



Key sustainability competences of university students: A systematic literature review of assessment methodologies, methods, and instruments

Eleni Sinakou^{*} , Maria Daskolia

Environmental Education Lab, Department of Educational Studies, National and Kapodistrian University of Athens, Greece

ARTICLE INFO

Keywords:

Key sustainability competences
Higher education
University students
Assessment

ABSTRACT

In recent years, empirical research has focused on the assessment of Key Sustainability Competences (KSCs) of university students. This study explored the methodologies, methods, and instruments used to measure the KSCs of university students. We identified the most efficient methodologies, methods, and instruments by investigating their strengths and weaknesses. We conducted a systematic literature review to analyze the methodologies, methods, and instruments used for assessing the KSCs of university students. A total of 37 reports, published from 2013 to 2026, were reviewed. The study employed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology to ensure rigor and transparency. Psychometric properties were assessed according to COSMIN criteria, and the certainty of the evidence was evaluated using the GRADE approach. The review revealed a broad range of methodologies, methods, and instruments applied for the assessment of KSCs of university students. Quantitative methodologies are favored in comparison to qualitative or mixed-method ones. Scaled instruments are often used for summative assessment purposes. Triangulation of methods is recommended for robust assessments. However, researchers seldom report the psychometric properties of the instruments employed; when such information was provided, it was typically supported by high-certainty evidence. The findings have implications for research in KSCs assessment.

1. Introduction

Growing global challenges such as climate change, limited resources, social inequality, and financial crises highlight an immediate need for change in how societies operate and how humans interact with the environment. Education for Sustainable Development (ESD) has emerged as a transformative pedagogical approach aiming to equip students with the knowledge, skills, values, and attitudes to address challenging issues towards a more sustainable future (UNESCO, 2018). HEIs have been charged with developing Key Sustainability Competences (KSCs) essential for promoting sustainability, thereby empowering students to contribute to a more sustainable future in both personal and professional contexts (Tójar-Hurtado et al., 2024).

HESD research has begun exploring how to enhance university ESD teaching. Robust assessment of teaching and learning in HESD is necessary for evaluating the effectiveness of academic programs (Redman et al., 2021; Valderrama-Hernández et al., 2020, 2025). HESD literature has recently addressed the assessment of students' learning outcomes, i.e., their KSCs (nde et al., 2020, 2025). While experts

largely agree on the KSCs that higher education should foster (e.g., Hammer & Lewis, 2023; [authors, under review]), empirical data are notably limited regarding the validity of this conceptual framework among university students (Hammer & Lewis, 2023). Therefore, Annelin and Boström (2022) call for the development of assessment instruments that can evaluate the KSCs of university students. Academic teachers could examine students' KSCs at the beginning and the end of a course or academic program to evaluate their progress. They could then rearrange ESD teaching accordingly.

However, there is some ambiguity regarding how KSCs can be assessed (Redman et al., 2021). While there has been some progress in creating assessment instruments to measure the KSCs of HEI students (e.g., see the reviews of Annelin & Boström, 2022; Redman et al., 2021), further work is required to make them efficient. This expanding body of research necessitates a systematic literature review that explores current assessment methodologies, methods, and instruments to guide researchers' choices.

This study distinguishes among research methodologies, methods, and instruments used to assess KSCs. Ontological and epistemological

^{*} Corresponding author.

E-mail addresses: esinakou@eds.uoa.gr, sinakou.eleni@hotmail.com (E. Sinakou).

<https://doi.org/10.1016/j.stueduc.2026.101609>

Received 13 October 2025; Received in revised form 13 April 2026; Accepted 17 April 2026

Available online 5 May 2026

0191-491X/© 2026 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

assumptions inform the selection of quantitative, qualitative, or mixed-method methodologies (Coe et al., 2025; Teddlie & Tashakkori, 1998). As the backbone of a study, methodology determines what evidence should be collected to assess students' KSCs. Learning outcomes, instruction and assessment tasks should be aligned to support valid inferences about learning (Biggs, 1996). This is particularly important for KSCs since the KSCs are conceptualized as integrated and context-dependent and thus, the selected methods and instruments should elicit appropriate evidence. In this regard, for instance, within a qualitative methodology, interviews might be selected as the method for generating data. To implement this method, an instrument such as a semi-structured interview guide is used to collect responses. Similarly, within a quantitative methodology, surveys may be used as the method, requiring a questionnaire as the instrument for gathering standardized data. These distinctions clarify the relationship between the overarching research philosophy (methodology), the procedures used to collect and analyze data (methods), and the specific tools used in practice (instruments). Previous HESD assessment research has identified conflation of the KSCs constructs and the instruments used to measure them. Redman et al. (2021) suggesting typologies that distinguish self-perception, observation-based, and test-based methodologies to emphasize the need for theoretically and methodologically justified methodologies, methods, and instruments.

