fox, 2015). The great majority of items have T-values that fall in the permissible range of –2 to 2 (Bond & Fox, 2015). We first look at and rely more on MNSQ because the T-value heavily depends on sample size,

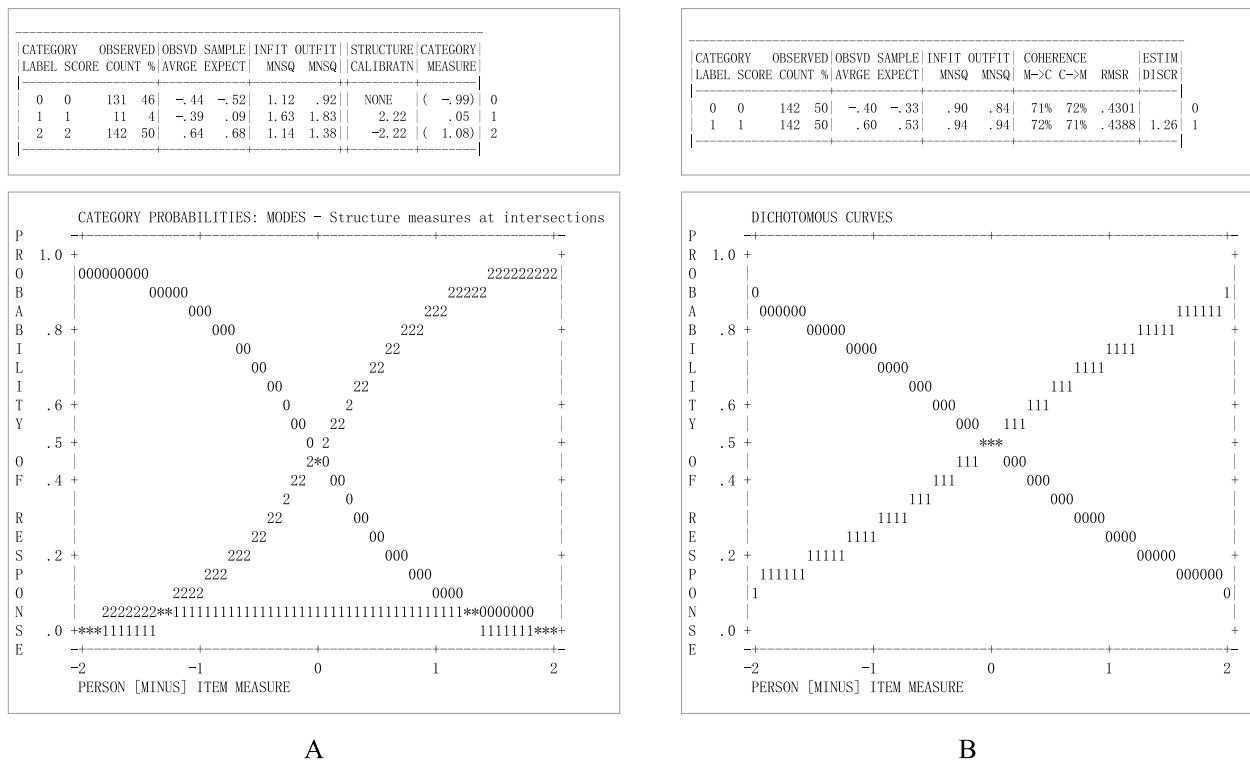


Fig. 4 The Item Characteristic Curve (ICC) of Item 45 before and after the revision of the rating level. (A is before the item revision, while B is after the revision)

and this study has a large sample size of 578 people. As long as the MNSQ value is within an acceptable fit range, we ignore the T-value (Boone et al., 2014). Overall, all of the instrument’s items generally fit the four-dimensional Rasch model well, indicating good validity to measure CSL.

Wright map

To intuitively convey the ability estimation of the person and the difficulty distribution of the item, as well as their corresponding degrees, the Wright map, also known as the person-item map, can place the difficulty of the item and the ability of the person on a Rasch scale (Boone et al., 2014). The Wright map of this study is shown in Fig. 5, where each “X” denotes 5.6 cases, and the four vertical lines represent the four dimensions of CSL. Person’s ability increases gradually from bottom to top. Figure 5 displays a good distribution of people throughout 5 to 6 logits in each dimension. On the right side of the vertical line is the distribution of item difficulty, which ranges from -2 to 2 logits, basically covering people with different ability. The findings demonstrate that, despite a lack of items measuring high and low levels of person’s ability, the instrument generally meets the requirements.

