



OPEN ACCESS

EDITED BY

Zara Teixeira,
University of Évora, Portugal

REVIEWED BY

Marek Oziewicz,
University of Minnesota Twin Cities,
United States
Miloslav Kolenatý,
Masaryk University, Czechia

*CORRESPONDENCE

Monika Martin
✉ monika.martin@ph-freiburg.de

RECEIVED 29 May 2025

ACCEPTED 22 July 2025

PUBLISHED 25 August 2025

CITATION

Martin M, Stadler M, Künsting J,
Schwichow M, Asshoff R, Bender U, Birke F,
Carrapatoso A, Grundmeier A-M, Höger C,
Schuler S, Stemmann J and Rieß W (2025)
Assessing climate literacy in secondary
schools: development and validation of an
interdisciplinary competence test.
Front. Educ. 10:1637522.
doi: 10.3389/feduc.2025.1637522

COPYRIGHT

© 2025 Martin, Stadler, Künsting, Schwichow,
Asshoff, Bender, Birke, Carrapatoso,
Grundmeier, Höger, Schuler, Stemmann and
Rieß. This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Assessing climate literacy in secondary schools: development and validation of an interdisciplinary competence test

Monika Martin^{1*}, Magdalena Stadler¹, Josef Künsting¹,
Martin Schwichow¹, Roman Asshoff², Ute Bender¹,
Franziska Birke¹, Astrid Carrapatoso¹, Anne-Marie Grundmeier¹,
Christian Höger³, Stephan Schuler⁴, Jennifer Stemmann¹ and
Werner Rieß¹

¹Research Center for Climate Change Education and Education for Sustainable Development, Freiburg University of Education, Freiburg, Germany, ²Centre for Biology Education, University of Münster, Münster, Germany, ³Institute for Religious Education, University of Luzern, Luzern, Switzerland, ⁴Department of Geography, Ludwigsburg University of Education, Ludwigsburg, Germany

Introduction: A key goal of Climate Change Education (CCE) in schools is promoting climate literacy in students, that is, equipping them with the skills needed to engage in climate-related discourse and actions in an informed way. To determine whether CCE achieves this goal, comprehensive assessment is essential. However, existing assessment instruments focus narrowly on factual scientific knowledge of climate change and offer limited insight into students' broader climate literacy. This study presents the development of an interdisciplinary climate literacy test for secondary students, integrating perspectives from nine school subjects across the natural sciences, social sciences, technology, and the humanities.

Method: The development process involved a cognitive pretest ($N = 20$), two pilot studies ($N_1 = 353$, $N_2 = 313$), a teacher survey ($N = 36$), and a validation study ($N = 825$). We provide a validity argument supporting interpretation of the test score as a meaningful measure of CCE outcomes in secondary education.

Results: The test difficulty is suited for 9th-grade students (approximately 15 years old) across all school types in Germany. Empirical results support the theoretically derived four-dimensional structure of climate literacy, covering the four competence facets of (1) Dealing with content knowledge, (2) Knowledge generation and evaluation, (3) Information and communication, and (4) Normative evaluation. Correlations with external variables suggest that the test captures a school-related competence that is relevant to students' everyday lives.

Discussion: The developed test provides an interdisciplinary and detailed assessment of secondary students' climate literacy. We recommend its use for comprehensive evaluation of CCE efforts, enabling the design of more targeted and effective interventions.

KEYWORDS

climate literacy, competence assessment, secondary education, interdisciplinary, test development, Rasch analysis

1 Introduction

Climate change poses an urgent threat to both natural ecosystems and human societies (Cook et al., 2013; IPCC, 2022). Immediate action is needed to reduce greenhouse gas emissions and adapt to increasingly severe impacts such as rising sea levels, extreme weather, and resource scarcity (IPCC, 2022). Addressing this global crisis requires

transformative changes across all levels of society. Climate Change Education (CCE) plays a crucial role in this transformation by equipping individuals with the knowledge and skills to take informed action and support systemic change (e.g., UNESCO and UNFCCC, 2016; Otto et al., 2020). To advance CCE in schools and develop both effective and efficient educational interventions, meaningful insights into the current state of students' competencies as well as detailed information about particular deficits are needed.

While educational systems consistently evaluate their students' ability in reading, science, or mathematics, for example, within the PISA assessments (OECD, 2017), less is known about young people's understanding of climate change and their ability to act. For both monitoring the current state of CCE and deriving further interventions, a valid and reliable instrument is needed. However, previous attempts to assess student outcomes in CCE have several shortcomings, providing only a limited view of students' actual competencies. Current instruments often focus primarily on factual knowledge and adopt a narrow natural science perspective while neglecting social science aspects (Lubej et al., 2025). Moreover, questions and problems are often presented in an abstract way with little connection to students' everyday lives (Lubej et al., 2025). Yet, school education aims not only to convey essential knowledge about the scientific processes and consequences of climate change but also to empower students to engage in public discourse and make informed decisions (e.g., Redman and Wiek, 2021; Kranz et al., 2022). Therefore, a comprehensive evaluation of students' climate change-related competencies should go beyond factual knowledge, integrate perspectives of both natural and social science as well as technology and the humanities, and state problems in a way that they are relevant to students' everyday experiences.

In the present study, we therefore took a comprehensive and interdisciplinary view on climate literacy (United States Global Change Research Program, 2009). We present a test designed to assess cognitive components of adolescents' climate literacy based on an interdisciplinary framework (Stadler et al., 2024). We further describe the process of validating its score interpretation using data collected from both student and teacher samples. The test can then be used to evaluate the effectiveness of CCE measures in school and to inform the development of targeted interventions.

1.1 Fostering climate literacy

While definitions of climate literacy vary, three key elements are commonly recognized: first, individuals must have a solid conceptual understanding of Earth's climate system, along with procedural and epistemic knowledge of climate science. Second, they need the ability to access and critically evaluate scientifically credible information on climate change, as well as effectively communicate about it. Finally, climate literacy includes attitudes and values that support responsible decision-making regarding adaptation and mitigation strategies (United States Global Change Research Program, 2009; DeWaters et al., 2014; Azevedo and Marques, 2017; Kuthe et al., 2019). This broad definition highlights the connection between climate literacy and the related concept of scientific literacy. A scientifically literate individual not only possesses knowledge of and about science but also demonstrates key attitudes, such

as a willingness to engage with scientific issues (OECD, 2017; Sjöström and Eilks, 2018). In this sense, climate literacy can be considered a specific application of the broader concept of scientific literacy (Bybee, 2012; Azevedo and Marques, 2017).

The nature of the literacy construct comprising both cognitive and affective-motivational components requires an assessment that distinguishes between cognitive components, such as knowledge, and affective-motivational components, such as attitudes (Klieme et al., 2008; Kahan, 2015). In this study, we focus on the cognitive part of climate literacy (i.e., conceptual knowledge and the ability to apply this knowledge to solve corresponding tasks), as we see a particular lack of validated instruments in this area compared to a broad research body on attitudinal measures stemming from environmental education (Kaiser et al., 2007; Bogner, 2018). Moreover, although CCE addresses affective-motivational parts as well (Cantell et al., 2019), the cognitive part of climate literacy is the primary focus of school curricula (Leve et al., 2023, 2025). As one intended purpose of our test instrument is to assess the extent to which school systems facilitate climate literacy, it must have the same focus as school curricula. In the following, we will use the term climate literacy to describe the cognitive aspects.

The central task of promoting climate literacy among students is not confined to a single school subject but is better understood as a cross-curricular effort that should be taught in an interdisciplinary manner (Leve et al., 2023; Hargis, 2024). Currently, climate change is primarily addressed in geography and science classes (Siegmund Space and Education, 2021). However, there is a growing call to also incorporate perspectives from the social sciences and humanities to fully capture the complexity of climate change as a socio-scientific issue (Shwom et al., 2017; Eames et al., 2024), including, for example, considerations of intergenerational justice or political measures for climate change mitigation (Kranz et al., 2022).

When it comes to the concrete enactment of CCE in school, researchers and political actors often call for "novel" or "innovative" methods (see Jorgenson et al., 2019; Monroe et al., 2019 for reviews). Especially holistic and action-oriented approaches like the Whole School Approach (e.g., Holst et al., 2024) are suggested as promising for advancing CCE, and thus their use is also supported on a policy level (UNESCO, 2020). However, their evaluation relies on self-report measures of knowledge or skills (e.g., Holst et al., 2024; Keller et al., 2024). Relying solely on self-report measures can be problematic as students typically do not or cannot assess their own knowledge validly (Stoutenborough and Vedlitz, 2014). Moreover, a recent meta-analysis summarized the effectiveness of interventions aimed at fostering climate literacy in formal elementary and secondary education (Aeschbach et al., 2025). Although the authors found an overall medium effect of CCE interventions on students' cognitive outcomes (here: knowledge), they point out that a large proportion of studies used *ad hoc* measures that had not previously been validated making it difficult to compare results across studies and to derive recommendations for practice (Aeschbach et al., 2025). Thus, in order to evaluate and compare the effectiveness of these highly demanded "novel" approaches to CCE, objective and valid measures of cognitive learning outcomes are needed.

1.2 Current assessment of climate literacy

Research on CCE outcomes, whether cognitive, motivational, or behavioral, has increased greatly during the past decades (García-Vinuesa et al., 2024; Segade-Vázquez et al., 2025). Thereby, various instruments are used to assess cognitive CCE outcomes in high school students (e.g., Stevenson et al., 2014; Bedford, 2016; Kuthe et al., 2019), educators (McNeal et al., 2014; Boon, 2016), and the general public (e.g., Tobler et al., 2012). However, when considering the comprehensive understanding of climate literacy as defined by the United States Global Change Research Program (2009) provided above, several shortcomings of the often-used tests and questionnaires can be noted. First, these instruments typically assess only factual knowledge—often in a true/false response format (e.g., Tobler et al., 2012; Stevenson et al., 2014; Kuthe et al., 2019). However, simple true/false questions fail to capture the specific alternative conceptions students have as well as difficulties students typically struggle with, for example, concerning the complex interconnections within the climate system (Shepardson et al., 2014; Schauss and Sprenger, 2021), the greenhouse effect (e.g., Schubatzky et al., 2024), or mitigation and adaptation strategies (Bofferding and Kloser, 2015). Moreover, such a question format bears the risk of oversimplifying the complexity of the climate system and its cause-effect relationships (e.g., “The floods and droughts in the Amazon are caused by climate change,” Higuchi et al., 2018; “Without greenhouse gases Europe would be an ice desert,” Deisenrieder et al., 2020). Finally, guessing probability is relatively high for true/false items, which may result in ceiling effects. In contrast, knowledge tests with single-choice or multiple-choice test items (e.g., Libarkin et al., 2018; Schubatzky et al., 2024) are less affected by guessing and therefore allow a more valid estimate of a person’s knowledge.

A second limitation of existing assessment instruments is their predominant focus on the natural science aspects of climate change, such as the greenhouse effect (e.g., Flora et al., 2014; Schubatzky et al., 2024), and on knowledge about individual mitigation behaviors like energy conservation and recycling (Tobler et al., 2012). However, if CCE is to be promoted as an interdisciplinary effort spanning multiple school subjects, including the social sciences, technology, and the humanities (e.g., Leve et al., 2023), then assessment tools must reflect this broader scope. Additionally, given that individual actions have limited impact on climate change mitigation, a valid assessment of action and effectiveness knowledge (Roczen et al., 2014) needs to be expanded to include knowledge of political and economic measures, which often have significantly greater impact (Kranz et al., 2022).

Finally, to effectively inform educational measures, outcome assessment should go beyond providing a single global indicator and instead offer nuanced insights into the various facets of the targeted competencies, as only the diagnosis of specific weaknesses enables adaptation of educational interventions. For example, in the domain of physical science conceptual knowledge, well-designed concept inventories can validly and reliably uncover common alternative conceptions students hold, such as misunderstandings about the greenhouse effect (Schubatzky et al., 2024). Insights gained from such detailed assessments can then inform the development of targeted interventions that address exactly those areas where students still need support. To advance

CCE, it would be valuable to extend this approach of fine-grained diagnostic assessment beyond the physical science content as well as beyond the focus on conceptual knowledge.

To summarize, there is a gap between the aims of CCE on the one hand and the instruments used for assessing whether these aims are met on the other hand. While the desired outcome of CCE, climate literacy, is defined as a multifaceted and interdisciplinary competence, current instruments merely assess reproduction and recognition of factual knowledge or self-reported knowledge, often with a strong focus on the natural sciences (Lubej et al., 2025). Systematically evaluated measurement instruments that enable a comprehensive, accurate, and valid assessment of cognitive outcomes of CCE are still lacking (Aeschbach et al., 2025).

Thus, we aimed to develop a test to assess climate literacy that meets three key criteria essential for advancing research on effective CCE. First, the test should cover the full range of knowledge crucial for dealing with climate change, incorporating not only the scientific basics but also the technological, social, political, and economic dimensions, reflecting the complex, interdisciplinary nature of climate change as a socio-scientific issue (Shwom et al., 2017). Second, it should be a valid measure of the competencies adolescents need to engage with climate change-related issues in their daily lives, requiring them not only to reproduce factual knowledge but also to apply it and critically engage with new information. Finally, the test results should provide detailed insights into the specific areas where educational efforts should be directed to in order to improve climate literacy among students.

To achieve these objectives, test development in the present study is grounded in an interdisciplinary framework of climate literacy. As research on sustainability education and CCE is currently criticized for a lack of connections to established theories and methods in educational research (Gräsel et al., 2013), the model closely aligns with existing definitions of the related construct of scientific literacy (Kampa and Köller, 2016; OECD, 2017). In the following section, we outline this interdisciplinary framework and describe how it informed the development of the assessment.

1.3 An interdisciplinary framework of climate literacy

In accordance with the definition of climate literacy (United States Global Change Research Program, 2009), Stadler et al. (2024) have constructed an interdisciplinary framework that serves as a theoretical foundation for the development of the competence test. The framework focuses on the cognitive part of climate literacy—namely, the knowledge and abilities essential for understanding and addressing climate change—while deliberately avoiding conflation with subjective beliefs or motivational aspects. To ensure compatibility with existing educational evaluation initiatives, both nationally and internationally, the authors sought to build on well-established competence models, including the scientific literacy framework of PISA (OECD, 2017) and the competence model used to evaluate the educational standards in Germany (Stanat et al., 2021).

While the literature on CCE goals typically refers to largely content-independent key competencies, such as critical thinking

TABLE 1 Description of the four competence facets of climate literacy.