We conducted a Systematic Literature Review (SLR) to investigate recent literature on assessment methodologies, methods, and instruments that measure the KSCs of university undergraduates. SLRs synthesize study findings transparently, systematically, and comprehensively (Bearman et al., 2012). This approach enables us to include not only well-known studies but also a broader array of literature through a structured procedure (Bearman et al., 2012). Furthermore, SLRs help reveal potential gaps in previous literature that may inform future research and policies.

HESD scholars agree that measuring HEI students' KSCs presents a highly intricate challenge (Redman et al., 2021; Sossé et al., 2021; Valderrama-Hernández et al., 2020, 2025). We explore the methodologies, methods, and instruments used by HESD researchers to assess university students' KSCs. We specifically focus on university students, as different KSCs are expected for students at the undergraduate or postgraduate level (Arene, 2022; Wiek et al., 2016). We examine their strengths, limitations and their psychometric properties to eventually decide on the most effective methodologies, methods, and instruments for our target population. In this regard, we set the following research questions for our review:

1. What are the methodologies, methods, and instruments used in HESD research to assess KSCs of university students?
2. What are the reported strengths and weaknesses of the various KSCs assessment methodologies, methods, and instruments?
3. To what extent do HESD studies report on psychometric properties (i.e., validity and reliability) of the KSCs instruments they employ?

In what follows, we first put forward our theoretical background on (a) what KSCs entail for university students, and (b) what assessment means in HEI context. We then set forth details on the methodology adopted for the SLR, report our findings, and finally, discuss possible implications for HESD teaching, research, and policymaking.

2. Theoretical background

2.1. Key sustainability competences in higher education

Key Sustainability Competences (KSCs) are necessary capabilities that help individuals enjoy balanced lives while striving for a better tomorrow for all (Gora et al., 2019). University students are anticipated to cultivate competencies for critically examining modern human values, behaviors, and policies, and to modify them to successfully deal

with sustainable development (SD) challenges (Dlouhá et al., 2019).

KSCs for university students, as defined by de Kraker et al. (2007), are the results of synergies among knowledge, skills, and attitudes that empower students to facilitate sustainability transitions. Glasser (2016) further defines KSCs as a complete, non-overlapping array of competencies vital for changing societal systems to improve collective well-being. Brundiers et al. (2021) define a KSC as: "a distinctive and multifunctional competence, which is composed of several sustainability competences that functionally relate to each other. It facilitates achieving successful performance and a positive outcome that progresses sustainability (given what is known, valued, and aspired at a given moment in time), while working on specific sustainability challenges and opportunities in a range of contexts" (p. 17).

Recent literature in the field of Higher Education for Sustainable Development emphasizes the Sustainability and Problem-Solving framework as a comprehensive set of KSCs of university students. This framework has been endorsed by HESD scholars (Bianchi et al., 2022; Michel et al., 2024) and it is apt for operationalization across multiple research methodologies (i.e., quantitative, qualitative, and mixed-methods) [(authors, under review)]. It consists of nine KSCs that HEIs should equip their students with (e.g., Brundiers et al., 2021; Redman et al., 2021; Wiek et al., 2011, 2016). Key Sustainability Competences encompass:

1. *Systems-thinking competence*: The ability to analyze intricate socio-ecological systems through time and space, accounting for systemic features such as impacts, feedback loops, and inertia.
2. *Anticipatory competence*: The ability to assess multiple possible futures, articulate visions for what lies ahead, assume responsibility for future generations, evaluate the consequences of actions, and manage potential risks or challenges.
3. *Normative competence*: The ability to discern the values and norms that shape individual or group behavior, to critically reflect on them, and to negotiate consensus or compromises amidst opposing interests.
4. *Strategic competence*: The ability to discern the underlying causes of unsustainability and to collectively design and execute novel transformative interventions at both global and local levels that foster sustainability.
5. *Interpersonal competence*: The ability to leverage others' knowledge and experience, recognize and respect their needs and viewpoints, effectively settle conflicts within groups or societies, and facilitate involvement to address SD issues.
6. *Intrapersonal competence*: The ability to self-regulate, motivate, and continually enhance oneself through competencies derived from emotional intelligence and social-emotional learning, with an awareness of one's own emotions, thoughts, and behaviors.
7. *Transdisciplinary competence*: The ability to address SD issues by efficiently applying transdisciplinary knowledge and methodologies within social contexts in a meaningful manner.
8. *Meta-competence or Integrated problem-solving competence*: The ability to recognize the necessity of integrating the aforementioned competencies and applying them into practice to address SD challenges.
9. *Implementation competence*: The ability to act decisively to achieve a focused strategy for addressing SD issues, informed by a well-defined future vision, while monitoring and evaluating outcomes and making necessary adjustments through iterative cycles of planning, execution, and assessment.

2.2. Assessment of key sustainability competences (KSCs) of University students

Establishing transparent assessment methodologies, methods, and instruments for sustainability competencies is essential for educators, scholars, program leaders, curriculum developers, and policymakers to plan, implement, and assess the effectiveness of Education for

Sustainable Development (ESD) pedagogies and educational programs aimed at promoting sustainability learning among students (Annelin & Boström, 2022; Redman et al., 2021). Assessment in education is the systematic collection of evidence through various techniques to observe student behaviour, make evaluations, and ascertain student knowledge, performance, and outcomes (Black & Wiliam, 1998). Researchers in educational science have criticized traditional assessment methods as insufficient for evaluating multidimensional and performance-based competencies (Frey et al., 2009).

In competence-based education, assessment serves two primary functions: formative assessment, which is administered throughout the learning process to offer continuous feedback to both students and teachers; and summative assessment, which takes place towards the end of a course to evaluate learners' outcomes, typically resulting in a grade or score (Black & Wiliam, 1998). Summative assessment instruments typically include tests, essays, portfolios, projects, presentations, or reports that evaluate learning following a given instructional period. Summative assessment facilitates standardization and its application across diverse student cohorts and at different time scales, enabling the aggregation of data at individual, group, program, school, and system levels (Erlinger, 2018). Formative assessment instruments include observations, classroom dialogues, written assignments, as well as self-evaluation and peer-assessment, all aimed at regulating student learning and the instructional process (Erlinger, 2018). Formative assessment directs the learning process, fostering student ownership and facilitating active engagement in self-assessment and peer evaluation (Looney, 2011). A challenge remains in the design and implementation of effective competence assessment instruments, encompassing both formative and summative assessments (Redman et al., 2021).

3. Methodology

3.1. Data collection

We conducted a systematic literature review to examine methodologies, methods, and instruments used in HESD research to measure Key

Sustainability Competences of university students. The PRISMA methodology (Preferred Reporting Items for Systematic Reviews and Meta-Analyses methodology (PRISMA), a widely recognized framework for conducting systematic literature reviews, was applied. Our review followed three stages, according to the PRISMA standards: (a) Identification, (b) Screening, and (c) Eligibility (Fig. 1).

The identification process involved a series of steps to select relevant reports. In this context, reports are papers or electronic documents that provide information on a specific topic (Page et al., 2021). The identification process was conducted in July 2025. To identify reports to include in the review, we searched electronic databases such as Scopus, Web of Science, Education Research Complete (via EBSCO), ScienceDirect, and Google Scholar. In each database, we applied a comprehensive combination of keywords using Boolean operators in the search string: ("higher education" OR "universit*" OR "tertiary education" OR "college") AND ("education for sustainability" OR "education for sustainable development" OR "sustainability education" OR "environmental and sustainability education") AND ("sustainability competenc*" OR "key sustainability competenc*" OR "competenc*") AND ("assessment" OR "evaluation" OR "effectiveness" OR "measurement"). This search yielded an initial pool of 244 reports. After deleting duplicates, we had 213 unique reports for review.

To confirm that the retrieved reports met our inclusion criteria, we screened the titles, abstracts, author-designated keywords, or the full text of each report. In addition to our initial search, we employed a snowballing strategy by examining the references cited in eligible reports (Wohlin et al., 2012). We checked these additional reports following the same procedure. The authors separately undertook the screening process for the selected reports. Any discrepancies were resolved through discussion among the authors.