The application of this instrument could help depict the distribution of a group’s ability in CSL.

Reliability

Separation reliability is an indicator of reliability used to evaluate the instrument’s characterization ability on different dimensions. According to the findings through ConQuest, the instrument has a separation reliability of 0.985, which means that the characterization ability in multiple dimensions can successfully separate people’s ability. By measuring scores on a particular dimension, EAP/PV reliability demonstrates the capacity to discern between various cognitive ability. Table 5 displays the total number of items and EAP/PV reliability in the instrument’s four dimensions. According to Table 5, the EAP/PV reliability distributions for the four dimensions range from 0.779 to 0.825, all higher than or close to 0.8, which indicates good reliability. Overall, the tool satisfies the criteria for reliability in science research and has good reliability as a whole and in different dimensions. Additionally, it shows that the multidimensional Rasch analysis technique has been properly implemented.

Table 4 Estimate, Error, and Fit statistics of items in the revised instrument of CSL. (An asterisk next to a parameter estimate indicates it is constrained)

Item		Estimate	Error	Unweighted		Weighted	
				MNSQ	T	MNSQ	T
1	SK_01	0.148	0.095	1.16	1.9	1.08	1.3
2	SK_02	-0.372	0.099	0.85	-1.9	0.95	-0.6
3	SK_03	0.426	0.076	1.06	1.1	1.05	1.5
4	SK_04	0.151	0.095	1.10	1.2	1.06	1.1
5	SK_05	1.281	0.092	1.13	1.6	1.07	1.5
6	SK_06	-0.463	0.082	1.05	0.9	0.97	-0.5
7	SK_07	-0.551	0.100	0.86	-1.7	0.99	0.0
8	SK_08	1.035	0.075	1.10	1.7	1.07	2.0
9	SK_09	-0.388	0.081	0.80	-3.6	0.88	-2.5
10	SK_10	0.837	0.075	1.20	3.2	1.15	4.5
11	SK_11	0.375	0.076	1.07	1.2	1.06	1.8
12	SK_12	0.418	0.076	0.97	-0.5	0.98	-0.4
13	SK_13	0.110	0.095	0.94	-0.7	0.99	-0.2
14	SK_14	-0.737	0.102	0.62	-5.3	0.83	-1.8
15	SK_15	0.067	0.095	0.92	-0.9	0.93	-1.1
16	SK_16	-1.028	0.103	0.70	-4.0	0.85	-1.6
17	SK_17	-0.480	0.098	0.75	-3.2	0.86	-2.1
18	SK_18	-0.377	0.097	0.86	-1.7	0.89	-1.7
19	SK_19	-1.051	0.103	0.90	-1.3	0.93	-0.7
20	SK_20	-0.294	0.096	1.06	0.8	1.07	1.1
21	SK_21	0.360	0.093	1.12	1.4	1.04	0.8
22	SK_22	0.362	0.093	1.01	0.1	1.01	0.3
23	SK_23	0.172*	0.429	0.86	-1.8	0.90	-2.1
24	SM_01	0.334	0.080	0.96	-0.6	0.99	-0.1
25	SM_02	-0.275	0.081	0.77	-4.2	0.86	-3.4
26	SM_03	0.909	0.098	0.96	-0.4	1.00	0.1
27	SM_04	-1.864	0.113	0.79	-2.7	1.00	0.0
28	SM_05	1.184	0.101	1.02	0.2	0.98	-0.3
29	SM_06	0.350	0.098	1.39	4.2	1.28	4.5
30	SM_07	-0.638*	0.234	1.10	1.2	1.03	0.6
31	PS_01	-1.463	0.101	0.81	-2.5	0.92	-1.0
32	PS_02	0.664	0.079	0.80	-3.6	0.88	-2.8
33	PS_03	0.553	0.094	0.93	-0.9	0.98	-0.4
34	PS_04	-0.032	0.094	0.98	-0.3	0.99	-0.1
35	PS_05	-1.136	0.096	0.95	-0.6	0.99	-0.1
36	PS_06	0.958	0.099	1.29	3.2	1.05	0.7
37	PS_07	0.418	0.095	1.35	3.8	1.28	4.5
38	PS_08	0.038*	0.249	1.00	0.0	1.00	0.1
39	SS_01	0.696	0.099	1.24	2.7	1.10	1.5
40	SS_02	-1.258	0.104	1.33	3.7	1.17	2.1
41	SS_03	0.754	0.099	1.16	1.9	1.11	1.7
42	SS_04	-0.861	0.101	0.95	-0.6	0.96	-0.6
43	SS_05	1.076	0.101	1.11	1.3	1.06	0.9
44	SS_06	-1.381	0.104	0.79	-2.7	0.96	-0.5
REV45	SS_07	-0.258	0.099	1.08	1.0	1.13	2.0
46	SS_08	-0.238	0.099	0.76	-3.1	0.81	-3.3
47	SS_09	0.898	0.103	1.27	3.0	1.21	2.9