Competence facet	Description
Dealing with content knowledge	Students know and understand phenomena of the climate system and climate change, explain them with reference to findings from climate science and make predictions based on them, and can derive, describe, and justify suitable actions (including strategies and measures) for dealing with climate change.
Knowledge generation and evaluation	Students know and understand how research is conducted in climate science to derive the most robust findings about climate change. They can interpret and evaluate climate-related statements in terms of their data and evidence. This competence area can further be differentiated into two aspects, namely (1) generating, processing, and interpreting questions and hypotheses, and (2) reflecting on the theory of science. The latter involves awareness of the characteristics of knowledge acquisition processes, climate science knowledge, and models.
Information and communication	Students can access information on climate change and select relevant and meaningful information and data from different forms of presentation (e.g., texts, tables, graphs, diagrams, and images). Moreover, they can examine various sources and forms of representation for their consistency as a prerequisite for engaging in argumentative exchange and communication with others based on this information. This competence area primarily addresses those forms of communication that occur through media (e.g., social networks, video platforms, newspapers, television) and play an important role in the public discourse on climate change.
Normative evaluation	Students evaluate findings of climate research and statements of people on climate change in normative terms. They can classify situations in which an evaluation is to be conducted, as well as the evaluation criteria relevant to these situations. A particular focus lies on evaluating action options and their impacts, which requires perspective-taking and impact assessment.

(Mochizuki and Bryan, 2015) and dealing with uncertainty (Stevenson et al., 2017), it is essential to acknowledge the context-dependency of competencies as an important factor in competence modeling (Klieme et al., 2008; Weinert, 2001). Accordingly, a summary of essential factual and conceptual knowledge forms the content base of the climate literacy framework. To this end, the framework is guided by an interdisciplinary approach (Stadler et al., 2024). The initial stage involved group discussions with experts from various disciplines and their didactics, namely biology, geography, physics, technology, politics, economics, food and consumption, fashion and textile, and religious and ethical education. Thereby, experts synthesized the content and objectives identified in existing literature on CCE into 12 fundamental concepts that represent the essential knowledge students should acquire about climate change (a translation of the German full list of the fundamental concepts formulated is provided in the [Supplementary material](#)). These 12 fundamental concepts can be subsumed under four content areas, namely (1) Scientific basics of the climate system, (2) Causes of climate change, (3) Impacts of climate change, and (4) Action options and barriers, which align with previous work by other authors (Adamina et al., 2018). An evaluation by another group of experts in the areas of climate sciences (meteorology, geology, climatology), geography, and sustainability governance was conducted to ensure coverage of all relevant content from the perspective of the academic disciplines (Stadler et al., 2024).

To further clarify how climate literacy entails applying this essential knowledge, Stadler et al. (2024) formulated four key competence facets. These competence facets draw upon the aforementioned established competence models of scientific literacy (OECD, 2017; Stanat et al., 2021) and include (1) Dealing with content knowledge, (2) Knowledge generation and evaluation, (3) Information and communication, and (4) Normative evaluation. Table 1 provides descriptions of the four competence facets. Following Stadler et al. (2024), a competence test assessing the broad construct of climate literacy should address all four competence facets. While conceptual knowledge as summarized in the four content areas presented above is most prominently addressed in the competence facet of *Dealing with*

content knowledge, it nevertheless provides the contextual frame for all competence facets (see Figure 1).

1.4 The present study

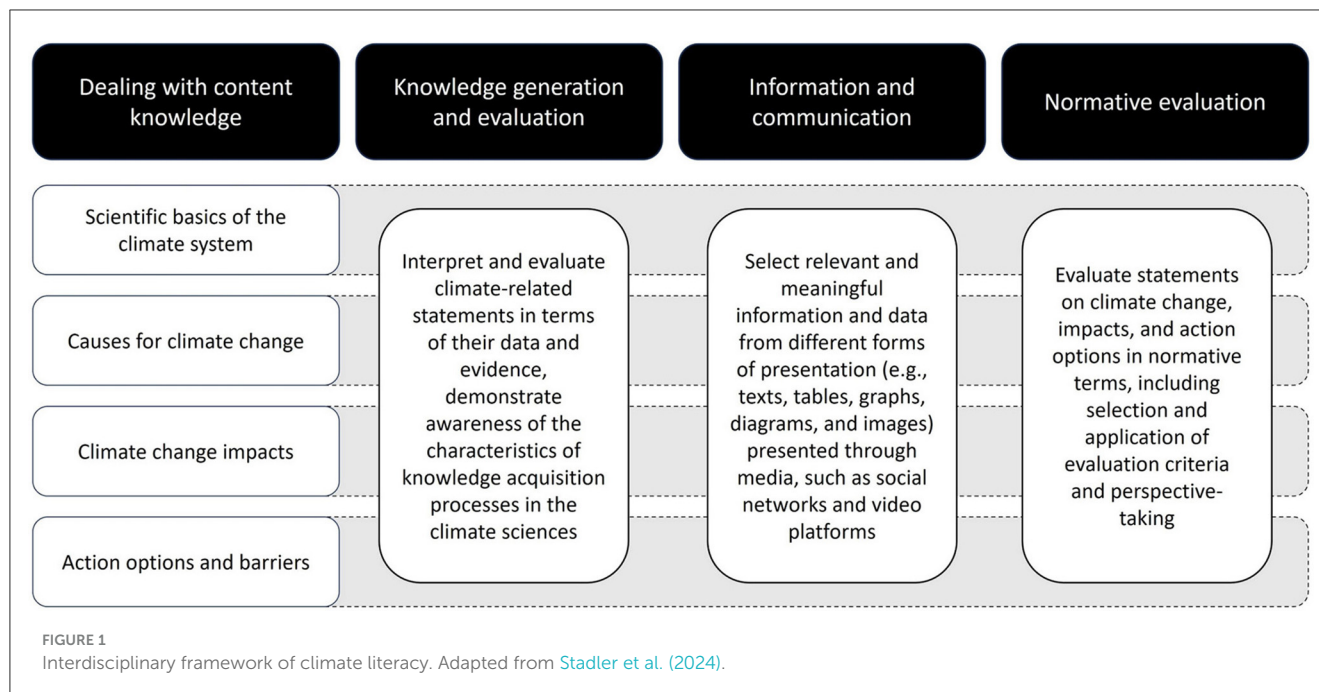
In the following, we present the development process of a climate literacy competence test and evaluate arguments for the validity of the intended interpretation of the test scores. Following the current discussion on test score validation, validity is not considered a feature of a test, but more a “process of constructing and evaluating arguments for and against the intended interpretation of test scores and their relevance to the proposed use” (AERA et al., 2014, p.11). Thus, in the validation process, researchers need to gather arguments that support the intended use and interpretation of a test score (Kane, 2013; AERA et al., 2014).

Following this argumentative approach of validity, we first define the intended use and interpretation of the test scores along with claims that back this intended interpretation. For each validity claim, we provide one or multiple theoretical and empirical sources of evidence that we will use to evaluate the claim. These sources include the item development procedure, a cognitive pretest and two pilot studies with students, an expert review of the test items with schoolteachers, and a main study with the final item pool.

1.5 Validity argument

1.5.1 Intended use and test score interpretation

The test should provide comprehensive insights into the climate literacy of students at the end of their compulsory schooling in Germany. It can be used to evaluate the primary outcome of CCE measures, that is, to what extent students are able to participate in social discourse of climate change and to engage with climate-change related issues in an informed way. Therefore, the test should be sensitive to detect differences in proficiency within the target group of German adolescents. Moreover, the test score should provide nuanced insights into students’ climate literacy to allow for targeted development of educational measures.



1.5.2 Claim 1: all aspects of climate literacy are represented; no aspect is over- or underrepresented. All items are relevant for the construct represented

Sources used for evaluation: We report on the item development process conducted by an interdisciplinary team, including perspectives from experts across nine school subjects. The interdisciplinary framework of climate literacy (Stadler et al., 2024) served as guidance for item development to ensure comprehensive coverage of competence and content aspects. To evaluate the test content's relevance for CCE in secondary school, a teacher sample of diverse backgrounds rated the items in terms of their suitability for assessing climate literacy. Following the standards of educational testing, this evidence can be regarded as evidence based on *test content* (AERA et al., 2014).

1.5.3 Claim 2: test difficulty is suited for German students in 9th and 10th grade. The test is sensitive to differences in proficiency levels typical for the target group

Sources used for evaluation: We conducted a cognitive pretest including think-aloud interviews to investigate whether students engaged in the intended cognitive processes. Two sequential pilot studies informed the adaptation of stimuli and answer options, providing evidence based on *response processes*. In a final validation study, we analyzed item difficulties using IRT modeling (Embretson and Reise, 2000) to evaluate their distribution relative to the proficiency levels of the student sample. To complement the empirical analysis, we collected schoolteachers' estimations of item difficulty for the respective age group. These expert judgments offer an additional perspective on the appropriateness of test difficulty for the target population. Finally, to ensure test fairness, we investigate differential item functioning (DIF; Ackerman, 1992)

across school types. While we expect differences in the overall test score between students from different educational tracks to be similar to those typically found for German students' scientific literacy (e.g., Stanat et al., 2021), high DIF items would constitute an unfair disadvantage for students of one school type. Test fairness addresses the evidence category of *consequences of testing*.

1.5.4 Claim 3: the assessment instrument represents the four-dimensional structure of climate literacy with four separable competence facets as outlined in the climate literacy framework

Sources used for evaluation: We conduct model comparisons on the main study data, evaluating relative fit indices such as Akaike Information Criterion (AIC) and sample-size adjusted Bayesian Information Criterion (BIC). We expect a four-dimensional model, where each item loads on a specific competence facet, to provide a better model fit compared to a one-dimensional model, in which all items load on a single latent trait. Comparing these concurrent models provides evidence based on *internal structure*.

1.5.5 Claim 4: the competence measured is distinct from generic cognitive abilities and can be shaped through school education

Sources used for evaluation: We investigate correlations of the test score with standardized tests of reading ability and general cognitive abilities. We expect moderate correlations of the test with general cognitive ability, aligning with findings from prior assessments of competencies, such as scientific literacy (Baumert et al., 2009). Furthermore, we expect the test scores to be moderately related to a standardized reading test, given that reading comprehension is a crucial prerequisite for engaging with

complex test items that assess application of knowledge rather than reproduction or recognition. However, we expect only moderate correlations (i.e., $0.20 < r < 0.40$; Cohen, 1988) of climate literacy with these generic constructs to ensure that the developed test measures a domain-specific construct rather than merely reflecting general intelligence. To further evaluate in what way the test can capture school efforts in fostering climate literacy, we compare mean test scores across school types. We expect students at upper-track secondary schools to outperform those at middle- and lower-track schools, with a moderate to large effect size (i.e., $d > 0.5$; Cohen, 1988), consistent with findings from national educational assessments in the area of scientific literacy (e.g., Stanat et al., 2021).

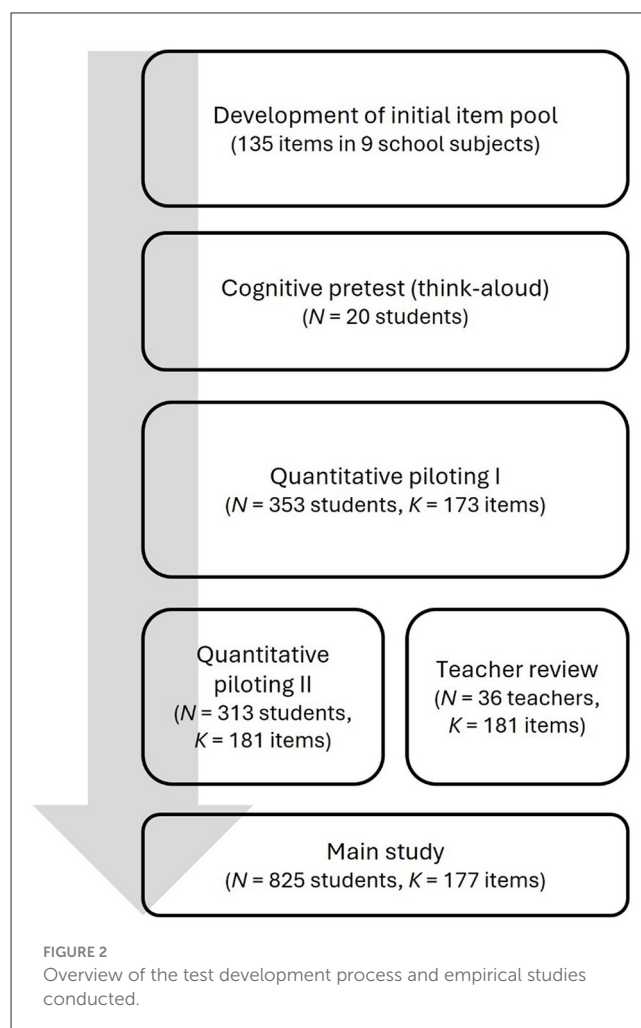
1.5.6 Claim 5: climate literacy assessed with this instrument is relevant for teenagers' active participation in social climate change discourse

Sources used for evaluation: We examine correlations of the test scores with motivational-affective and behavioral components of climate literacy, thereby addressing the evidence category of *relations to external variables*. Previous studies have shown that knowledge of causes and consequences of climate change was significantly related to attitudes about climate change (Tobler et al., 2012; Stevenson et al., 2014; Loy, 2018). Moreover, prior research suggests positive relations of more general environmental knowledge and pro-environmental behavior (Roczen et al., 2014). Thus, we expect medium positive correlations of the climate literacy test score with both climate change-related attitudes and behavior.

2 Method

2.1 Item development and pilot testing

The item development process was grounded in the interdisciplinary climate literacy framework presented above. Figure 2 provides an overview of the steps involved in the development process. First, teacher educators from nine educational disciplines (biology, geography, physics, technology, politics, economics, food and consumption, fashion and textile, and religious and ethical education) developed 15 test items each, resulting in 135 items in total. The items were clustered into units of three to seven items addressing the same topic (e.g., renewable energies, fast fashion, sea level rise, etc.). To achieve a balanced set of items across the competency model as well as to standardize the item features, item construction followed several rules. First, each item should focus on one of the four competence facets and cover primarily one of the content aspects (see Figure 1). Second, the items should use different answer formats (single-choice, multiple-choice, and short response) and vary in the cognitive processes evoked (Bloom et al., 1956; Krathwohl, 2002). Thereby, we tried to prevent that items follow primarily a certain style that is typical for a particular subject. For example, an item from the area of religious and ethical education required the students to assign typical statements about consumption (e.g., meat-based diet) to their underlying values (e.g., tradition) and thereby resembled in its structure a typical item from the natural sciences that required



students to assign typical statements to whether they referred to the concept of weather or climate. Third, to ensure relevance to adolescents' everyday lives, items should be embedded in real-world contexts (e.g., social media comments and peer discussions), whenever possible. Finally, the 15 items from each subject area should cover a broad range of difficulties.