Aligned with the focus of our study, we established specific inclusion criteria. This review elaborates on methodologies, methods, and instruments used to measure KSCs of university students across all disciplines. We included reports that focus on both one KSC alone and more than one. Furthermore, we excluded reports focusing on the sustainability competences of school students, poststudents, academic staff, or

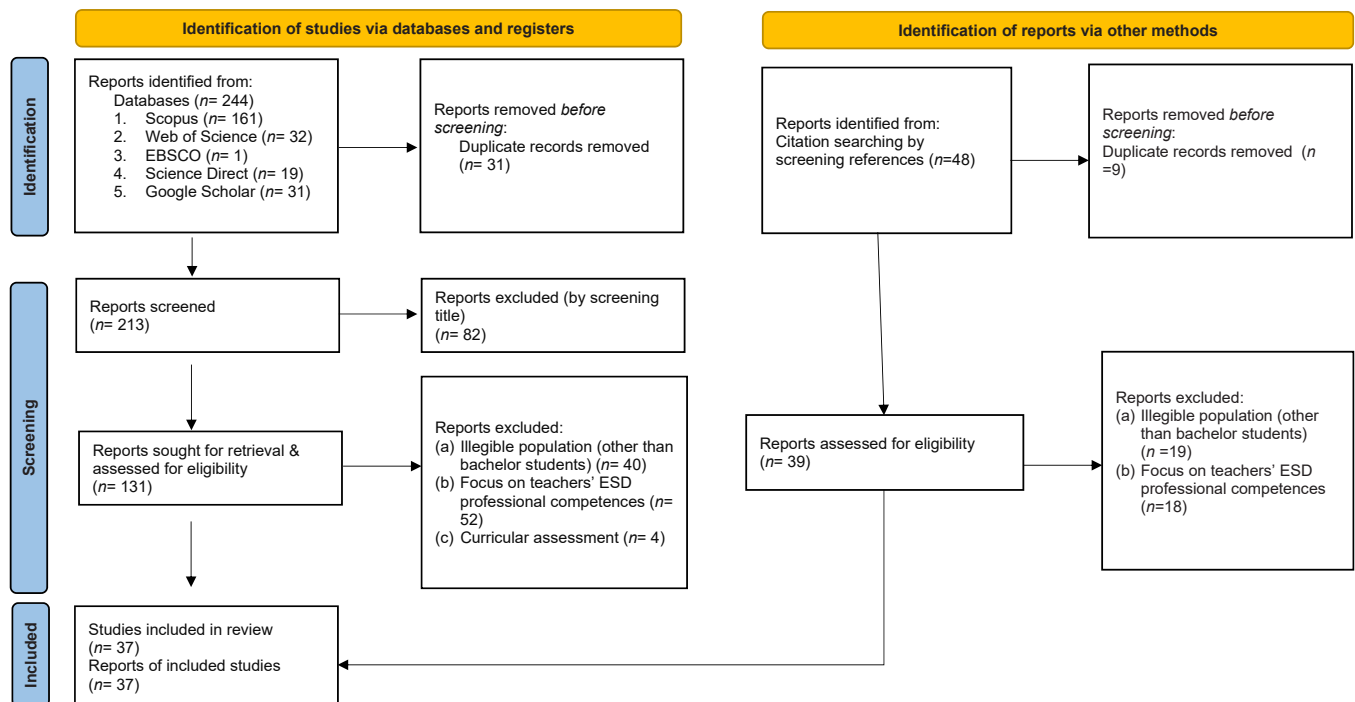


Fig. 1. PRISMA 2020 Flow Diagram for Literature Reviews (Page et al., 2021). From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>.

professionals. Reports on curriculum assessments were also not included in this review. Reports addressing teachers’ professional sustainability competences, that is, their skills to teach ESD, were also excluded. All retrieved reports, regardless of publication year needed to be fully accessible and written in English. Finally, we included 37 reports for analysis in this review.