Table 4 (continued)

Item		Estimate	Error	Unweighted		Weighted	
				MNSQ	T	MNSQ	T
48	SS_10	0.766	0.102	0.79	-2.6	0.84	-2.5
49	SS_11	-0.057	0.100	0.72	-3.7	0.80	-3.5
50	SS_12	-0.137*	0.335	0.77	-3.0	0.85	-2.5

Correlation between dimensions of CSL

By providing the opportunity to fit a multidimensional item response model, ConQuest allows for evaluating correlations between potential variables, reflecting the relationships between various dimensions in Rasch measurements through correlation. Table 6 shows the correlation (values below the diagonal) of the CSL sub-dimensions generated by ConQuest. The correlation range between dimensions ranges from 0.676 to 0.877. Specifically, the correlation coefficients between SK and SM, as well as between SM and PS, are 0.836 and 0.877, both greater than 0.8, indicating a high degree of correlation; The correlation coefficients between SK and PS, SK and STSS, SM and STSS, and PS and STSS are 0.782, 0.696, 0.676, and 0.799, respectively, all greater than 0.5, indicating a moderate correlation (Bond & Fox, 2015). The correlation between any two dimensions is statistically significant ($df=46, P<0.05$) and is still within the acceptable range of 0.3 to 0.9 (Lin et al., 2019), indicating that there is sufficient discriminant validity between dimensions.

Impact of demographics factors on CSL

The nonparametric test results seen in Tables 7 and 8 demonstrate that while there was no statistically significant difference in performance between males and females, some demographic characteristics, such as age, location type, educational background, and occupation, have a remarkable impact on CSL (the significance level is 0.05).

Comparing distributions across groups provides more details. In terms of age variables, residents aged 60 and older perform significantly worse on the SK, SM, PS, and STSS dimensions than citizens in all other age categories. Urbanites perform much better than rural citizens across all four CSL dimensions. Citizens’ CSL gradually rises as their educational background does. However, there is no significant difference in PS and STSS between citizens with a college degree and those with a master’s degree, nor is there any significant difference in STSS between citizens who only attended high school and citizens with a master’s degree. Citizens with a bachelor’s degree or above are significantly higher than those with a low degree in SK, SM, PS, and STSS dimensions.

Regarding occupation, the CSL of unemployed citizens is significantly lower than that of employed citizens and students waiting for further study.

Discussions

Discussion of the instrument

The four-dimensional instrument was developed to assess CSL. It can be used online or via a paper and pencil format. 50 items make up the instrument, and it takes about 30 minutes for the participants to complete the survey. According to the *Standards for Educational and Psychological Testing* (American Educational Research Association et al., 2014), we have provided validity evidence based on test content, response process, and internal structure. The validity evidence collected during the instrument development process of this study includes content evidence based on the expert review, response process validity evidence based on cognitive interviews, and internal structure validity evidence based on four-dimensional model fit. The findings support that this instrument can be used to assess CSL.

In some ways, the instrument’s high quality suggests that the elements we considered for the CSL framework and tools’ evaluation are reasonable. They are crucial components of CSL that every citizen should possess since they aid in problem-solving and engagement in public affairs. However, additional research will be necessary to better understand what SK, SM, and STSS are needed by citizens, and how to address issues and participate in science in more practical situations.

Discussion of different groups’ CSL

In exploring the differences in CSL among citizens with different demographic characteristics, our findings are in agreement with several prior studies that have used other instruments to assess CSL.