The initially developed test items of each subject area were thoroughly reviewed by a second educational expert with a background in a different subject to ensure clarity and to eliminate ambiguity in the intended correct answers. This step was especially valuable for social science items, addressing topics such as responsibilities for mitigation efforts or the consequences of political decisions that are often more ambiguous than physical or biological facts, such as the concentration of greenhouse gases in the atmosphere.

We pretested the initially developed items in think-aloud cognitive pretests with 20 9th-grade students from various school types. Thereby, each student received two to three units resulting in a testing time of about 30 min per student. Additionally, we handed out test booklets to five school classes during substitution lessons. We asked the students to circle any words they did not understand. We used the information gained from both the think-aloud sessions and the written feedback to revise the items. Moreover, we used

TABLE 2 Overview of the student samples in the two quantitative pilot studies and the main study.

Sample information	Pilot 1	Pilot 2	Main study
Data collection			
Time period	Nov 2022 – Dec 2022	Mar 2023 – Apr 2023	Jun 2023 – Jul 2023
N students	353	313	825
School grade	10	10	9
Distribution of students in upper/middle/lower track schools (in %)	36/52/12	79/21/-	59/29/12
Student characteristics			
Age mean (SD)	15.9 (0.7)	16.1 (0.6)	15.6 (1.0)
Gender distribution (in %): Male/Female/Other/Not indicated	47/47/6/-	44/52/4/-	44/47/5/4
Main Language spoken at home (in %): German/Other/Not indicated	80/20/-	83/17/-	77/19/4

the students' associations to the items uttered during think-aloud to develop another 38 items covering topics not yet addressed so far.

Building on this work, we conducted two quantitative pilot studies with students in grade 9 and 10 ($N_1 = 353$ and $N_2 = 313$; see Table 2 for details on the samples) to obtain information about the functioning of the individual items and of the test. Specifically, we analyzed item functioning within item response theory (IRT; Embretson and Reise, 2000) and relied on item difficulty and discrimination parameters as well as item infit (WMNSQ) to identify poorly functioning items.

Simultaneously to the second pilot study, we presented the test items to 36 teachers (18 male, 18 female) in an online survey. All teachers taught at German secondary schools (upper track) and had on average 14.6 years of teaching experience ($SD = 9.5$; min = 1, max = 34). We ensured a diverse background of the teachers in both the natural and the social sciences. All teachers specialized in two or more subjects with most mentions in chemistry (9), biology (8), geography (7), and physics (6). Technology, religious and ethical education, economics/business, and social studies were named by four teachers each. As it would take too long to thoroughly comment on all items for one person, we presented each teacher only a subset of 20 to 25 items. The questionnaire presented each test item on one page along with a question on whether the test item is suited for assessing climate literacy (rated as 1 = no, 2 = rather no, 3 = rather yes, 4 = yes), a question on the estimated difficulty of the item for 9th graders (rated on a five point scale from "very easy" to "very difficult"), and an open question to provide an additional comment on the item (e.g., potential problems with assessment, improvements in wording).

We used both the data from the pilot studies and the teacher survey to remove and revise suboptimal test items. While a detailed presentation of the pilot study data and the resulting revisions would exceed the scope of this article, Table 3 provides a short overview of exemplary decisions and actions taken based on the pilot studies and teacher survey.

The final item pool for the main study consisted of 177 items in 43 thematically grouped units. Figure 3 shows two example items. The item "CO₂-tax" is a multiple-choice item. It is assigned to the content area of *Action options and barriers* and the competence facet of *Information and communication*. For this item, all essential information needed for answering correctly is provided in the item stem and the graph. The item "Food waste" is a single-choice item and addresses the role of food waste in climate change. It is assigned to the content area of *Causes for climate change* and the competence aspect of *Dealing with content knowledge*. Here, students must apply their prior knowledge of energy consumption and greenhouse gas emissions to arrive at the correct answer.

2.2 Main study

In the following, we report about the sample and assessment procedure of the main study. Additionally, we report on initial item quality analyses based on item response theory that led to the final selection of items for further analysis.

2.2.1 Sample

We obtained data from 825 9th-grade students (357 male, 387 female, 44 other, 37 not answered) in 17 secondary schools in the federal state Baden-Württemberg. In Germany, the secondary schools typically are organized in three tracks (upper, middle, and lower secondary school), which typically differ in their demographic characteristics. Of the 17 schools, nine were in the upper or academic track ("Gymnasien"), five were in the middle track ("Realschulen"), and three were a mixed form of middle and lower track students ("Haupt-/Werkrealschulen" and "Gemeinschaftsschulen"). In the total sample, the mean age was 15.6 years ($SD = 0.98$), and 154 students (18.7%) indicated another language than German as the main language spoken at home (for details on the main study sample, see Table 2). All studies involving school students (i.e., cognitive pretests, quantitative pilot studies, and main study) were approved by the ministry of education and cultural affairs of the federal state Baden-Württemberg. Informed consent for all studies was obtained from the students' parents. We did not pay any compensation, but the participating schools received feedback on the results of their school compared to the other schools as an incentive.

We had to exclude 40 students because their response times to the closed items as well as written responses to the open response items indicated that they did not take the test seriously. Specifically, we excluded students if either their answers to three or more open response items were coded as "not serious" (e.g., random letter series or insults against classmates or politicians) or the proportion of items answered in <3 s exceeded 20%. After these exclusions, the final sample consisted of 785 students.

TABLE 3 Exemplary decisions on item revisions based on the pilot studies in student and teacher samples.

Item information	Pilot 1	Pilot 2	Teacher survey
N items	173	181	181
Item difficulty ^a (<i>P</i>)	<i>M</i> = 0.42 <i>SD</i> = 0.21 Min = 0.06 Max = 0.88	<i>M</i> = 0.45 <i>SD</i> = 0.23 Min = 0.03 Max = 0.96	Not applicable
Exemplary revisions in item pool based on the data collected	<ul style="list-style-type: none">• Reduction of reading load by splitting longer items• Focused on developing easier items to cover the full range of student abilities	<ul style="list-style-type: none">• Used student answers to former short response items to create closed items• Decided to use dichotomous items only instead of a mixture of dichotomous and partial credit items	<ul style="list-style-type: none">• Added useful distractors representing typical student conceptions brought up by teachers• Removed items criticized unanimously by teachers of multiple subjects

^aItems in pilot study 1 and pilot study 2 contained 35 and 28 partial credit items, respectively, which were scored 0/1/2. The item difficulty (*P*) provided here indicates the percentage of students scoring at least one point. Thus, the mean and range cannot be directly compared to the data from the main study which contained dichotomous items only.

2.2.2 Materials

When measuring broad constructs, such as climate literacy, requiring many items to cover all relevant sub-facets, the number of items and thus testing time for each student can be reduced by assigning items to different test booklets (Frey et al., 2009). We arranged the 43 test units to seven booklets using a Balanced Incomplete Block Design (Frey et al., 2009; Frey and Bernhardt, 2012) to reduce test time to a feasible level. The booklet design also allowed us to control for position effects, that is, for example, overestimating the difficulty of the items positioned at the end of the test. Each student was randomly assigned to one of seven test booklets, each comprising between 73 and 78 items. The average difficulty (*P*) was similar for all seven booklets with mean difficulties ranging from 0.32 for the most difficult booklet to 0.40 for the easiest booklet. A detailed description of how we compiled the booklets can be found in the [Supplementary material](#).

To investigate relations of the climate literacy test score with other cognitive constructs, we assessed general cognitive abilities and reading ability with standardized tests. Moreover, we assessed climate change attitudes and self-reported behavior with questionnaires.

2.2.2.1 General cognitive abilities

We used the non-verbal subscale *matrices* of a general cognitive ability test (CFT 20-R; Weiß, 2006) to measure a central part of fluid intelligence. We selected this subscale because it has shown a high correlation with the total score in the CFT (Visser et al., 2022; Weiß, 2006). The scale consists of 15 items, with each item presenting three connected patterns. Students select one pattern out of five options to complete the set provided. The students had 3 min to complete the test and were automatically forwarded to the next test. Reliability for the matrices test was good with Cronbach's $\alpha = 0.82$.

2.2.2.2 Reading speed and fluency

We used the reading speed and fluency test for students in 5th to 12th grade (Schneider et al., 2017). A particular advantage of this test is that it measures reading comprehension economically and independently of prior knowledge (Schlagmüller et al., 2022). The test comprises a fairy tale of 2,006 words, including a total of 47 gaps spread across the text. For each gap, students have to select the word that fits the context best out of three options. After 6 min, they were automatically redirected to the next page. We assessed the reading

speed and fluency test only in a randomly drawn subsample of 218 students. Reliability was good with Cronbach's $\alpha = 0.83$.

2.2.2.3 Climate change attitude

We assessed attitude toward climate change with an adapted version of the General Ecological Behavior scale (GEB-40; Kaiser et al., 2007), which has been demonstrated to be an effective tool for measuring general attitudes in the environmental domain. This scale is grounded in the assumptions of the Campbell paradigm, which posits that an individual's attitudes toward a particular topic (here: climate change mitigation) are reflected in their engagement in behaviors that support the goal associated with that topic (Kaiser, 2021). Specifically, the stronger a person's attitudes, the greater their willingness to overcome potential barriers in pursuit of this goal. Additionally, the overarching attitude should become apparent in behaviors across different areas (Kaiser et al., 2007). Building on these assumptions, a person's attitude can be estimated from their self-reported behavior.

The adapted scale GEB-40-climate consists of 40 items covering everyday behavior, of which the students answer 34 on a five-point scale (1 = never, 5 = very often; example item: "When I make notes on paper, I use sheets of used paper that have already been printed on one side.") and six in a dichotomous format (yes/no; example item: "I am a member of an environmental organization."). The scale includes behavior in various areas (e.g., food and consumption, mobility, energy use, non-activist public behavior). In total, 29 items are positively worded, and 11 items are negatively worded. We computed person estimates using a one-dimensional Rasch model (see Kaiser et al., 2007 on the scaling of the GEB-40).

2.2.2.4 Climate-friendly behavior in the public sphere

Following the current discussion on collective agency in the context of CCE (Kranz et al., 2025), we focused on students' public mitigation behavior rather than their private consumption behaviors. As a single-item measure for a high impact public mitigation behavior (Nielsen et al., 2021), we asked the students whether they have already participated in the Fridays for Future demonstrations.

Carbon Pricing

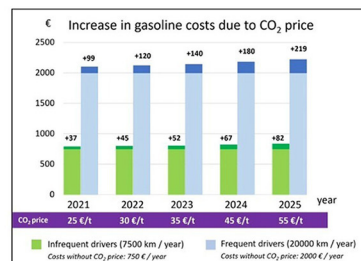
Question 4 of 6

CO₂ tax

Governments may introduce a CO₂ tax to mitigate climate change. This means that companies must pay a certain amount of tax to the state for each ton of CO₂ a certain amount to the state in the form of a tax.

How the CO₂ tax works

25€ per ton of CO₂ produced - that's what companies will have to pay from January 1, 2021, if they sell diesel and gasoline, heating oil and natural gas. By 2025, this amount will rise to 55€ per ton.



Which conclusions can you draw from the graph provided?

Select one or multiple options.

- ☐ In 2023, CO₂ will cost twice as much as in 2021.
- ☒ Frequent drivers are especially impacted by the CO₂ tax.
- ☐ The price of gasoline is higher for frequent drivers than for infrequent drivers.
- ☒ The CO₂ price will rise more sharply from 2024 than in previous years.

Next

Food waste

Question 1 of 5

Food waste

Worldwide, around a third of the food produced for humans is not consumed by humans. This has an impact on the climate.

What is the main reason that the large amount of food waste is a driver of climate change?

Select one answer.

- ☒ The production of food that is thrown away emits unnecessary greenhouse gases.
- ☐ When food waste is left in nature, it pollutes the environment.
- ☐ When food rots, harmful gases are released.
- ☐ Food waste often ends up in the trash along with the packaging, making it difficult to recycle.

Next

FIGURE 3

Example items. The figure shows the items as they were provided in the computer-based assessment environment the students worked in. In the gray area on the left, the title of the current unit as well as the item number within the unit is displayed. The dark blue bars indicate the number of completed and upcoming units, respectively. Rectangular checkboxes allow selecting multiple answers, round boxes allow one answer only.

2.2.3 Procedure

Data collection took place in classrooms over three consecutive lessons, including two short breaks. Students worked individually on laptops or desktop computers. After an introduction into the different question formats, students

were given up to 70 min to work on the competence test. We instructed them to read and answer the questions carefully but not spend excessive time on single items. After the competence test, students worked on the standardized cognitive ability and reading test, and finally answered

the questionnaires on climate-change related attitudes and behavior.

2.2.4 Scoring

We scored all items dichotomously as solved completely (1) or not solved (0). The 13 short response items were coded according to a coding manual comprising definitions of wrong and correct responses as well as multiple examples per category. One trained rater coded all student responses. To determine interrater reliability, a second trained rater coded a random sample of 30% of the responses per item. For all items, the raters reached sufficient to excellent agreement with $0.76 < \kappa < 0.98$ (Landis and Koch, 1977).

We treated missing values similar to PISA (OECD, 2024): items at the end of the booklet that were not reached within the test time was coded as missing. The test is designed as a power test and not as a speed test (Kaplan and Saccuzzo, 2001), meaning that we gain only limited information about a person's ability from the sheer number of items completed. Moreover, our main goal in this study is to assess the quality of the test items rather than a precise measure of person ability. Therefore, if we had coded items that were not answered due to a lack of time as incorrect, we would have overestimated their difficulty. In contrast, items within the test that were skipped without entering a response were coded as incorrect (0). Additionally, items that were answered in up to 3 s were coded as missing regardless of if the given answer was correct or incorrect. Our rationale for this decision was that giving a serious answer in < 3 s was not possible, and thus, the submitted answer would not provide valid information about a student's ability to solve this item. While for the PISA assessments this threshold for rapid-guessing typically is set to 5 s (Michaelides and Ivanova, 2022), we chose the more conservative criterion of 3 s to make sure not to lose too much information.