The reports chosen for this SLR are journal articles ($n = 35$) or book chapters ($n = 2$). The final pool of studies comprises 5 literature reviews and 32 empirical papers. These reports were published between 2013 and 2026 (Fig. 2). Table S1 in the Supplementary Material lists the reports. Then, we used critical appraisal tools to assess the quality of the selected reports. Specifically, we utilized the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018) for studies employing qualitative, quantitative, or mixed methods designs, and the CASP Systematic Review Checklist for systematic literature reviews (Critical Appraisal Skills Programme Systematic Review Checklist, 2023). The retrieved reports demonstrated moderate to high quality and were therefore accepted for analysis. Tables S2 & S3 in the Supplementary Material present the outcomes of the critical appraisal in detail. The screening process for the selected reports was carried out independently by each author, with discrepancies resolved through discussion. Using this method also helps to reduce the risk of bias in the included studies. We found that none of the included systematic reviews employed critical appraisal tools to evaluate the risk of bias of the primary studies they reviewed (Annelin & Boström, 2022; Cebrian & Segalas, & Hernandez, 2019; Redman et al., 2021; Segalàs & Hernández, 2019; Ssossé et al., 2021), diminishing their trustworthiness. Moreover, several reviews did not clearly articulate their research questions (Albareda-Tiana et al., 2024; Galt et al., 2013; Xiao et al., 2024). Without explicit research questions, it is difficult to assess the alignment between aims, methods, and outcomes, which increases the risk of misinterpretation and reduces confidence in the synthesized evidence.

3.2. Analytical approach

The authors have separately extracted the data from the included reports. They cross-checked their extractions and any discrepancies were resolved through discussion among the authors. To address research questions RQ1 and RQ2, we meticulously analyzed their content of the selected reports by utilizing spreadsheets. The data were thematically categorized and then synthesized to extract meaningful interpretations (Thompson, 2022). To answer RQ3, we evaluated the studies by taking three essential steps: a) appraising the methodological quality of the reports, b) analysing the outcomes of each study in relation to the criteria for good measurement properties, and c) verifying the overall strength of the evidence.

Methodological quality appraisal. We evaluated the methodological

quality of each of the included studies by utilising the COSMIN Risk of Bias Checklist (Mokkink et al., 2024). The COSMIN Risk of Bias Checklist assesses dimensions of outcome measure development and validation: reliability (internal consistency, inter/intra-rater reliability, test- retest reliability, measurement error), validity (content validity, criterion validity, construct validity-including structural validity and cross-cultural validity/measurement invariance-, responsiveness). Each psychometric property received a score on a 4-point scale: *very good* (V), *adequate* (A), *doubtful* (D), and *inadequate* (I). The total risk of bias rating for each psychometric property was established by the lowest rating among its items.

Psychometric properties quality appraisal. We assessed each psychometric property of each of the included studies according to the COSMIN criteria for good psychometric properties (Mokkink et al., 2024) and rated them as *sufficient* (+), *indeterminate* (?), or *insufficient* (-).

Grading the quality of evidence. We assessed the quality of evidence for each psychometric property by using the COSMIN-modified Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) methodology (Mokkink et al., 2024). The quality of each psychometric property was categorised as *high*, *moderate*, *low*, or *very low* according to the combined results of the above-described risk of bias assessment and criteria for good measurement properties.

4. Results

4.1. Research methodologies, methods, and instruments employed

The selected reports showcase a wide range of methodologies, methods, and instruments for collecting data on Key Sustainability Competences of university students. Some studies use solely qualitative or quantitative methodologies, while others employ mixed-method approaches. To address RQ1, we analyze the retrieved reports based on their methodological approach (qualitative, quantitative or mixed-

Table 1
Methodologies and methods used.

Methodologies	Methods	Number of Reports
Literature reviews	systematic literature reviews	5
Qualitative approaches	semi-structured interviews, role-playing, written course assignments, open-ended questionnaires, reflective learning journals	6
Quantitative approaches	scaled assessment instruments often with in conjunction with a rubric	17
Mixed-method approaches	course evaluations, vignette survey, self-assessment rubrics combined with a reflective essay	9