According to earlier studies, numerous demographic variables have a significant impact on CSL. For example, citizens with higher educational backgrounds and younger ages have higher CSL (Miller, 1998, 2006; Ren et al., 2022), which is consistent with our research findings. This is in line with our general understanding that citizens who receive higher levels of education can learn more scientific knowledge and methods, have

	Dimension				Terms in the Model (excl Step terms)
	Dim 1	Dim 2	Dim 3	Dim 4	
4					
3	X	X		X	
	XX	X		X	
	XX	XXX		X	
	XXXX	XX		X	
2	XXXXX	XXXX	X	XX	
	XXXXXX	XXX	X	X	
	XXXXXXXX	XXXXX	XX	XXXX	
	XXXXXXXXXX	XXXXX	XXX	XXXX	5
	XXXXXXXXXX	XXXXX	XXXX	XXXX	28
1	XXXXXXXXXX	XXXXXX	XXXX	XXXX	8 26 36 43 47
	XXXXXXXXXX	XXXXXXXX	XXXX	XXXX	10 41 48
	XXXXXX	XXXXX	XXXXX	XXXX	32 33 39
	XXXXXXXXXX	XXXXX	XXXXXXXXXX	XXXX	3 11 12 21 22 24 29 37
	XXXXX	XXXXXX	XXXXXXXXXX	XXX	1 4 23
0	XXXXXX	XXXXX	XXXXXXXXXX	XXXXXXXX	13 15 34 38 49
	XXX	XXXXXXXX	XXXXXXXX	XXXXXXXX	45 46 50
	XXXXX	XXXXX	XXXXXX	XXXXXX	2 9 18 20 25
	XXX	XXXX	XXXXXX	XXXXXXXX	6 7 17 30
	XX	XXXX	XXXXXXXXXX	XXXXXX	14
-1	XX	XXX	XXXX	XXX	16 42
		XX	XXXX	XXXX	19 35
	X	XXXX	XXX	XXX	40 44
		XXX	XXXXX	XXX	31
		X	XX	XX	
-2		XX	X	XXX	27
		X	XXX	X	
			XX	XX	
		X	X	X	
			X	X	
-3				X	
				X	
				X	
-4				X	

Each 'X' represents 5.6 cases

Fig. 5 Wright map for the four-dimensional instrument. (Dimensions 1, 2, 3, and 4 are person ability estimations for SK, SM, PS, and STSS)

more opportunities for scientific participation, and have a deeper understanding of the nature of science. Therefore, they can demonstrate higher levels of CSL.

This is consistent with the widespread belief that people with higher levels of education are better equipped to learn new scientific concepts and techniques,

Table 5 EAP/PV reliability for the four dimensions of CSL

Dimension	Number of Items	EAP/PV reliability
SK	23	0.821
SM	7	0.825
PS	8	0.813
STSS	12	0.779

Table 6 Correlation between dimensions of CSL

Dimension	Dimension			
	Dim1(SK)	Dim2(SM)	Dim3(PS)	Dim4(STSS)
Dim1(SK)	–	–	–	–
Dim2(SM)	0.836*	–	–	–
Dim3(PS)	0.782*	0.877*	–	–
Dim4(STSS)	0.696*	0.676*	0.799*	–

* $p < .05$

participate in more scientific activities, and comprehend the fundamental principles of science. As a result, they can exhibit increased levels of CSL. In terms of senior adults, particularly Chinese citizens over 60, the general level of education is quite poor. Ren indicates that the education reform in China in 1977 had a long-term and significant impact on people born after 1960 (Ren et al., 2022). This part explains why Chinese citizens under the age of 60 have a significantly higher CSL in the survey.

The performance of individuals from urban areas is considerably better than that of individuals from rural areas in the four dimensions of CSL when considering the characteristics of the location type. This is in line with common opinion. Urban and rural communities have unequal access to formal and informal education resources, including funding for education, qualified teachers, science and technology museums, planetariums, etc. This results in an uneven distribution of CSL between urban and rural areas.

However, our results also contradict some previously published findings on CSL. Previous national surveys from numerous countries found a statistically significant difference between males and females in CSL (Garner-O’Neale et al., 2013; Miller, 2000; Miller, 2016; Zhang et al., 2016). Our study’s findings, however, are in line with those of a survey conducted in Japan, showing that there is no appreciable distinction between men’s and women’s CSL (Naganuma, 2017). Given the lack of statistical differences between boys and girls in PISA scientific literacy tests (OECD, 2007), Naganuma believes this is a relatively new finding (Naganuma, 2017). Compared to the gender differences reflected in CSL assessments that focus on SK, Naganuma’s tool places making decisions at the center of CSL. It seems that there is no significant difference between men and women when CSL emphasizes “science engagement”.