2.2.5 Analyses

We utilized the Rasch model (Rasch, 1960; Boone, 2016) to compute person and item measures which were used in further analyses to answer the research questions. Based on our theory driven operationalization of climate literacy and our systematic item development we expect that the items of our instruments fulfill the requirements of Rasch analyses: the test items can be answered independently from each other and their difficulty varies within each competence facet. Moreover, we expect climate literacy to be multi-dimensional as it consists of related but nevertheless different competence facets. The Rasch model allows testing the dimensionality of test instruments by comparing models (one-dimensional model vs. a four-dimensional model representing the competence facets) regarding the final deviance of the model estimation and information criteria such as AIC and BIC. However, to investigate the dimensionality of our test measures, theoretical considerations about the dimensionality of climate literacy were considered too, as we utilized the latent modeling in a confirmatory manner (Edelsbrunner and Dablander, 2019).

We performed an initial Rasch-scaling with all items using the R package TAM version 4.2-21 (Robitzsch et al., 2024) with a marginal maximum likelihood (MML) estimator. We specified a one-dimensional Rasch model with dichotomously scored items.

We had to exclude one item before starting the analyses because we noticed an erroneous item formulation after data collection making the item impossible to solve, leaving 176 items for the initial scaling. We examined item infit (WMNSQ) to assess item quality. Typically, items with a WMNSQ between 0.8 and 1.2 demonstrate an acceptable fit to the Rasch model (Bond and Fox, 2007). Moreover, we visually inspected item characteristic curves (ICC) for unusual patterns deviating from the expected slope. Finally, we examined item-test correlations as a classical measure of item discrimination. While we excluded all items with negative discrimination as this indicates poor item quality in all cases, we did not use a strict cutoff for low discrimination, such as $r_{it} > 0.25$ (e.g., Kampa and Köller, 2016; OECD, 2024). Rather, we examined whether exclusion of an item would lead to a meaningful increase of reliability. Thus, in some cases, we decided to keep items with relatively low discrimination in the test to preserve broad coverage of the construct climate literacy.

The initial scaling with 176 items showed a reliability of $EAP_{Rel} = 0.89$ and item infit values between 0.84 and 1.32. The discrimination of the items was $-0.15 < r_{it} < 0.54$. After exclusion of 21 items with unacceptable infit or discrimination, or with irregular ICC, the model with the remaining 155 items showed a reliability of $EAP_{Rel} = 0.90$ and item infit values between 0.84 and 1.20. The discrimination of the items was $0.11 < r_{it} < 0.55$ with 15 items having a discrimination value below $r_{it} = 0.20$. However, we decided to keep these items in the test as they were important from a content perspective. Of the remaining 155 items, 100 were multiple choice (e.g., multiple true/false statements on one stimulus), 42 were single choice (i.e., only one correct answer with three incorrect distractors), and 13 were short response questions. Moreover, 85 items addressed the competence aspect *Dealing with content knowledge*, 14 items addressed *Knowledge generation and evaluation*, 27 items addressed *Information and communication*, and 29 items addressed *Normative evaluation*. An overview of the items is provided in the [Supplementary material](#). We performed all further analyses with these remaining 155 items.

3 Results

We report the results alongside the claims made for the validation argument. Thereby, we include different sources of evidence, namely, the test development process (claim 1), a cognitive pretest with students in the target age group (claim 2), two quantitative pilot studies with students in the target age group (claim 2), a main study with students in the target age group (claims 2, 3, 4, and 5), and a teacher survey (claims 1 and 2).

3.1 Evaluation of claim 1: all aspects of climate literacy are represented; no aspect is over- or underrepresented. All items are relevant for the construct represented

The test items were developed each addressing two features described in the interdisciplinary framework of climate literacy (Stadler et al., 2024), namely content aspects and competence facets.

Thereby, each test item (primarily) represents one content aspect and one competence facet. Figure 4 shows the distribution of items according to the framework categories. Overall, the test items cover all competence facets and content aspects. However, there are some disbalances in the number of items in single cells. In the following, we will outline reasons for the cell disbalances.

First, there is a strong focus on the competence facet *Dealing with content knowledge*. While knowledge alone is not enough to ensure responsible action (Kollmuss and Agyeman, 2002), a fundamental understanding of climate change is essential. For instance, without grasping how carbon dioxide influences global temperatures, individuals may struggle to comprehend the rationale behind political measures like fossil fuel taxes. Moreover, without a proper understanding of the scientific basics of climate change, people get susceptible to misinformation they encounter, for example through social media (Lutzke et al., 2019; Treen et al., 2020). However, research indicates that students have significant knowledge gaps and inadequate alternative conceptions about climate change (e.g., Wildbichler et al., 2024). Therefore, we put an emphasis on content knowledge to ensure that all relevant aspects of climate change (scientific basics of the climate system, causes and impacts of climate change, and action options and barriers) are adequately covered in the test. If only a few items were used, important content areas might be left out.

In contrast, for the other three competence facets, fewer items are needed because they assess domain-general competencies rather than specific content knowledge. For example, an item asking students to interpret a graph on the impacts of the CO₂ tax on gasoline prices (see example item, Figure 3) is assigned to the content aspect of *Action options and barriers*. However, students need little prior knowledge of this topic, as all necessary information is provided in the graph. Similarly, evaluating a study design from a methodological perspective (competence facet *Knowledge generation and evaluation*) or relating statements about traveling to their underlying values and norms (*Normative evaluation*) requires only minimal engagement with specific content. As a result, while a few well-designed items can provide a reliable estimate of students' competence in these facets, assessing content knowledge requires a greater number of items to ensure comprehensive coverage of all relevant topics.

Second, regarding content, we particularly emphasized the aspect of action options and barriers. This content aspect can be described in more detail by referring to four basic conceptions, namely, *mitigation and adaptation strategies*, *individual action in social structures*, *barriers to action*, and *ethical and religious justification for actions* (for details, see Supplementary material and Stadler et al., 2024). To date, both educational interventions and assessment instruments primarily consider climate change from a natural science perspective (Aeschbach et al., 2025; Westphal et al., 2025). Moreover, when knowledge about actions is assessed (Tobler et al., 2012; Roczen et al., 2014), questions typically address single actions, for example, for energy conservation. However, to promote a holistic view on climate change as a socio-scientific issue, greater emphasis on the social sciences and the political perspective (Kranz et al., 2022) as well as on the interrelations of social and individual actions and the barriers to action (e.g., commons dilemma; Aitken et al., 2011) is required. To answer

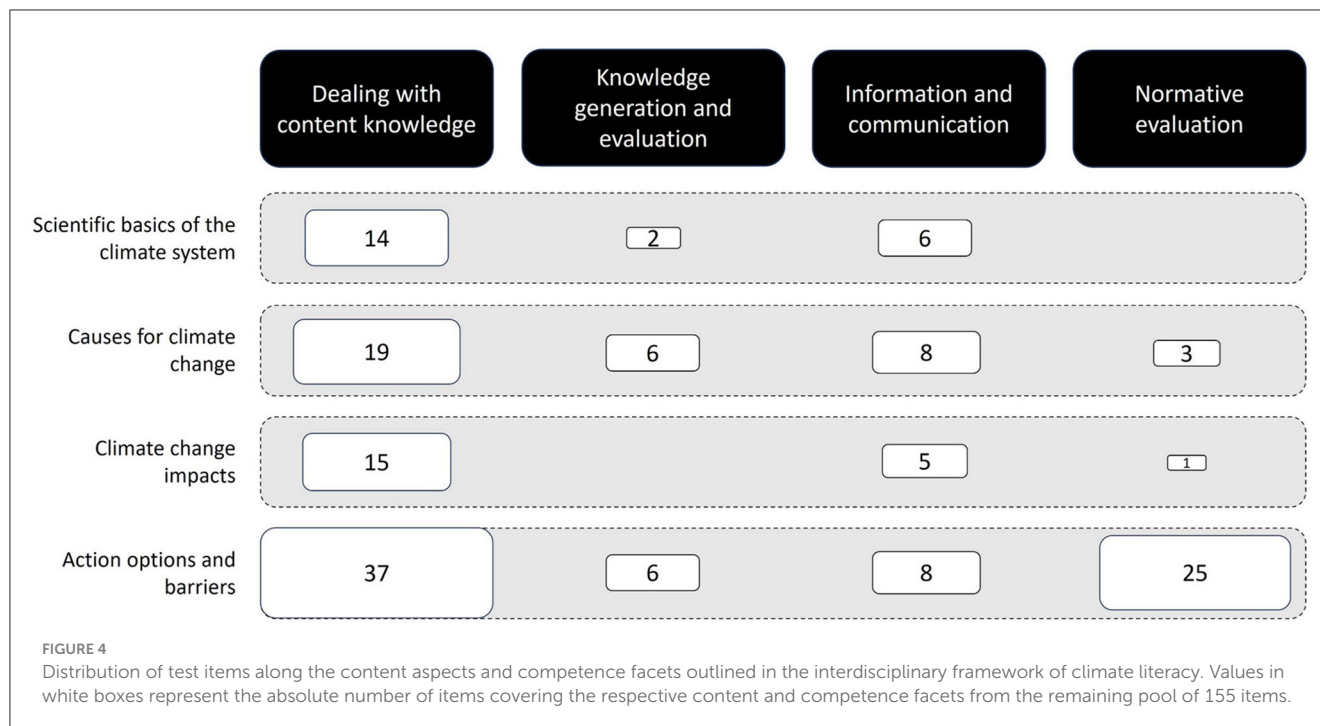
this call, about half of the test items in the climate literacy test presented focus on this social (e.g., political, economic) and ethical perspective.

Third, some combinations of content aspects and competence facets are not meaningful. For example, the competence facet *Normative evaluation* focuses on the application of normative criteria rather than the scientific evaluation of data or facts. As a result, it cannot be applied to just any content area. Evaluating CO₂ levels in the atmosphere from a normative perspective or considering different perspectives on the distinction between weather and climate, for instance, would be inappropriate. However, normative evaluation is particularly relevant when assessing action options, such as understanding the perspectives of people affected by political measures. Therefore, most items addressing this competence facet were concentrated in the *Action options and barriers* content category.

Finally, some cells in the climate literacy framework remained less frequented or even empty not on purpose. Specifically, during the process of item development, pretesting, and pilot testing, we encountered difficulties with the functioning of items covering the competence facet of *Knowledge generation and evaluation*. These items often showed low or even negative discrimination and were either directly excluded or first refined and then still excluded. Low discrimination, that is, low correlations of the item score with the test total score, means that students with high ability levels typically do not solve these items more often or even less often than students with comparably low ability levels. One potential reason for that these items have insufficient discrimination might be that they require to ignore prior knowledge while focusing on generation of new knowledge (Amsel et al., 2008).

The teacher survey revealed mixed perceptions regarding the appropriateness of the test items for assessing climate literacy. Overall, item ratings were relatively high across all teachers ($M = 3.10$, $SD = 0.48$ on a scale from 1 to 4). However, 25 items received mean ratings of 2.5 or below, suggesting that many teachers would not (1) or would rather not (2) use these items to evaluate their students' climate literacy. A closer look at the data showed that teachers tended to rate items outside their subject area as less appropriate. For instance, a physics teacher might consider an item on climate-friendly food choices irrelevant for assessing climate literacy. It is important to note that while the test was designed to assess a school-related competence, it is not strictly aligned to a curriculum, as in Germany the curriculum makes relatively few concrete specifications regarding CCE (see Leve et al., 2025). Instead, the test is grounded in a normative competence model of climate literacy, which may not fully reflect what is currently taught in schools. Therefore, while the teacher ratings might back the use of most of the items, low appropriateness ratings alone do not warrant the removal of the items from the assessment.

To summarize, the test items cover all content areas and competence facets of the interdisciplinary climate literacy framework with a particular focus on the content area of *Action options and barriers*. A teacher sample with a background in various disciplines evaluates the majority of the items to be appropriate for assessing climate literacy in 9th-grade students.



3.2 Evaluation of claim 2: test difficulty is suited for German students in 9th and 10th grade. The test is sensitive to differences in proficiency levels typical for the target group

On average, the dichotomous items were solved by 37.7% ($SD = 18.4$) of the students with the easiest item solved by 79% and the most difficult item solved by only 4% of our sample. Thus, the test difficulty can be regarded as appropriate for the ability level of the sample with neither ceiling nor floor effects. For a closer evaluation of item difficulties across the test components, we examined Wright Maps of the item difficulties and person abilities, separated for the four competence facets *Dealing with content knowledge*, *Knowledge generation and evaluation*, *Information and communication*, and *Normative evaluation* (Figure 5). Wright Maps show item difficulties and person abilities on the same scale and therefore allow for examination whether the difficulties of the test items sufficiently cover the range of person abilities in a sample. A person with a particular ability level (i.e., a logit score) will solve items in the same difficulty with a probability of 0.5, items below that level with a higher probability, and items above that level with a lower probability. The mean person ability is fixed to 0 on the logit scale with abilities typically ranging from -3 to $+3$. Thus, the mean and range of item difficulties can be evaluated in terms of fit to the assessed sample.