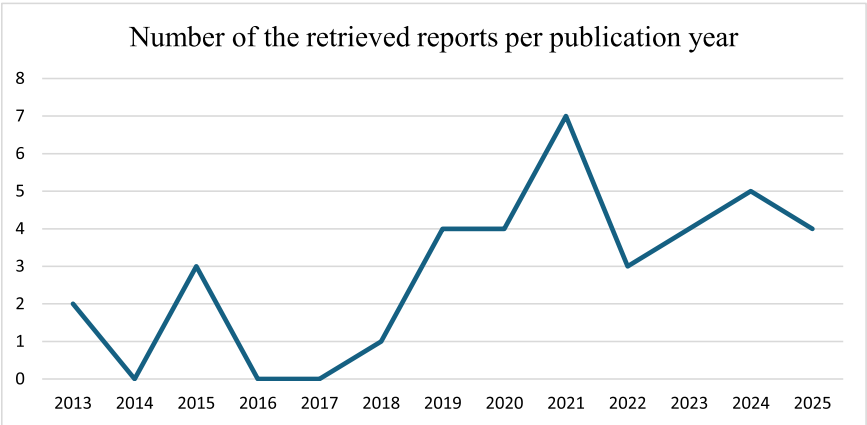


Fig. 2. Number of the retrieved reports ($n = 37$) per Publication year (2013–2026).

method). Table 1 outlines the methodologies and methods used. Subsequently, we discuss the specific methodologies, methods, and instruments they employ. To address RQ2, we describe reported strengths and limitations. As for RQ3, we report, later in this section, on the psychometric properties of the reported instruments.

4.1.1. Literature reviews

Some of the retrieved articles include systematic literature reviews that focus on assessment methodologies, methods, and instruments for undergraduate students in HEIs (Annelin & Boström, 2022; Cebrian & Segalas, & Hernandez, 2019; Redman et al., 2021; Segalàs & Hernández, 2019; Sossé et al., 2021). Although these reviews address the same issue, they adopt different perspectives. Cebrian, and Segalas, and Hernandez (2019) and Segalàs and Hernández (2019), for instance, searched for studies focusing on undergraduate students. The selected studies focus on students' competencies, skills, outcomes, perceptions, attitudes and their behaviors. The goal in most of these studies is summative rather than formative evaluation or self-assessment (Cebrian & Segalas, & Hernandez, 2019; Segalàs & Hernández, 2019). The review by Cebrian, and Segalas, and Hernandez (2019) investigates research on assessing sustainability competencies and related capabilities, learning outcomes, attributes of students of all educational levels (including HEI students), alumni and even professionals (e.g., farmers). The review by Segalàs and Hernández (2019) focuses on HEIs students of all disciplines. Annelin and Boström (2022) and Sossé et al. (2021) gathered studies with an exclusively quantitative design. Meanwhile, Cebrian, and Segalas, and Hernandez (2019), Redman et al. (2021) and Segalàs and Hernández (2019) include research with a variety of methodologies.

Sossé et al. (2021) selected studies employing a quasi-experimental design to detect the impact of ESD on students' behavior regarding SD issues. They found that the impact of ESD is indeed measurable. They also pinpoint that KSCs are often used as a means to measure it in educational research. All included studies prioritize the utilization of quantitative methodologies, which generally follow a positivistic approach. These studies use instruments that are either adapted from existing instruments or developed by the authors themselves. In all instances, the instruments include a list of (sub)competencies with various levels of mastery and are checked for validity and reliability.

Annelin and Boström (2022) found that most of the selected studies did not elaborate on all KSCs; instead, they focused on specific ones. The retrieved scales either measure general competence for sustainability instead of KSCs or they are validated in discipline-specific contexts. In the latter case, no information is provided regarding their validation in different contexts. They also noted that little research has been done on how to measure the integrated problem-solving competence, which is considered one of the KSCs according to the framework of Wiek and colleagues (2011; 2016). While this competence might be built upon other KSCs, the authors argue it is also a distinct competence. The literature reviews mentioned above reveal the application of a broad range of methods to assess KSCs in HEI students.

Indeed, Redman et al. (2021) clustered the research methods applied into three major categories based on who examines students' KSCs: a) self-perception-based instruments (Scaled self-assessment instruments, reflective writing, focus groups/interviews), b) observation-based instruments (performance observation and coursework analysis) and c) test-based instruments (concept mapping and scenario/case tests).