Research results also reveal differences in CSL among citizens with different occupations. This is consistent with previous international comparative research results (Wu et al., 2018). Individuals with occupations exhibit higher levels of CSL than those who have not participated in social work, which seems to illustrate the important role of social engagement in improving CSL.

Nowadays, CSL plays an increasingly important role at the national and individual levels. How to issue practical policy documents and carry out effective CSL promotion actions at the national level is of great significance. Focusing on the differences in CSL among different groups with different demographic characteristics will help us have a clear target. We need to bridge the gap in CSL among different groups while promoting the overall improvement of CSL for all.

Limitations and implications

This study is an essential inquiry into the construction of the CSL assessment framework and instrument development. The findings presented above should be accepted with caution. The purpose of this study is to develop and validate an instrument. The ability values of the samples in the study are evenly distributed and have a large span, which can serve the development of the instrument well.

Table 7 Nonparametric tests of demographic factors (Gender and Location Type)

Factor	Test	SK			SM			PS			STSS		
		U	Z	Sig.	U	Z	Sig.	U	Z	Sig.	U	Z	Sig.
Gender	Mann-Whitney	37,208.500	−0.972	0.331	37,520.500	−0.811	0.417	37,533.000	−0.805	0.421	37,207.500	−0.972	0.331
Location Type		17,805.000	−5.318	0.000*	18,865.000	−4.652	0.000*	18,505.000	−4.878	0.000*	17,503.000	−5.508	0.000*

U. Mann-Whitney U

Z. Standardized Test Statistic

* $p < .05$

Table 8 Nonparametric tests of demographic factors (Age, Educational Background, and Occupation)

Factor	Test	SK			SM			PS			STSS		
		Test Statistic	df	Sig.	Test Statistic	df	Sig.	Test Statistic	df	Sig.	Test Statistic	df	Sig.
Age	Kruskal-Wallis	31.204 ^a	5	0.000*	28.701 ^a	5	0.000*	25.862 ^a	5	0.000*	16.884 ^a	5	0.000*
Educational Background		65.517 ^a	5	0.000*	54.535 ^a	5	0.000*	48.564 ^a	5	0.000*	42.804 ^a	5	0.000*
Occupation		65.304 ^a	4	0.000*	55.321 ^a	4	0.000*	51.891 ^a	4	0.000*	47.715 ^a	4	0.000*

^aThe test statistic is adjusted for ties

df. Degree of freedom

*p<.05

However, as a populous country, the sample in this study may not represent the entire population of Chinese citizens. Therefore, it is necessary to select larger and more representative samples for further instrument improvement and investigation analysis. Although this instrument is already of good quality, there is still a lack of extremely difficult and low-difficulty items. Future revisions and additions of items are both possible to raise the instrument's quality.

From the perspective of science education research, this study has certain theoretical and practical significance. We incorporated STSS into the framework of CSL and obtained effective validation from tools, enriching the connotation and extension of CSL in theory. In practice, although the instrument was developed in Chinese, the formulation of the assessment framework and the substantial collaboration of outside experts throughout the instrument's development might, in some ways, be seen as an international consensus on CSL and its assessment. Future translations of the instrument into additional languages for regional use in many nations will help to advance CSL horizontal comparability. Future studies may look more closely at the performance level of CSL and the application of the instrument to various samples. All or some of the items may be used during the assessment process when utilizing the instrument.

The development of the new CSL framework and the effective validation of the instrument have certain implications for education within and across science disciplines. Currently, global science education widely focuses on problem-solving ability and advocates for the participation of all citizens in scientific problem-solving in countries and regions (Kolstø, 2001). The importance of PS as one of the dimensions of CSL in this study has been validated. This responds to the focus of science education on PS. The STSS newly incorporated into the CSL framework is also particularly important for science education with SL as its objective. Therefore, when teaching science, teachers should pay attention to the development of students' PS and their comprehension of STSS as well as the students' mastery of SK and SM.