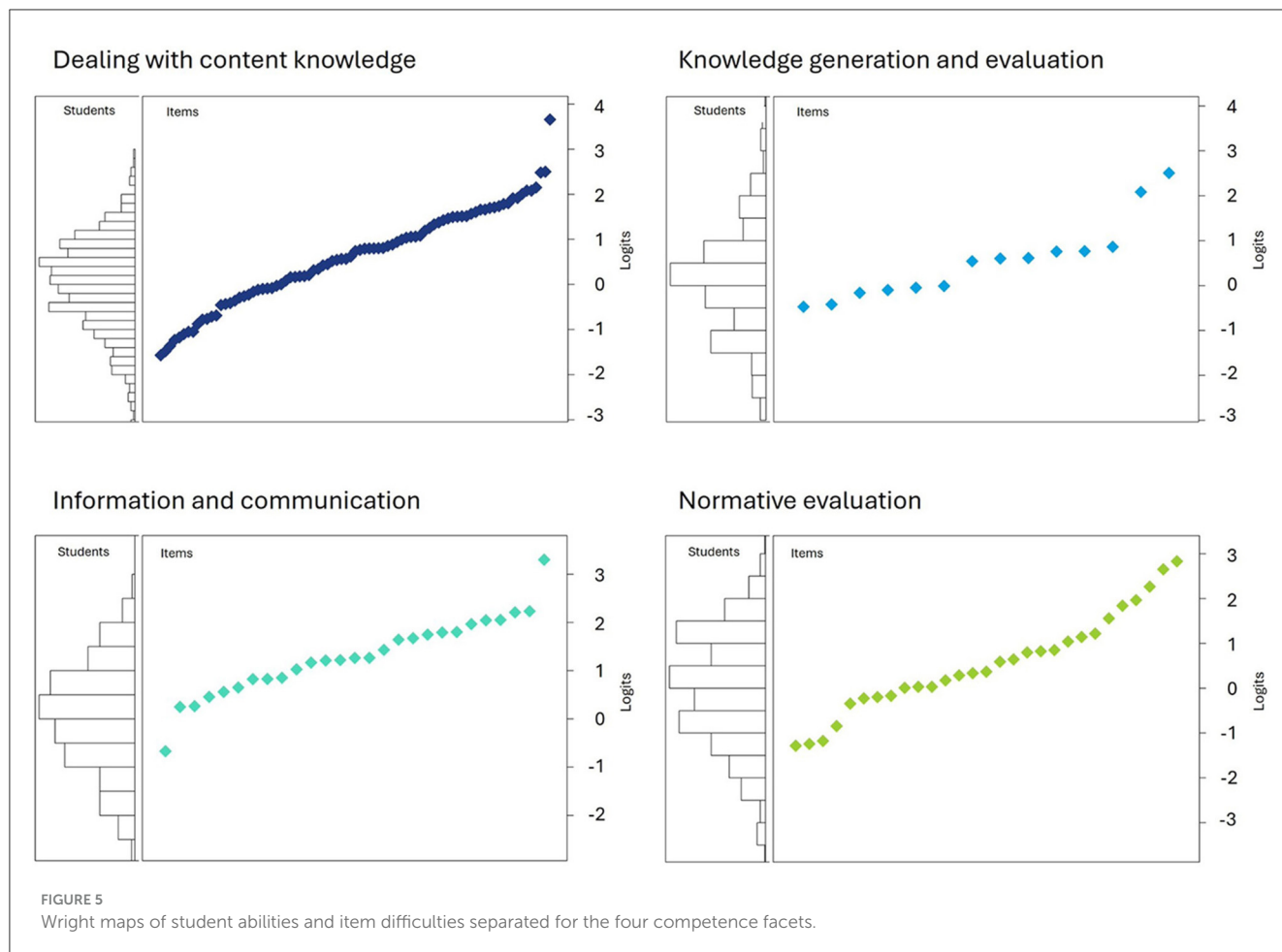
The Wright Maps show that for the competence facets *Dealing with content knowledge* and *Normative evaluation*, the items cover nearly the full range of person abilities in our sample of 9th-graders. The items of the remaining two competence facets were relatively difficult with mean item difficulties of 0.54 logits (range: -0.47 ; 2.46) for *Knowledge generation and evaluation* and 1.29

logits (range: -0.65 ; 3.27) for *Information and communication*, respectively. A four-way ANOVA showed significant differences between the item difficulties of the four competence facets, $F_{(3,151)} = 3.63$, $p = 0.014$. *Post-hoc* tests with Bonferroni adjusted p -values showed that the competence facet *Information and communication* was significantly more difficult than *Dealing with content knowledge* ($p = 0.014$, $d = 0.68$) and *Normative evaluation* ($p = 0.033$, $d = 0.79$) but not different from *Knowledge generation and evaluation* ($p = 0.12$). Differences between the other competence aspects were not statistically significant.

Additionally, we intended the four content aspects to be covered by items of diverse difficulties. Figure 6 shows a closer examination of only the competence facet *Dealing with content knowledge* items. For three of the four content areas, namely the *Scientific basics of the climate system*, *Causes for climate change*, and *Action options and barriers*, the items cover the full range of abilities in our sample with an equal number of items of low, medium, and high difficulty. Only for the content area *Climate change impacts*, there is a lack of relatively easy items.

To assess test fairness, we analyzed differential item functioning (DIF) by calibrating the instrument separately for students from different school types. As shown in Figure 7, the item difficulty order was largely consistent across school types. While the overall test was expectedly easier for students in upper-track schools, individual item difficulties were highly correlated ($r = 0.90$), supporting the validity of score comparisons between school types. Further DIF analyses using the Mantel-Haenszel method with the package *diffR* (Magis et al., 2010) did not reveal any statistically significant DIF items, indicating that no test items exhibited systematic bias across school types.

To complement our analyses of item difficulties, we investigated the teacher survey data. On a five-point scale from *very easy* (1) to *very difficult* (5), the mean difficulty ratings ranged



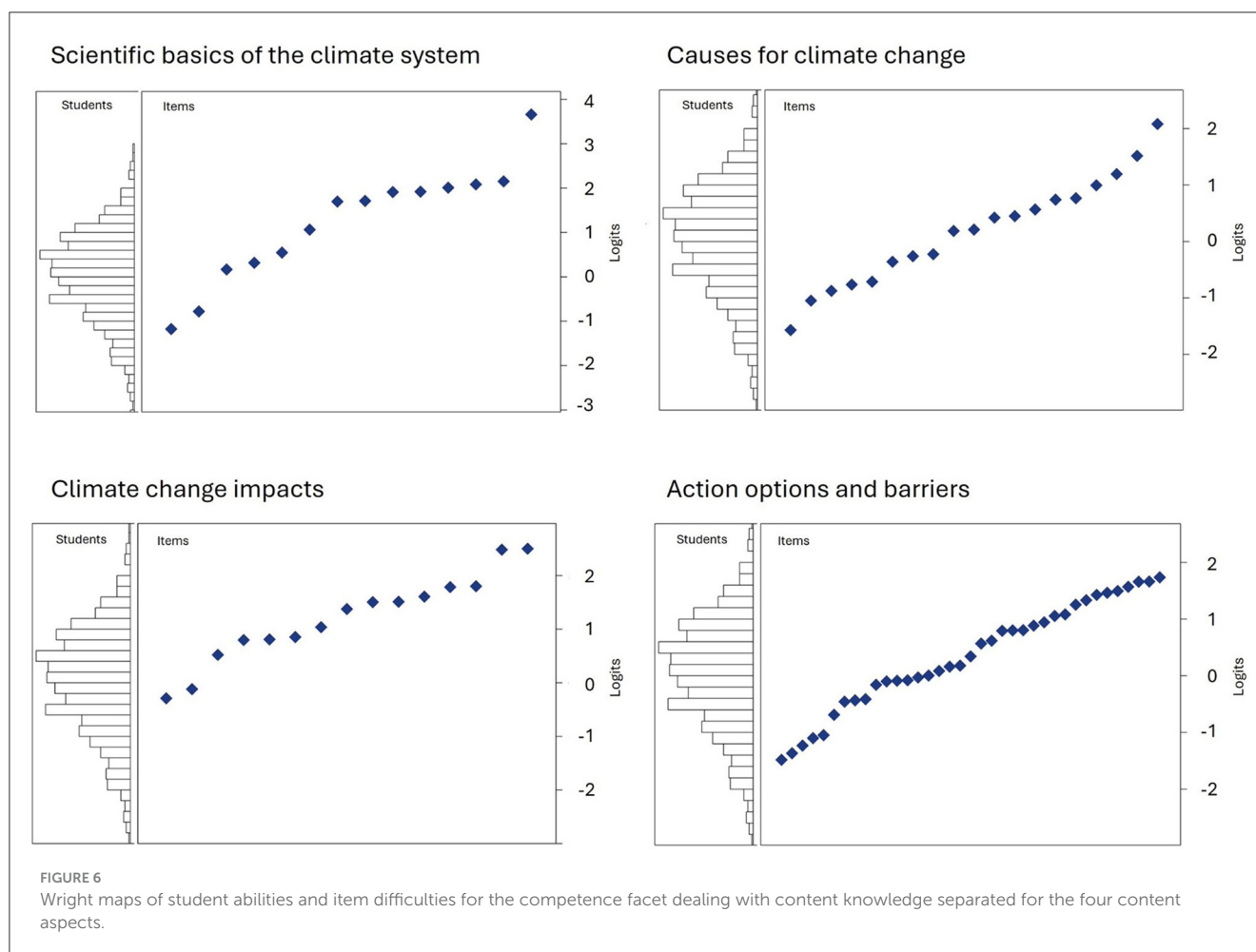
from 1.20 to 4.75. Thus, the teacher sample confirmed the wide range of difficulties covered by the test items. The mean and median ratings were 3.04 and 3.20, respectively, which is slightly above the midpoint of the scale. When looking at the difficulty estimations for the four competence facets, the teacher ratings resemble the findings from the student data with the competence facets *Knowledge generation and evaluation* and *Information and communication* showing slightly higher mean difficulty ratings ($M_{KG} = 3.40$ and $M_{IC} = 3.18$) than *Dealing with content knowledge* and *Normative evaluation* ($M_{CK} = 3.00$ and $M_{NE} = 2.84$).

While the qualitative teacher ratings cannot be directly compared to the empirical difficulty estimates obtained in our student sample, we compared the relative distribution of the items. As our teacher sample consisted exclusively of educators from upper-track schools, we computed empirical difficulties separately for upper-track students. Looking at the item distribution across the respective midpoints of the teacher rating scale and the Rasch scale, it becomes clear that the teachers' expectations of what students in the respective age group should be able to solve aligns with the empirical difficulties. While in the Rasch scaling of the student data, 37% of items showed difficulty estimates below the theoretical midpoint of 0 logits, experts rated 35% of the items below their scale-midpoint of 3.

3.3 Evaluation of claim 3: the assessment instrument represents the four-dimensional structure of climate literacy with four separable competence facets as outlined in the climate literacy framework

To test whether the empirical data support the assumed four-dimensional competence structure, we estimated two models and compared them using relative fit indices AIC and sample-size adjusted BIC. The one-dimensional model (Figure 8A) assumes that all items measure a single latent ability, namely, climate literacy. In contrast, the four-dimensional model assigns each item to one of four competence facets of climate literacy, meaning that the test score reflects students' abilities in four distinct dimensions (Figure 8B).

As shown in Table 4, both AIC and adjusted BIC were lower for the four-dimensional model than for the one-dimensional model. Additionally, a likelihood ratio test confirmed that the four-dimensional model provided a significantly better fit to the data [$\chi^2(6) = 49.28$, $p < 0.001$]. Thus, the empirical data support the assumed four-dimensional structure of climate literacy, consisting of the four competence facets as outlined in the interdisciplinary framework by Stadler et al. (2024). The EAP reliabilities for all



four dimensions were sufficiently high ($0.72 < \text{EAP}_{\text{Rel}} < 0.88$; see Table 4). Although the latent correlations between the dimensions were relatively high (all $r > 0.79$; see Table 5), previous research on the similar construct of scientific literacy suggests that even constructs with latent correlations of $r > 0.90$ between dimensions can still be considered as multidimensional (Baumert et al., 2009; Kampa and Köller, 2016).

However, a four-dimensional model might naturally fit better than a one-dimensional model simply because of its higher complexity, which reduces residual error (Bollen, 1989). To address this concern, we estimated a random four-dimensional model, in which the 155 test items were randomly assigned to four groups of the size of our competence facets (see Kampa and Köller, 2016, for a similar approach). If the random four-dimensional model did not fit the data better than the one-dimensional model, but the theoretically derived four-dimensional model did, this would suggest that the better fit is not just due to the higher number of parameters but rather reflects a meaningful structure. However, the random four-dimensional model failed to converge within a feasible number of iterations (max. 500). Although this prevents a direct comparison of fit indices, the non-convergence itself may indicate that not just any complex model improves fit. Instead, our findings suggest that the theoretically assumed four-dimensional

structure of climate literacy is meaningful and supported by the data.

3.4 Evaluation of claim 4: the competence measured is distinct from generic cognitive abilities and can be shaped through school education

Table 6 shows correlations between the climate literacy test scores separately for the four competence facets and relevant cognitive, affective, and behavioral variables. As expected, all competence facets were significantly related to general cognitive abilities and reading ability as assessed by standardized tests. While these generic skills are highly relevant for solving competence items that go beyond the mere reproduction of factual knowledge (Baumert et al., 2009), these only medium-sized correlations of $0.29 < r < 0.46$ indicate that the test scores capture specific competencies related to dealing with climate-change related problems that go beyond generic skills.

Moreover, to evaluate whether the assessed competencies can be affected by schooling, we compared students attending different school types. The mean person ability for upper track secondary

school students and for medium/lower track school students are displayed in Table 7. For better readability, we standardized ability scores according to the PISA scale (OECD, 2024) with a sample mean of 500 and a standard deviation of 100. As expected, students at higher track schools outperformed those at middle/lower track schools in all four competence facets. This finding aligns with results for German students in scientific literacy as retrieved from national and international assessments (Schiepe-Tiska et al., 2019; Stanat et al., 2021). While in national assessments in Germany, typically students at upper track schools show more homogeneous performances, as reflected by smaller standard deviations (Stanat et al., 2021), in the present study this is only the case for the competence facet of *Dealing with content knowledge*. One possible explanation for this finding is the absence of detailed curricular guidelines for teaching the multifaceted construct of climate literacy (Leve et al., 2025). As a result, while the student population at upper-track schools is generally more homogeneous than those in middle- or lower-track schools, their climate literacy, as measured by the presented test instrument, may vary more significantly based on the specific emphasis of their individual school and teachers.

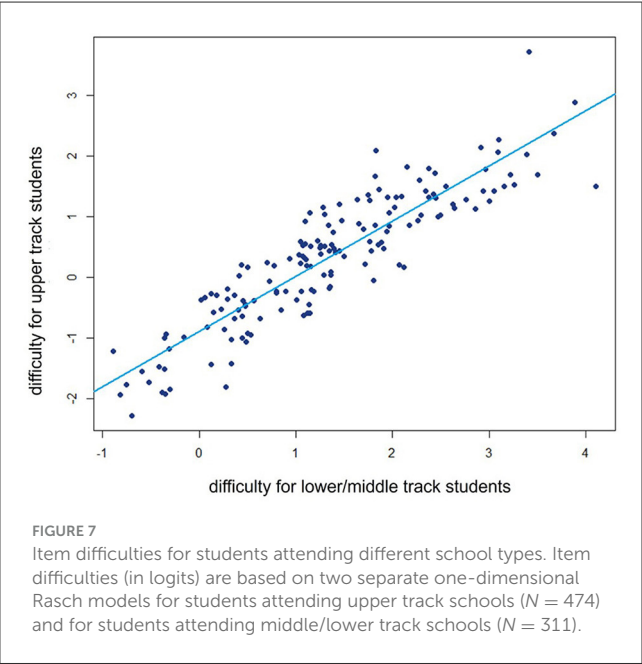


TABLE 4 Model comparison and quality information.