4.1.2. Qualitative approaches

Few researchers have applied qualitative methodologies to study bachelor students' Key Sustainability Competences. Azeiteiro et al. (2015), for instance, used semi-structured interviews to assess the effectiveness of a sustainability course in terms of their students' KSCs. Gordon and Thomas (2018) assessed their students via role-playing after the completion of a semester course. The students participated in a debate where they adopted several conflicting roles and argued on an SD-related topic. The authors analyzed their argumentation along with a

written course assignment. Giangrande et al. (2019) developed a conceptual framework for KSCs to serve as a basis for assessment instruments targeting learners, practitioners, teachers, and policy makers. For each KSC, they created open-ended questions for evaluation and self-reflection. The suggested competence framework provides a route to assess ESD outputs, aligning with SDG 4.7's goals of human rights, gender equality, peace, and global citizenship. Vesala-Varttala et al. (2021) asked students to write *reflective learning journals* during an international sustainability course with a focus on normative competence. To guide students on what to note down, the researchers provided them with reflection themes as guidelines. The data were examined using a theory-driven content analysis applying the KSCs definitions. This analysis involved KSCs divided into knowledge, attitudes, and skills. Konrad et al. (2021) assessed the development of students' interpersonal competence through project-based sustainability courses. They adopted a multi-case approach studying project-based sustainability courses in Germany, USA, Switzerland and Spain. They involved their students and teachers in series of semi-structured interviews, focus groups and observations. Finally, Muhamad Noor et al. (2024) organized semi-structured interviews with undergraduates in Malaysian universities. They investigated their perceptions of which KSCs were cultivated through their engagement in extracurricular university activities. The students responded to open-ended questions and the collected data then were thematically analyzed.

4.1.3. Quantitative approaches

Many of the reported studies use *scaled assessment instruments* to investigate HEI students' KSCs. However, relatively few of these instruments are designed for students to evaluate their *self-perceptions* regarding KSCs related to knowledge, attitudes, and behavior concerning environmental or sustainability issues. For instance, the questionnaire by Muñoz-Rodríguez et al. (2020) comprises 19 statements that require responses based on a four-point Likert scale: strongly disagree, disagree, agree, and strongly agree. The authors measure HEI students KSCs at the beginning and the end of their studies. Chaikovska et al. (2024) examined the development of KSCs of Ukrainian student teachers during an English course. They report the structure of a questionnaire consisting of 36 statements referring to KSCs, with and a four-point Likert scale for student responses. They collected data using the questionnaire before and after the completion of the English language course. Baena-Morales, Urrea-Solano, et al. (2023) developed a scale to measure future physical education teachers' KSCs. The Physical Education Scale for Sustainable Development in Future Teachers (PESD-FT) questionnaire includes 20 statements and an eight-point Likert scale. The items are grouped into the three aspects of SD – the social, the economic and the environmental – and are directly associated with the SDGs. The data collected with PESD-FT by Baena-Morales, Urrea-Solano, et al. (2023) were subsequently used by Baena-Morales, García-Taibo, et al. (2023) to determine the curriculum's influence on students' KSCs. Xiao et al. (2024) constructed and validated a scale for assessing students' KSCs. The questionnaire includes (a) a mini-scenario as an introduction to make all students familiar with sustainability practices and (b) 29 statements with a six-point Likert scale. Lately, (Annelin, 2026) developed and validated a scale to measure students' perceptions of their KSCs, the Sustainability Competence Mindset Map (SCMM). The SCMM consists of 21 items and students were asked to rank their agreement with the statements on a five-point Likert scale, from strongly disagree to strongly agree.

Other studies examine more generic issues such as relevance and interest in sustainability. For instance, Risopoulos-Pichler et al. (2020) developed a scale comprising 24 statements with a six-point Likert-type scale. The 24 items refer to the personal relevance of sustainability, relevance of ESD, organizational initiatives to encourage sustainability, individual or community impact, and national or global perspectives on sustainability. Hyytinen et al. (2023) developed a questionnaire to collect data regarding students' interest in learning KSCs. The

questionnaire includes 24 statements and a seven-point Likert scale for student responses. Hammer and Lewis (2023) applied a survey to university students, asking them how important they considered the promotion of KSCs in their studies. The questionnaire consists of 13 statements, each representing a KSC, and respondents used a six-point Likert scale. Martínez Casanovas et al. (2022) employed cluster analysis to create profiles of graduate students' KSCs in conjunction with their perceptions of a series of active learning methodologies. To measure KSCs, they developed a scale based on the conceptual framework by Wiek et al. (2011).