Meanwhile, the significant differences in CSL among groups with different educational backgrounds and occupations suggest the important role of formal education and informal in cultivating individual CSL. Although individuals spend relatively little amount of time in formal education (Liu, 2013), informal education has a lower effect on developing CSL than formal education. Therefore, how to make good use of informal education (such as scientific venues, media information, etc.) to promote CSL for all should receive broader attention.

Abbreviations

CSL	Civic scientific literacy
SE	Science engagement
SL	Scientific literacy
TIMSS	Trends in International Mathematics and Science Study
PISA	Programme for International Student Assessment
OECD	Organization for Economic Co-operation and Development
SK	Scientific knowledge
SM	Scientific method
PS	Problem-solving
STSS	Scientific thought and spirit of science
NOS	Nature of science
5G	5th Generation mobile communication technology
CNNIC	China Internet Network Information Center
ICC	Item Characteristic Curve
Cr	Authoritative coefficient
Cv	Coefficient of variation

Acknowledgements

The research was supported by the "Research on the Assessment System of International Civic Scientific Literacy in 2021" (grant 21EBR014). We would like to thank all the participants for their efforts in helping us understand their civic scientific literacy. We wish to express sincere gratitude to those who took the time to help improve our work and provide many helpful suggestions.

Authors' contributions

JW contributed to the study conception and design, the drafting and substantial review of the work; and has approved the submitted version. YML, ZMZ, JCW, TL, SL, JXL, and SMX contributed to the material preparation and data analysis. The first draft of the manuscript was written by YML. JW and TL substantively revised the work. The authors read and approved the submitted manuscript.

Funding

This research was made possible by the financial support of the China Research Institute for Science Popularization (Grant No. 21EBR014).

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 7 July 2023 Accepted: 26 November 2023

Published online: 01 February 2024

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *The standards for educational and psychological testing*. American Educational Research Association.
- Bell, R. L., & Lederman, N. G. (2003). Understandings of the nature of science and decision making on science and technology based issues. *Science Education*, 87(3), 352–377. <https://doi.org/10.1002/sce.10063>
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.
- 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>