Model	Dimensions			Model quality			
	Name	Items	EAP-Rel	LogLik	Par	AIC	aBIC
1	Climate literacy (general)	155	0.900	−27,128	155	54,566	54,796
2	Dealing with content knowledge	85	0.879	−27,103	161	54,528	54,767
	Knowledge generation and evaluation	14	0.728				
	Information and communication	27	0.787				
	Normative evaluation	29	0.837				

Par, Number of parameters estimated. Lower AIC and aBIC values indicate better model fit and are printed in bold font.

Finally, as students at higher track schools typically demonstrate superior general cognitive abilities and reading abilities (Baumert et al., 2006), we were interested in the additional explanatory power of school type beyond generic skills.¹ A hierarchical linear regression showed that general cognitive abilities alone explained 18% of the variance in the climate literacy overall score ($R^2 = 0.18$). Adding school type significantly improved the model, $\Delta R^2 = 0.15$, $p < 0.001$, resulting in a final model that explained 33% of the variance. Thus, whether students attend an upper-track or middle- or lower-track school substantially predicts their scores, suggesting that the competencies measured in this climate literacy test can be fostered through CCE in schools.

3.5 Evaluation of claim 5: climate literacy assessed with this instrument is relevant for teenagers’ active participation in social climate change discourse

To assess whether the cognitive component of climate literacy measured by this instrument is meaningfully linked to students’ ability to actively and responsibly engage in social discourse on climate change, as defined by the United States Global Change Research Program (2009), we examined its relationship with the other key components of climate literacy, namely the affective-motivational and behavioral component. We found significant medium-sized correlations between all four facets of climate literacy and climate change attitudes (see Table 6), with the largest correlation for *Dealing with content knowledge* ($r = 0.45$).

Regarding behavior, we examined relations to students’ participation in public climate protest which can be regarded as an individual high impact behavior (Haugestad et al., 2021). We found significant small correlations of participation in the Fridays for Future protests and the climate literacy score for three of the four competence facets. Specifically, when looking at the standardized WLE scores, students who reported that they have participated at least once in the Fridays for Future protests scored significantly higher than the sample mean of 500 points

1 As we obtained reading ability data from only a small subsample, we computed hierarchical regressions only with general cognitive abilities to preserve statistical power.

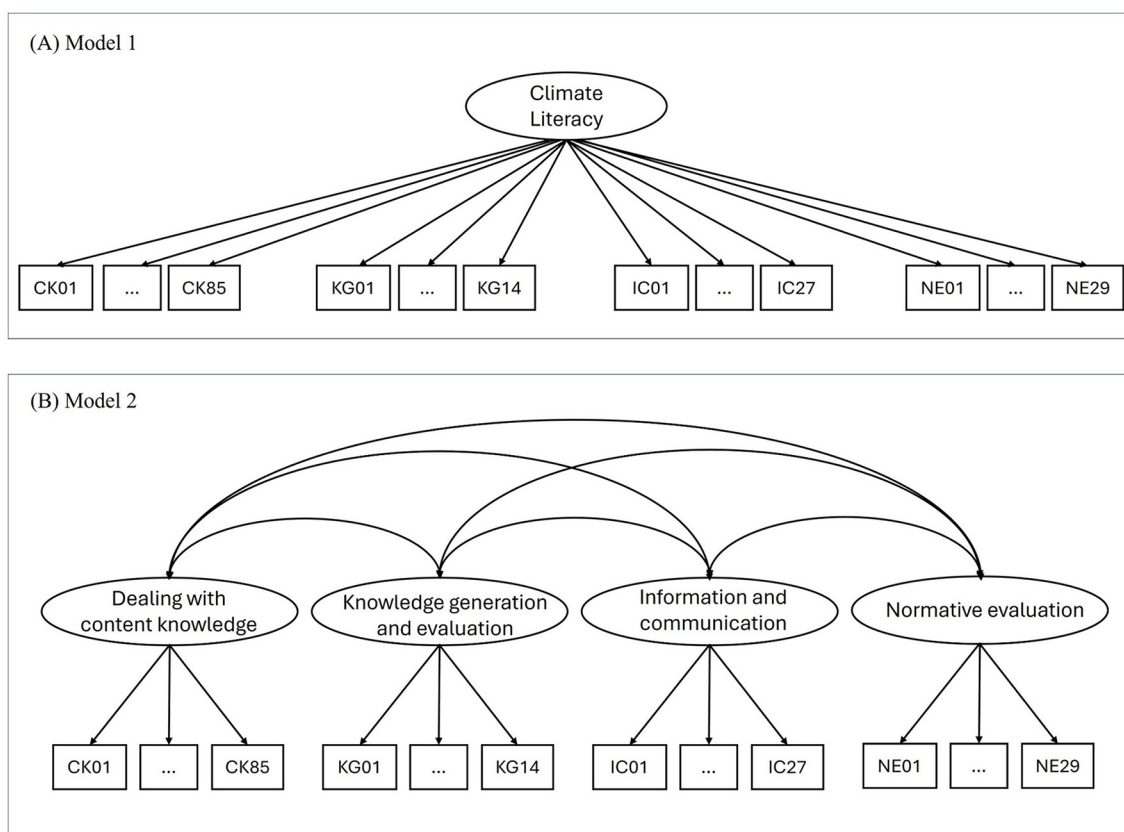


FIGURE 8

Schematic representation of the one-dimensional model (Model 1, A) and the four-dimensional model (Model 2, B). CK, Dealing with content knowledge (85 items); KG, Knowledge generation and evaluation (14 items); IC, Information and communication (27 items); NE, Normative evaluation (29 items).

TABLE 5 Latent correlations between competence facets.

Competence facet		1	2	3	4
1	Dealing with content knowledge	-			
2	Generating and evaluating knowledge	0.814	-		
3	Information and communication	0.858	0.794	-	
4	Normative evaluation	0.910	0.836	0.861	-

with 545, 536, and 540 points for *Dealing with content knowledge* [$t_{(97.59)} = 4.27, p < 0.001$], *Information and communication* [$t_{(93.00)} = 3.19, p = 0.002$], and *Normative evaluation* [$t_{(94.23)} = 3.61, p < 0.001$], respectively.

4 Discussion

One key objective of climate change education (CCE) is to promote climate literacy in students, that is, equipping them with the knowledge and skills to engage with climate-related challenges. To determine whether CCE measures meet this aim,

it is essential to assess students' climate literacy. Therefore, the present study introduced a comprehensive climate literacy test for secondary school students. In developing this assessment instrument, we aimed at addressing three key features that current assessment instruments of CCE outcomes are lacking, namely interdisciplinarity, relevance to everyday life, and the opportunity to provide differentiated insights into students' competencies. In the following, we will discuss in what way we have achieved these three aims. Therefore, we will summarize the validation argument and evaluate the theoretical and empirical evidence supporting the test's intended use. Additionally, we identify gaps in the validation argument that future research should address and outline potential practical applications of the instrument.

4.1 Final evaluation of the validity argument

To validate the intended test score interpretation of providing detailed insights into the climate literacy of secondary school students, we formulated five validity claims referring to construct representation, test difficulty, test fairness, internal structure, and relations to external variables. First, we ensured adequate

construct representation by grounding the test development in an interdisciplinary framework of climate literacy (Stadler et al., 2024). The test items were developed by educational experts from nine disciplines, including the natural and social sciences, as well as technology and the humanities. Thus, the newly developed test overcomes a shortcoming of current instruments that typically focus on the physical aspects of climate change (Lubej et al., 2025). However, despite this interdisciplinary scope, we did not include all school subjects that might be relevant in promoting climate literacy. Other disciplines that could make an interesting contribution to expanding the test would be, for example, mathematics, chemistry or artistic subjects such as music or fine arts.

The competence test developed in this study is the first comprehensive measure of adolescents' cognitive aspects of climate literacy. By focusing on climate literacy (see also scientific literacy, Sjöström and Eilks, 2018) instead of climate change knowledge, the test captures the competencies adolescents need to navigate climate change in their daily lives, for example, when interpreting information encountered on social media, or by taking diverse perspectives when discussing mitigation actions. When looking at the competence facets covered, we see a particularly relevant contribution in developing test items covering the competence facet of *Normative evaluation*. It captures whether students can apply normative principles, without conflating this ability with their personal support for those norms. Thereby, the instrument focuses on the cognitive part of climate literacy without blending knowledge with attitudes or beliefs, a frequent challenge when addressing controversial topics like climate change mitigation actions (Kahan, 2015).

Second, cognitive labs including think-aloud assessment as well as two quantitative pilot studies with students in the target age group allowed us to modify test difficulty to suit students at the end of 9th grade in all secondary school types. Moreover, DIF analyses ensured that the test was fair for students attending different school types. The broad range of item difficulties covered by the developed instrument prevents effects of restricted variance often observed in existing instruments (e.g., Stevenson et al., 2014; Kuthe et al., 2019). Thereby, it fulfills an essential prerequisite for capturing development, for example, through the course of an educational intervention. However, further research is required to examine to what extent the developed test is sensitive to instruction (Naumann et al., 2019).

While in total, the test items covered the whole range of students' abilities, the competence aspect of *Information and communication* seemed especially difficult for the students. This finding aligns with previous research indicating that secondary school students have difficulties with interpreting information from graphical representations (e.g., Nieminen et al., 2010; Scheid et al., 2017). The complexity of the natural climate system (e.g., tipping points) as well as a variety of social interdependencies in mitigation actions require higher level interpretation and integration of information extracted from diagrams. However, further development of the test could also address lower complexity levels of information extraction to allow for differentiation among the lower ability levels.

Third, the four competence facets proposed in the interdisciplinary framework of climate literacy can also be separated empirically. Thus, the problems presented in the test

items require demonstration of four competence facets, namely (1) *Dealing with content knowledge*, (2) *Knowledge generation and evaluation*, (3) *Information and communication*, and (4) *Normative evaluation*. However, high latent correlations ($0.79 < r < 0.91$)² indicate a close relationship between the four competence facets. Future studies might therefore investigate the required competencies needed for solving the climate literacy test items in more detail. For example, it would be interesting to examine within-item multidimensionality (Hartig and Höhler, 2009), that is, whether specific items cover more than one competence facet.

Fourth, as expected, test scores across all competence facets varied by school type. The largest difference emerged in *Dealing with content knowledge*, where students at upper-track schools outperformed those at middle- or lower-track schools by one standard deviation. The finding that school type explains a substantial amount of variance in the test scores beyond general cognitive abilities supports the idea that the climate literacy test measures a school-related competence. Establishing this link to school-related variables is crucial for validating the test's use in monitoring formal CCE measures, as it suggests that climate literacy, as assessed by this instrument, is shaped by school experiences and thus can be influenced through educational interventions. Based on this assumption, researchers might use the climate literacy test to investigate whether differences in curricula (e.g., Leve et al., 2025) translate into meaningful variations in student outcomes.

Finally, climate literacy scores showed significant correlations with climate change-related attitude (medium-sized correlations) and behavior (small correlations), highlighting the assessed construct's relevance to students' daily lives. Interestingly, these correlations were more pronounced for the competence facets *Dealing with content knowledge* and *Normative evaluation* than for *Knowledge generation and evaluation* and *Information and communication*. This pattern not only reinforces the four-dimensional structure proposed in the interdisciplinary climate literacy framework (see also Claim 3 on structural validity) but also aligns with previous findings in environmental knowledge research, where different types of knowledge exhibit varying degrees of association with behavior (Roczen et al., 2014; Braun and Dierkes, 2019).

However, the correlations between the climate literacy test scores and behavior were somewhat lower than expected. One reason for this result might be that we explicitly focused on public sphere behavior while previous research on pro-environmental behavior focused on private consumption behavior, such as energy conservation and choices of transportation (e.g., Roczen et al., 2014). Thus, further research is needed to uncover the relationship between students' cognitive climate literacy and their engagement in collective climate action. Moreover, we assessed public sphere behavior with only one item (i.e., participation in climate strikes). Further studies might address this limitation by assessing public sphere behavior with multiple items, covering not only multiple potential actions (e.g., signing petitions,

² It should be noted that, due to correction for measurement error, latent correlations are typically higher than the manifest correlations (Hartig and Höhler, 2009).

TABLE 6 Correlations of the four climate literacy competence facets with related constructs.

Variable		<i>M</i>	<i>SD</i>	1	2	3	4
1	Dealing with content knowledge ^a	0.00	1.00				
2	Knowledge generation and evaluation ^a	0.08	1.24	0.52**			
3	Information and communication ^a	0.06	1.14	0.63**	0.45**		
4	Normative evaluation ^a	0.01	1.28	0.71**	0.48**	0.58**	
5	Reading ability ^b	13.11	6.28	0.45**	0.36**	0.35**	0.44**
6	General cognitive abilities ^c	9.49	3.29	0.41**	0.29**	0.35**	0.36**
7	Attitude toward climate change ^d	0.01	0.99	0.45**	0.30**	0.34**	0.39**
8	Participation in FFF demonstrations ^e	0.10	0.30	0.14**	0.07	0.12**	0.13**

**Indicates $p < 0.01$. ^aClimate literacy competence facet scores on logit scale, four-dimensional Rasch model. ^b47 items (max. 47 points), the reading test was only presented to a subsample of 218 students. ^c15 items (max. 15 points). ^dLogit scale, one-dimensional Rasch model. ^eDichotomous item: 0 = have never participated, 1 = have participated at least once.

TABLE 7 Climate literacy scores for students attending different school types.

Climate literacy competence facet	Standardized score <i>M</i> (<i>SD</i>)		Difference		
	Upper track students (<i>n</i> = 474)	Middle/lower track students (<i>n</i> = 311)	Δ	<i>t</i>	<i>p</i>
Climate literacy overall	540 (84)	439 (92)	101	15.5	<0.001
Dealing with content knowledge	538 (82)	442 (97)	96	14.5	<0.001
Knowledge generation and evaluation	524 (97)	463 (93)	61	8.76	<0.001
Information and communication	529 (96)	457 (90)	72	10.7	<0.001
Normative evaluation	534 (91)	449 (92)	85	12.8	<0.001

Scores are scaled with a sample mean of 500 and sample standard deviation of 100 points.

getting involved in a youth organization) but also their quantity and intensity.