- Carlson, J. L. (2008). *Effect of theme-based, guided inquiry instruction on science literacy in ecology* [Master dissertation]. Michigan Technological University.
- Chakravartty, A. (2022). Scientific knowledge vs. knowledge of science: Public understanding and science in society. *Science & Education*, 32, 1795–1812. <https://doi.org/10.1007/s11191-022-00376-6>
- CNNIC. (2022). *The 50th statistical report on China's internet development*. CNNIC. Retrieved May 25, 2023, from <https://www.cnnic.net.cn/NMediaFile/2022/0926/MAIN1664183425619U2MS433V3V.pdf> (in Chinese)
- Council of Canadian Academies. (2014). *Science culture: Where Canada stands*. Council of Canadian Academies. Retrieved June 10, 2023, from <https://rapports-cac.ca/wp-content/uploads/2019/05/FullReport-Science-Culture.pdf>
- DeBoer, G. E. (2000). Scientific literacy: Another look at its historical and contemporary meanings and its relationship to science education reform. *Journal of Research in Science Teaching*, 37(6), 582–601. [https://doi.org/10.1002/1098-2736\(200008\)37:6%3c582::AID-TEA5>3.0.CO;2-L](https://doi.org/10.1002/1098-2736(200008)37:6%3c582::AID-TEA5>3.0.CO;2-L)
- Educational Policy Commission. (1970). Education and the Spirit of science. *ETC: A Review of General Semantics*, 27(2), 139–164.
- Fives, H., Huebner, W., Birnbaum, A. S., & Nicolich, M. (2014). Developing a measure of scientific literacy for middle school students. *Science Education*, 98(4), 549–580. <https://doi.org/10.1002/sce.21115>
- Galloway, J. (1992). Madness to the method. *Nature*, 358, 464. <https://doi.org/10.1038/358464a0>
- Garner-O'Neale, L., Maughan, J., & Ogunkola, B. (2013). Scientific literacy of undergraduate chemistry students in the University of the West Indies, Barbados: Individual and joint contributions of age, sex and level of study. *Academic Journal of Interdisciplinary Studies*, 2(10), 55. <https://www.richtmann.org/journal/index.php/ajis/article/view/1696>
- Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education*, 11(4), 364–377. <https://doi.org/10.1187/cbe.12-03-0026>
- He, W. (2019). From inheritance to innovation: The development of the supervision and evaluation of Chinese National Survey of scientific literacy. *Studies on Science Popularization*, 14(5), 15–22+33+108. <https://doi.org/10.19293/j.cnki.1673-8357.2019.05.002>. (in Chinese)
- Holbrook, J., & Rannikmae, M. (2009). The meaning of scientific literacy. *International Journal of Environmental & Science Education*, 4(3), 275–288.
- Jones, L. R., Wheeler, G., & Centurino, V. A. S. (2015). *TIMSS 2015 science framework*. Retrieved May 5, 2023, from https://timss.bc.edu/timss2015/downloads/T15_FW_Chap2.pdf
- Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science Education*, 85(3), 291–310. <https://doi.org/10.1002/sce.1011>
- Lamar, M. M. (2012). Using and developing measurement instruments in science education: A Rasch modeling approach. *Science Education*, 96(1), 183–185. <https://doi.org/10.1002/sce.20477>
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331–359. <https://doi.org/10.1002/tea.3660290404>
- Lin, J., Liu, X., Chang, C.-Y., Sun, T., & Shi, X. (2019). *Developing a Measurement Instrument for Teacher Key Competencies* [Conference]. Toronto, Canada: 2019 AERA.
- Liu, E. (2009). *Middle school biology pedagogy* (2nd ed.). Higher education press. (in Chinese)
- Liu, X. (2009). Beyond science literacy: Science and the public. *International Journal of Environmental & Science Education*, 4(3), 301–311.
- Liu, X. (2013). Expanding notions of scientific literacy: A reconceptualization of aims of science education in the knowledge society. In N. Mansour & R. Wegerif (Eds.), *Science education for diversity: Theory and practice* (pp. 23–39). Springer. https://doi.org/10.1007/978-94-007-4563-6_2
- McEneaney, E. H. (2003). The worldwide cachet of scientific literacy. *Comparative Education Review*, 47(2), 217–237. <https://doi.org/10.1086/376539>
- Miller, J. D. (2000). The development of civic scientific literacy in the United States. In K. C. Cohen (Ed.), *Innovations in science education and technology* (pp. 21–47). Springer. https://doi.org/10.1007/978-94-011-3992-2_3
- Miller, J. D. (1983). Scientific literacy: A conceptual and empirical review. *Daedalus*, 112(2), 29–48.
- Miller, J. D. (1998). The measurement of civic scientific literacy. *Public Understanding of Science*, 7(3), 203–223. <https://doi.org/10.1088/0963-6625/7/3/001>
- Miller, J. D. (2006). *Civic scientific literacy in Europe and the United States* [conference]. Montreal, Canada: World Association for Public Opinion Research.
- Miller, J. D. (2016). *Civic scientific literacy in the United States in 2016: A report prepared for the National Aeronautics and Space Administration by the University of Michigan*. Retrieved March 14, 2023, from https://smd-prod.s3.amazonaws.com/science-red/s3fs-public/atoms/files/NASA%20CSL%20in%202016%20Report_0_0.pdf
- Miller, J. D. (2022). Public understanding of science and technology in the internet era. *Public Understanding of Science*, 31(3), 266–272. <https://doi.org/10.1177/09636625211073485>
- Mönch, C., & Markic, S. (2022). Science teachers' pedagogical scientific language knowledge—A systematic review. *Education Sciences*, 12(7), 497. <https://doi.org/10.3390/educsci12070497>
- Naganuma, S. (2017). An assessment of civic scientific literacy in Japan: Development of a more authentic assessment task and scoring rubric. *International Journal of Science Education, Part B: Communication and Public Engagement*, 7(4), 301–322. <https://doi.org/10.1080/21548455.2017.1323131>
- National Academies of Sciences, Engineering, and Medicine. (2016). *Science literacy: Concepts, contexts, and consequences*. Washington, DC: The National Academies Press.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, 87(2), 224–240. <https://doi.org/10.1002/sce.10066>
- Norris, S. P., & Phillips, L. M. (2009). Scientific literacy. In D. R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 271–285). Cambridge University Press.
- OECD. (2007). *PISA 2006 Science Competencies for Tomorrow's World: Volume 1: Analysis*. OECD Publishing. Retrieved May 5, 2023, from https://www.oecd-ilibrary.org/education/pisa-2006_9789264040014-en
- OECD. (2017). *PISA 2015 assessment and analytical framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*. OECD Publishing. Retrieved May 5, 2023, from <https://www.oecd.org/publications/pisa-2015-assessment-and-analytical-framework-9789264281820-en.htm>
- OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing. Retrieved May 5, 2023, from https://www.oecd-ilibrary.org/education/pisa-2018-assessment-and-analytical-framework_b25efab8-en
- OECD. (2023). *PISA 2025 Science Framework (Draft)*. OECD Publishing. Retrieved May 5, 2023, from https://pisa-framework.oecd.org/science-2025/assets/docs/PISA_2025_Science_Framework.pdf
- Ren, L., Zhang, C., & Guo, F. (2022). A study on the age-period-cohort effect of Chinese civic scientific literacy. *Studies in Science of Science*, 40(9), 1544–1554. <https://doi.org/10.16192/j.cnki.1003-2053.2022.09.001>. (in Chinese)
- Ren, L., Zhang, C., & He, W. (2013). Constructing and analysis of the model of how the factors affect the scientific literacy of Chinese citizens and a comparative investigation. *Studies in Science of Science*, 31(07), 983–990. <https://doi.org/10.16192/j.cnki.1003-2053.2013.07.004>. (in Chinese)
- Rios, J., & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, 26(1), 108–116.
- Roberts, D. A. (2010). Competing visions of scientific literacy: The influence of a science curriculum policy image. In C. Linder, L. Östman, D. A. Roberts, P. Wickman, G. Erickson, & A. MacKinnon (Eds.), *Exploring the landscape of scientific literacy* (pp. 11–27). New York: Routledge.
- Rutherford, F. J., & Ahlgren, A. (1991). *Science for all Americans* (1st ed.). Oxford University Press.
- Shen, B. S. P. (1975). Views: Science literacy: Public understanding of science is becoming vitally needed in developing and industrialized countries alike. *American Scientist*, 63(3), 265–268.
- The State Council of PRC. (2006). *Notice of the State Council on the Action Plans Outline for Nation's Science Literacy (2006—2010—2020 Year)*. Retrieved May 10, 2023, from http://www.gov.cn/gongbao/content/2006/content_244978.htm (in Chinese)
- The State Council of PRC. (2008). *Decision of the State Council on Implementing the Scientific Outlook on Development and Strengthening Environmental Protection*. Retrieved June 8, 2023, from https://www.gov.cn/zhengce/zhengceku/2008-03/28/content_5006.htm (in Chinese)