Nevertheless, previous investigations with measures of general environmental knowledge and relatively narrow measures of climate change knowledge, such as the simple recognition of scientific facts about greenhouse gases, show only weak and non-significant correlations with attitudinal measures (e.g., Stoutenborough and Vedlitz, 2014) and public sphere actions such as climate activism (e.g., Barbosa et al., 2021; Knupfer et al., 2023). Thus, the competence test presented here, which covers four competence facets that go beyond factual knowledge provides a more comprehensive view of students' climate literacy that can also predict relevant external criteria like climate change-related attitudes and meaningful behaviors. However, as these findings are correlational, we cannot infer causality. It is possible that higher climate literacy encourages participation in demonstrations, that participation in such activities enhances climate literacy by informal education, or that a shared underlying factor drives both. For example, interest in climate-related topics might lead to both activism and information-seeking (see also Taube et al., 2021; Baierl et al., 2022). Future research, such as longitudinal studies or experimental interventions, is needed to disentangle these relationships.

To sum up, the developed test instrument adopts an interdisciplinary approach to climate literacy, aligning with calls for greater integration of the social sciences in CCE (Shwom et al., 2017; Kranz et al., 2022; Aeschbach et al., 2025; Westphal

et al., 2025). Furthermore, by distinguishing between four specific competence facets, the test scores gained by this instrument offer nuanced insights into the broad construct of climate literacy, enabling the identification and targeting of areas where educational efforts are most needed.

4.2 Intended use

The test instrument aims at measuring climate literacy broadly by incorporating items from nine different school subjects. The large number of items made a booklet design necessary, where each student completed only about 40% of the total test items. This approach is commonly used in large-scale assessments (e.g., OECD, 2024) to gather population-level data, such as national trends. However, it is not suitable for diagnosing individual students, as no single student completes items covering all aspects of the construct to be assessed. Thus, only the entire sample collectively covers the breadth of the construct. Similarly, the results of this climate literacy assessment should not be used to determine an individual student's proficiency. Instead, they provide insights at a more aggregated level, such as comparisons between school types or across federal states. The findings from this study, while not representative, offer an initial look at patterns in climate literacy. For example, differences observed between school types in other domains, such as science (Stanat et al., 2021), appear to be just as pronounced—if not more so—in climate literacy.

This comprehensive assessment can serve as a valuable tool for evaluating the effectiveness of novel and promising educational approaches. For instance, in (quasi)-experimental studies, it can help assess whether holistic initiatives such as the *Whole School Approach* (e.g., Holst et al., 2024) and programs that complement formal school education, such as *Public Climate School* (Keller et al., 2024) successfully facilitate climate literacy in students.

Moreover, the competence assessment provides nuanced insights into the climate literacy construct and its four competence facets. Thus, the four separate test scores enable the identification of specific deficits and the development of targeted interventions. For example, the present study's results suggest that irrespective of the school type, students struggle when it comes to extracting and interpreting information (e.g., about climate change causes and mitigation strategies) from complex representations. This finding might provide an indication for the development of educational interventions targeted specifically to the competence facet of *Information and communication*.

4.3 Outlook and conclusion

The test instrument and accompanying materials are available for research purposes at FDZ Bildung (DOI: <https://doi.org/10.7477/1231-424-63>). Looking ahead, further development of the instrument could include the establishment of competence levels, making test results more interpretable for educators and policymakers (Hartig and Klieme, 2006; Neumann et al., 2010). Additionally, the development of a short scale or an adaptive testing format (e.g., Frey and Seitz, 2009) would enhance usability in settings with limited time or resources, without compromising the test's diagnostic value. Including more low-difficulty items could improve the instrument's sensitivity at the lower end of the ability spectrum. Beyond refinement, the test offers promising practical applications: it enables large-scale monitoring of educational interventions, supports curriculum development, and can help identify correlations and influences between climate literacy and other constructs. In sum, the presented operationalization of an interdisciplinary climate literacy framework into tangible test items offers both a theoretical advancement and a practical resource for advancing CCE in schools.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ministerium für Kultus, Jugend und Sport Baden-Württemberg. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

MM: Data curation, Methodology, Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft. MSt: Data curation, Methodology, Conceptualization, Investigation, Writing – review & editing. JK: Methodology, Supervision, Conceptualization, Project administration, Writing – review & editing. MSc: Methodology, Conceptualization, Formal analysis, Writing – review & editing. RA: Conceptualization, Writing – review & editing. UB: Conceptualization, Writing – review & editing. FB: Conceptualization, Writing – review & editing. AC: Conceptualization, Writing – review & editing. A-MG: Conceptualization, Writing – review & editing. CH: Conceptualization, Writing – review & editing. SS: Conceptualization, Writing – review & editing. JS: Conceptualization, Writing – review & editing. WR: Methodology, Supervision, Conceptualization, Formal analysis, Project administration, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. Generative AI was used for language editing (grammar and style) of the manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2025.1637522/full#supplementary-material>

References

- Ackerman, T. A. (1992). A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective national council on measurement in education. *J. Educ. Meas.* 29, 67–91. doi: 10.1111/j.1745-3984.1992.tb00368.x
- Adamina, M., Hertig, P., Probst, M., Reinfried, S., Stucki, P., and Vogel, J. (2018). *Klimabildung in allen Zyklen der Volksschule und in der Sekundarstufe II*. Available online at: <https://www.globeswiss.ch/files/Downloads/1567/Download/Summary%20CCESO%201.pdf> (Accessed March 13, 2022).
- AERA, APA, and NCME (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association. Available online at: <https://www.testingstandards.net/open-access-files.html> (Accessed April 19, 2025).
- Aeschbach, V. M.-J., Schwichow, M., and Rieß, W. (2025). Effectiveness of climate change education—a meta-analysis. *Front. Educ.* 10:1563816. doi: 10.3389/feduc.2025.1563816
- Aitken, C., Chapman, R., and McClure, J. (2011). Climate change, powerlessness and the commons dilemma: assessing New Zealanders' preparedness to act. *Glob. Environ. Change* 21, 752–760. doi: 10.1016/j.gloenvcha.2011.01.002
- Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E., et al. (2008). A dual-process account of the development of scientific reasoning: the nature and development of metacognitive intercession skills. *Cogn. Dev.* 23, 452–471. doi: 10.1016/j.cogdev.2008.09.002
- Azevedo, J., and Marques, M. (2017). Climate literacy: a systematic review and model integration. *Int. J. Glob. Warming* 12, 414–430. doi: 10.1504/IJGW.2017.084789
- Baierl, T.-M., Kaiser, F. G., and Bogner, F. X. (2022). The supportive role of environmental attitude for learning about environmental issues. *J. Environ. Psychol.* 81:101799. doi: 10.1016/j.jenvp.2022.101799
- Barbosa, R. d. A., Randler, C., and Robaina, J. V. L. (2021). Values and environmental knowledge of student participants of climate strikes: a comparative perspective between Brazil and Germany. *Sustainability* 13:8010. doi: 10.3390/su13148010
- Baumert, J., Lüdtke, O., Trautwein, U., and Brunner, M. (2009). Large-scale student assessment studies measure the results of processes of knowledge acquisition: evidence in support of the distinction between intelligence and student achievement. *Educ. Res. Rev.* 4, 165–176. doi: 10.1016/j.edurev.2009.04.002
- Baumert, J., Stanat, P., and Watermann, R. (2006). "Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus," in *Herkunftsbezogene Disparitäten im Bildungswesen*, eds. J. Baumert, P. Stanat, and R. Watermann (Wiesbaden: VS Verlag für Sozialwissenschaften), 95–188.
- Bedford, D. (2016). Does climate literacy matter? A case study of U.S. students' level of concern about anthropogenic global warming. *J. Geogr.* 115, 187–197. doi: 10.1080/00221341.2015.1105851
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). "Taxonomy of educational objectives: the classification of educational goals," in *Handbook 1: Cognitive Domain*, ed. B. S. Bloom (New York: David McKay), 1103–1133.
- Bofferding, L., and Kloser, M. (2015). Middle and high school students' conceptions of climate change mitigation and adaptation strategies. *Environ. Educ. Res.* 21, 275–294. doi: 10.1080/13504622.2014.888401
- Bogner, F. X. (2018). Environmental values 2-MEV and appreciation of nature. *Sustainability* 10:350. doi: 10.3390/su10020350
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Hoboken, NJ: John Wiley and Sons.
- Bond, T. G., and Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Boon, H. J. (2016). Pre-service teachers and climate change: a stalemate? *Aust. J. Teach. Educ.* 41, 39–63. doi: 10.14221/ajte.2016v41n4.3
- Boone, W. (2016). Rasch analysis for instrument development: why, when, and how? *CBE Life Sci. Educ.* 15:rm4. doi: 10.1187/cbe.16-04-0148
- Braun, T., and Dierkes, P. (2019). Evaluating three dimensions of environmental knowledge and their impact on behaviour. *Res. Sci. Educ.* 49, 1347–1365. doi: 10.1007/s11165-017-9658-7
- Bybee, R. W. (2012). "Scientific literacy in environmental and health education," in *Science / Environment / Health: Towards a Renewed Pedagogy for Science Education*, eds. A. Zeyer, and R. Kyburz-Grabner (Netherlands: Springer), 49–67.
- Cantell, H., Tolppanen, S., Aarnio-Linnanvuori, E., and Lehtonen, A. (2019). Bicycle model on climate change education: presenting and evaluating a model. *Environ. Educ. Res.* 25, 717–731. doi: 10.1080/13504622.2019.1570487
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences, 2nd Edn*. New York: Lawrence Erlbaum Associates.
- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., et al. (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environ. Res. Lett.* 8:24024. doi: 10.1088/1748-9326/8/2/024024
- Deisenrieder, V., Kubisch, S., Keller, L., and Stötter, J. (2020). Bridging the action gap by democratizing climate change education-The case of k.i.d.Z.21 in the context of Fridays for future. *Sustainability* 12:1748. doi: 10.3390/su12051748
- DeWaters, J. E., Andersen, C., Calderwood, A., and Powers, S. E. (2014). Improving climate literacy with project-based modules rich in educational rigor and relevance. *J. Geosci. Educ.* 62, 469–484. doi: 10.5408/13-056.1
- Eames, C. W., Monroe, M. C., White, P. J., and Ardoin, N. M. (2024). *Engaging environmental education through PISA: leveraging curriculum as a political process*. *Aust. J. Environ. Educ.* 40, 1–11. doi: 10.1017/ae.2024.40
- Edelsbrunner, P. A., and Dablander, F. (2019). The psychometric modeling of scientific reasoning: a review and recommendations for future avenues. *Educ. Psychol. Rev.* 31, 1–34. doi: 10.1007/s10648-018-9455-5
- Embretson, S. E., and Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah: Erlbaum.
- Flora, J. A., Saphir, M., Lappé, M., Roser-Renouf, C., Maibach, E. W., and Leiserowitz, A. A. (2014). Evaluation of a national high school entertainment education program: the alliance for climate education. *Clim. Change* 127, 419–434. doi: 10.1007/s10584-014-1274-1
- Frey, A., and Bernhardt, R. (2012). On the importance of using balanced booklet designs in PISA. *Psychol. Test Assess. Model.* 54, 397–417.
- Frey, A., Hartig, J., and Rupp, A. A. (2009). An NCME instructional module on booklet designs in large-scale assessments of student achievement: theory and practice. *Educ. Meas.* 28, 39–53. doi: 10.1111/j.1745-3992.2009.00154.x
- Frey, A., and Seitz, N. N. (2009). Multidimensional adaptive testing in educational and psychological measurement: current state and future challenges. *Stud. Educ. Eval.* 35, 89–94. doi: 10.1016/j.stueduc.2009.10.007
- García-Vinuesa, A., Meira-Carrea, P. Á., and Caride-Gómez, J. A. (2024). Climate change education and secondary school students: a meta-synthesis (1993-2017). *Educ. Knowl. Soc.* 25:e31358. doi: 10.14201/eks.31358
- Gräsel, C., Bormann, I., Schütte, K., Trempler, K., and Fischbach, R. (2013). Outlook on research in education for sustainable development1. *Policy Futures Educ.* 11, 115–127. doi: 10.2304/pfie.2013.11.2.115
- Hargis, K. (2024). "Practicing climate action in a K-12 school using a whole institution approach," in *Sustainable Development Goals Series*, eds. A. E. Wals, B. Bjønness, A. Sinnes, and I. Eikeland (New York: Springer), 247–259.
- Hartig, J., and Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Stud. Educ. Eval.* 35, 57–63. doi: 10.1016/j.stueduc.2009.10.002
- Hartig, J., and Klieme, E. (2006). "Kompetenz und Kompetenzdiagnostik [Competence and competence diagnosis]," in *Leistung und Leistungsdiagnostik [Performance and assessment of performance]*, ed. K. Schweizer (Berlin: SpringerVerlag), 127–143.
- Haugestad, C. A. P., Skaug, A. D., Kunst, J. R., and Power, S. A. (2021). Why do youth participate in climate activism? A mixed-methods investigation of the #FridaysForFuture climate protests. *J. Environ. Psychol.* 76:101647. doi: 10.1016/j.jenvp.2021.101647
- Higuchi, M. I. G., Paz, D. T., Roazzi, A., and Souza, B. C. D. (2018). Knowledge and beliefs about climate change and the role of the amazonian forest among university and high school students. *Ecopsychology* 10, 106–116. doi: 10.1089/eco.2017.0050
- Holst, J., Grund, J., and Brock, A. (2024). Whole institution approach: measurable and highly effective in empowering learners and educators for sustainability. *Sustain. Sci.* 19, 1359–1376. doi: 10.1007/s11625-024-01506-5
- IPCC (2022). *Climate Change 2022 - Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press.
- Jorgenson, S. N., Stephens, J. C., and White, B. (2019). Environmental education in transition: a critical review of recent research on climate change and energy education. *J. Environ. Educ.* 50, 160–171. doi: 10.1080/00958964.2019.1604478
- Kahan, D. M. (2015). Climate-science communication and the measurement problem. *Polit. Psychol.* 36, 1–43. doi: 10.1111/pops.12244
- Kaiser, F. G. (2021). Climate change mitigation within the Campbell paradigm: doing the right thing for a reason and against all odds. *Curr. Opin. Behav. Sci.* 42, 70–75. doi: 10.1016/j.cobeha.2021.03.024
- Kaiser, F. G., Oerke, B., and Bogner, F. X. (2007). Behavior-based environmental attitude: development of an instrument for adolescents. *J. Environ. Psychol.* 27, 242–251. doi: 10.1016/j.jenvp.2007.06.004
- Kampa, N., and Köller, O. (2016). German national proficiency scales in biology: internal structure, relations to general cognitive abilities and verbal skills. *Sci. Educ.* 100, 903–922. doi: 10.1002/sce.21227