- Tian, F., & Dai, H. (2018). A study on the macro-influential factors of civic scientific literacy. *Globalization*, 4, 89–99+117+134. <https://doi.org/10.16845/j.cnki.ccieeqqh.2018.04.007>. (in Chinese)
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage Publications.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER Conquest Version 2.0*. ACER Press.
- Wu, S., Zhang, Y., & Zhuang, Z.-Y. (2018). A systematic initial study of civic scientific literacy in China: Cross-National Comparable Results from scientific cognition to sustainable literacy. *Sustainability*, 10(9), 3129. <https://doi.org/10.3390/su10093129>
- Wu, T., He, H., Wei, S., Zhu, P., Feng, Q., & Tang, Z. (2022). How to establishing an indicators framework for evaluating the performances in primary TB control institutions under the new TB control model? Based on a Delphi study conducted in Guangxi, China. *BMC Public Health*, 22, 2431. <https://doi.org/10.1186/s12889-022-14865-4>
- Zeidler, D. L., Walker, K. A., Ackett, W. A., & Simmons, M. L. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science Education*, 86(3), 343–367. <https://doi.org/10.1002/sce.10025>
- Zhang, F., Du, G., & He, W. (2016). On the status of civic scientific literacy of Chinese women and countermeasures: From a perspective of female citizens' social rights. *Studies on Science Popularization*, 11(03), 39–44+117. <https://doi.org/10.19293/j.cnki.1673-8357.2016.03.005>. (in Chinese)
- Zhang, F., Ren, L., & He, W. (2013). A study on the measurement of fair development index for the construction of Chinese civic scientific literacy. *Science & Technology Progress and Policy*, 30(16), 134–137. (in Chinese)
- Zhou, J., Yu, J., Qiu, J., & Zhou, Y. (2018). *The Spirit of science: Attribute and structural model*. Preprints. <https://doi.org/10.20944/preprints201808.0338.v1>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.