- Kane, M. (2013). The argument-based approach to validation. *Sch. Psych. Rev.* 42, 448–457. doi: 10.1080/02796015.2013.12087465
- Kaplan, R. M., and Saccuzzo, D. P. (2001). *Psychological Testing: Principles, Applications, and Issues, 5th Edn.* Belmont, CA: Wadsworth/Thomson Learning.
- Keller, J., Eichinger, M., Bechtoldt, M., Liu, S., Neuber, M., Peter, F., et al. (2024). Evaluating the public climate school, a multi-component school-based program to promote climate awareness and action in students: a cluster-controlled pilot study. *J. Clim. Change Health* 15:100286. doi: 10.1016/j.joclim.2023.100286
- Klieme, E., Hartig, J., and Rauch, D. (2008). “The concept of competence in educational context,” in *Assessment of Competencies in Educational Contexts*, eds. J. Hartig, E. Klieme, and D. Leutner (Göttingen: Hogrefe & Huber Publishers), 3–22.
- Knupfer, H., Neureiter, A., and Matthes, J. (2023). From social media diet to public riot? Engagement with “greenfluencers” and young social media users’ environmental activism. *Comput. Human Behav.* 139:107527. doi: 10.1016/j.chb.2022.107527
- Kollmuss, A., and Agyeman, J. (2002). Mind the Gap: why do people act environmentally and what are the barriers to pro-environmental behavior? *Environ. Educ. Res.* 8, 239–260. doi: 10.1080/13504620220145401
- Kranz, J., Breitenmoser, P., Laherto, A., Krug, A., Schwichow, M., and Tasquier, G. (2025). Science Education for collective agency in the climate crisis: a social identity approach. *Res. Sci. Educ.* 55, 1149–1168. doi: 10.1007/s1165-025-10282-w
- Kranz, J., Schwichow, M., Breitenmoser, P., and Niebert, K. (2022). The (Un)political Perspective on climate change in education—a systematic review. *Sustainability* 14:74194. doi: 10.3390/su14074194
- Krathwohl, D. R. (2002). A revision of bloom’s taxonomy: an overview. *Theory Pract.* 41, 212–218. doi: 10.1207/s15430421tip4104_2
- Kuthe, A., Keller, L., Körfgen, A., Stötter, H., Oberrauch, A., and Höferl, K. M. (2019). How many young generations are there?—A typology of teenagers’ climate change awareness in Germany and Austria. *J. Environ. Educ.* 50, 172–182. doi: 10.1080/00958964.2019.1598927
- Landis, J. R., and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 33, 363–374. doi: 10.2307/2529786
- Leve, A.-K., Garrecht, C., and Harms, U. (2025). Curriculare Einbindung der Klimabildung – wie ist der Stand und wo soll es hingehen? *Z. f. Didakt. Nat.* 31:3. doi: 10.1007/s40573-025-00178-7
- Leve, A. K., Michel, H., and Harms, U. (2023). Implementing climate literacy in schools — what to teach our teachers? *Clim. Change* 176:134. doi: 10.1007/s10584-023-03607-z
- Libarkin, J. C., Gold, A. U., Harris, S. E., McNeal, K. S., and Bowles, R. P. (2018). A new, valid measure of climate change understanding: associations with risk perception. *Clim. Change* 150, 403–416. doi: 10.1007/s10584-018-2279-y
- Loy, L. S. (2018). *Communicating Climate Change - How Proximating Climate Change and Global Identity Predict Engagement. [Dissertation]*. University of Hohenheim, Stuttgart, Germany.
- Lubej, M., Petraš, Ž., and Kirbiš, A. (2025). Measuring climate knowledge: a systematic review of quantitative studies. *iScience* 28:111888. doi: 10.1016/j.isci.2025.111888
- Lutzke, L., Drummond, C., Slovic, P., and Árvai, J. (2019). Priming critical thinking: simple interventions limit the influence of fake news about climate change on Facebook. *Glob. Environ. Change* 58:101964. doi: 10.1016/j.gloenvcha.2019.101964
- Magis, D., Béland, S., Tuerlinckx, F., and de Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behav. Res. Methods* 42, 847–862. doi: 10.3758/BRM.42.3.847
- McNeal, K. S., Walker, S. L., and Rutherford, D. (2014). Assessment of 6- to 20-Grade educators’ climate knowledge and perceptions: results from the climate stewardship survey. *J. Geosci. Educ.* 62, 645–654. doi: 10.5408/13-098.1
- Michaelides, M. P., and Ivanova, M. (2022). Response time as an indicator of test-taking effort in PISA: country and item-type differences. *Psychol. Test Assess. Model.* 64, 304–338.
- Mochizuki, Y., and Bryan, A. (2015). Climate change education in the context of education for sustainable development: rationale and principles. *J. Educ. Sustain. Dev.* 9, 4–26. doi: 10.1177/0973408215569109
- Monroe, M. C., Plate, R. R., Oxart, A., Bowers, A., and Chaves, W. A. (2019). Identifying effective climate change education strategies: a systematic review of the research. *Environ. Educ. Res.* 25, 791–812. doi: 10.1080/13504622.2017.1360842
- Naumann, A., Musow, S., Aichele, C., Hochweber, J., and Hartig, J. (2019). Instructional sensitivity of tests and items. *Z. für Erzieh.* 22, 181–202. doi: 10.1007/s11618-018-0832-0
- Neumann, K., Fischer, H. E., and Kauertz, A. (2010). From PISA to educational standards: the impact of large-scale assessments on science education in Germany. *Int. J. Sci. Math. Educ.* 8, 545–563. doi: 10.1007/s10763-010-9206-7
- Nielsen, K. S., Clayton, S., Stern, P. C., Dietz, T., Capstick, S., and Whitmarsh, L. (2021). How psychology can help limit climate change. *Am. Psychol.* 76, 130–144. doi: 10.1037/amp0000624
- Nieminen, P., Savinainen, A., and Viiri, J. (2010). Force concept inventory-based multiple-choice test for investigating students’ representational consistency. *Phys. Rev. Spec. Top. – Phys. Educ. Res.* 6, 1–12. doi: 10.1103/PhysRevSTPER.6.020109
- OECD (2017). *PISA 2015 Assessment and Analytical Framework, Revised Edition*. Paris: OECD Publishing.
- OECD (2024). *PISA 2022 Technical Report*. Paris: OECD Publishing.
- Otto, I. M., Donges, J. F., Cremades, R., Bhowmik, A., Hewitt, R. J., Lucht, W., et al. (2020). Social tipping dynamics for stabilizing earth’s climate by 2050. *Proc. Natl. Acad. Sci. USA* 117, 2354–2365. doi: 10.1073/pnas.1900577117
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Danmarks Paedagogiske Institut.
- Redman, A., and Wiek, A. (2021). Competencies for advancing transformations towards sustainability. *Front. Educ.* 6:785163. doi: 10.3389/feduc.2021.785163
- Robitzsch, A., Kiefer, T., and Wu, M. (2024). *TAM: Test Analysis Modules*. R package version 4.2-21, Available online at: <https://CRAN.R-project.org/package=TAM> (Accessed August 18, 2024).
- Roczen, N., Kaiser, F. G., Bogner, F. X., and Wilson, M. (2014). A competence model for environmental education. *Environ. Behav.* 46, 972–992. doi: 10.1177/0013916513492416
- Schauss, M., and Sprenger, S. (2021). Students’ conceptions of uncertainties in the context of climate change. *Int. Res. Geograp. Environ. Educ.* 30, 332–347. doi: 10.1080/10382046.2020.1852782
- Scheid, J., Müller, A., Hettmannsperger, R., and Kuhn, J. (2017). Erhebung von repräsentationaler Kohärenzfähigkeit von Schülerinnen und Schülern im Themenbereich Strahlenoptik. *Z. für Didakt. Nat.* 23, 181–203. doi: 10.1007/s40573-017-0065-4
- Schiepe-Tiska, A., Rönnebeck, S., and Neumann, K. (2019). “Naturwissenschaftliche Kompetenz in PISA 2018,” in *PISA 2018: Grundbildung im internationalen Vergleich*, eds. K. Reiss, M. Weis, E. Klime, and O. Köller (Münster: Waxmann), 211–240.
- Schlagmüller, M., Ennemoser, M., and Usanova, I. (2022). “Diagnostics of reading speed, reading comprehension, and reading accuracy using the LGVT 5–12+,” in *Language Development in Diverse Settings*, eds. H. Brandt, M. Krause, and I. Usanova (Wiesbaden: Springer VS), 99–132. doi: 10.1007/978-3-658-35650-7_4
- Schneider, W., Schlagmüller, M., and Ennemoser, M. (2017). *LGVT 5-12: Lesegeschwindigkeits- und -verständnistest für die Klassen 5–12*. Göttingen: Hogrefe.
- Schubatzky, T., Wackermann, R., Haagen-Schützenhöfer, C., and Wöhlke, C. (2024). How Well do German a-level students understand the scientific underpinnings of climate change? *Sustainability* 16:7264. doi: 10.3390/su16177264
- Segade-Vázquez, M., García-Vinuesa, A., Rodríguez-Groba, A., and Conde, J. J. (2025). Characterising educational research on climate change in the climate emergency era (2017–2024). *Rev. Esp. Pedagog.* 83, 179–198. doi: 10.22550/2174-0909.4110
- Shepardson, D. P., Roychoudhury, A., Hirsch, A., Niyogi, D., and Top, S. M. (2014). When the atmosphere warms it rains and ice melts: Seventh grade students’ conceptions of a climate system. *Environ. Educ. Res.* 20, 333–353. doi: 10.1080/13504622.2013.803037
- Shwom, R., Isenhour, C., Jordan, R. C., McCright, A. M., and Robinson, J. M. (2017). Integrating the social sciences to enhance climate literacy. *Front. Ecol. Environ.* 15, 377–384. doi: 10.1002/fee.1519
- Siegmund Space and Education (2021). *Analyse zur Verankerung von Klimabildung in den formalen Lehrvorgaben für Schulen und Bildungseinrichtungen in Deutschland*. Available online at: <https://www.siegmund-se.de/klimabildung> (Accessed May 27, 2022).
- Sjöström, J., and Eilks, I. (2018). “Reconsidering different visions of scientific literacy and science education based on the concept of Bildung,” in *Cognition, Metacognition, and Culture in STEM Education. Innovations in Science Education and Technology*, eds. Y. J. Dori, Z. R. Mevarech, and D. R. Baker (New York: Springer), 65–88.
- Stadler, M., Martin, M., Schuler, S., Stemmann, J., Rieß, W., and Küsting, J. (2024). “Entwicklung eines Kompetenzstrukturmodells für climate literacy,” in *Bildung für nachhaltige Entwicklung, Edition ZfE*, eds. H. Kminek, V. Holz, M. Singer-Brodowski, H. Ertl, T. S. Idel, and C. Wulf (Wiesbaden: Springer VS. © Springer Nature), 3–36. Available online at: <https://link.springer.com/book/10.1007/978-3-658-46596-4>
- Stanat, P., Schipolowski, S., Schneider, R., Sachse, K. A., Weirich, S., and Henschel, S. (2021). *IQB-Bildungstrend 2021*. Available online at: <https://www.iqb.hu-berlin.de/bt/BT2018/Bericht/> (Accessed May 27, 2022).
- Stevenson, K. T., Peterson, M. N., Bondell, H. D., Moore, S. E., and Carrier, S. J. (2014). Overcoming skepticism with education: interacting influences of worldview and climate change knowledge on perceived climate change risk among adolescents. *Clim. Change* 126, 293–304. doi: 10.1007/s10584-014-1228-7

- Stevenson, R. B., Nicholls, J., and Whitehouse, H. (2017). What is climate change education? *Curric. Perspect.* 37, 67–71. doi: 10.1007/s41297-017-0015-9
- Stoutenborough, J. W., and Vedlitz, A. (2014). The effect of perceived and assessed knowledge of climate change on public policy concerns: an empirical comparison. *Environ. Sci. Policy* 37, 23–33. doi: 10.1016/j.envsci.2013.08.002
- Taube, O., Ranney, M. A., Henn, L., and Kaiser, F. G. (2021). Increasing people's acceptance of anthropogenic climate change with scientific facts: is mechanistic information more effective for environmentalists? *J. Environ. Psychol.* 73:101549. doi: 10.1016/j.jenvp.2021.101549
- Tobler, C., Visschers, V. H. M., and Siegrist, M. (2012). Consumers' knowledge about climate change. *Clim. Change* 114, 189–209. doi: 10.1007/s10584-011-0393-1
- Treen, K. M. d. I., Williams, H. T. P., and O'Neill, S. J. (2020). Online misinformation about climate change. *Wiley Interdiscip. Rev. Clim. Change* 11, 1–20. doi: 10.1002/wcc.665
- UNESCO (2020). *Education for Sustainable Development: A Roadmap*. Paris: UNESCO.
- UNESCO and UNFCCC (2016). *Action for Climate Empowerment: Guidelines for Accelerating Solutions Through Education, Training and Public Awareness*. Paris: UNESCO.
- United States Global Change Research Program (2009). *Climate Literacy: The Essential Principles of Climate Science*. Available online at: www.globalchange.gov (Accessed May 27, 2022).
- Visser, L., Rothe, J., Schulte-körne, G., and Hasselhorn, M. (2022). Evaluation of an online version of the CFT 20-R in third and fourth grade children. *Children* 9:40512. doi: 10.3390/children9040512
- Weinert, F. E. (2001). "Concept of competence: a conceptual clarification", in *Defining and Selecting Key Competencies*, eds. D. S. Rychen and L. H. Salganik (Göttingen: Hogrefe & Huber), 45–65.
- Weiß, R. H. (2006). *Grundintelligenztestskala 2 Revision (CFT20-R)*. Göttingen: Hogrefe.
- Westphal, A., Kranz, J., Schulze, A., Schulz, H., Becker, P., and Wulff, P. (2025). *Climate change education: bibliometric analysis of the Status-Quo and future research paths*. *J. Environ. Educ.* 56, 307–326. doi: 10.1080/00958964.2025.2475299
- Wildbichler, S., Haagen-Schützenhöfer, C., and Schubatzky, T. (2024). *Students' ideas about the scientific underpinnings of climate change: a systematic review of the literature*. *Stud. Sci. Educ.* 61, 117–169. doi: 10.1080/03057267.2024.2395206