Researchers in Spain developed an instrumentarium with self-assessment scales used in conjunction with a rubric known as the 'Sustainability Map'. Valderrama-Hernández et al., (2020, 2025) along with Fuertes-Camacho et al. (2019), Sánchez-Carracedo et al. (2021) and Valderrama-Hernández et al. (2020) present an 18-statement self-assessment scale with a four-point Likert scale. This scale is designed for students in education degree programmes to report their perceptions regarding their KSCs. Sánchez-Carracedo, Sabate, et al. (2021) and Sánchez-Carracedo et al. (2022) describe a similar scale, but for engineering students. The latter scale contains 34 statements and a four-point Likert scale for student responses. Both scales, for education and engineering degrees, are used in conjunction with their respective 'Sustainability Maps'. A Sustainability Map is a rubric that analyzes KSCs at three levels: *Knowing*, *Knowing how*, and *Showing and doing*, as described by Miller's pyramid (1990). For each of the three levels, there are three indicators (low, medium and high). The statements in both instruments are informed by the cells of the *Sustainability Maps*. In Sánchez-Carracedo et al. (2022), the researchers employed these instruments with a sample of both education and engineering students to facilitate comparisons.

Beyond student perspectives, Valdes-Ramirez et al. (2024) asked tutors to examine their students' KSCs. They utilized quantitative data collected by the Tecnológico de Monterrey, the Mexico's national university, to investigate students' KSCs. The collected data contained information on students' KSCs, their academic program, course, and the semester. Tutors assigned each KSC as 'observed' or 'not observed' based on a rubric. Similarly, Albareda-Tiana et al. (2024) used 'the Sustainability Map' for all university students, regardless of their discipline. Tutors of the academic courses in which the students participated used the rubric to assess their students' KSCs based on their oral and written assignments. The aim of this study was to evaluate the effectiveness of the teaching methodologies applied.

4.1.4. Mixed-method approaches

Many of the selected reports describe instruments structured using mixed-method methodologies. These studies typically collect (a) both qualitative and quantitative data that complement each other, or (b) qualitative data that are subsequently quantified. Often, the selected studies utilize *course evaluations* to assess students' acquisition of KSCs. In such cases, data collection processes are frequently carried out before and after the completion of a sustainability course (e.g., Albareda Tiana & Alférez Villarreal, 2016; Galt et al., 2013; Oh, 2019; Remington-Doucette et al., 2013; Remington-Doucette & Musgrove, 2015). For example, Galt et al. (2013) assessed the development of KSCs in university students through two complementary ways: (a) a self-assessment rubric and (b) a reflective essay where students described their course experience. Albareda Tiana and Alférez Villarreal (2016) evaluated the cultivation of student teachers' KSCs using both qualitative analysis of students' self-reported perceptions and an assessment by the tutors of the course. Oh (2019) combined data from course evaluation and student interviews to evaluate the development of interpersonal competence based on a rubric scoring system. Kioupi (2025) developed a survey with 14 items and a four-point Likert scale for students to assess their KSCs. The survey included also open-ended questions to illuminate the results of the closed-ended questions. The researcher has finally captured students' reflections via online tools such as Padlet and Vevox.

Schmidt-Ross and Hensel-Börner (2025) first collected qualitative data via interviews and reflection essays from students. Based on these data, they created a questionnaire, which was administrated to students at the beginning and the end of a sustainability course. The questionnaire included 36 items and a ten-point likert-type scale from 'low' to 'high'.

Some studies utilize vignette surveys, which often involve more than one data collection method. In a *vignette* survey, students are presented with a small number of short vignettes describing educational scenarios. They are then asked to reflect on these scenarios and respond to open-ended, close-ended questions, or scales. For instance, the instrument of Sandri et al. (2018) includes four vignettes, each followed by multiple-choice questions for students to respond to, along with one close-ended question. The answers are accompanied by a rubric scoring system designed to indicate students' capacity to apply KSCs in practice. Similarly, Renta-Davids et al. (2020) collected responses on four vignettes using multiple choices questions, which were then statistically analyzed. They supplemented this with one open-ended question, and these data are analyzed thematically (content analysis). Remington-Doucette et al. (2013) and Remington-Doucette and Musgrove (2015) employed two vignettes presenting sustainability challenges and potential solutions. Students were required to respond to accompanying open-ended questions. These questions were supported by an answer key and a rubric scoring system based on specific criteria, further categorized at five levels ranging from "exceptional" to "no skill".

4.1.5. Strengths of the reported methods

Surveys are widely utilized due to their efficiency, ease of distribution, and ability to gather vast amounts of information from larger student populations (Cebrian & Segalas, & Hernandez, 2019; Redman et al., 2021; Segalàs & Hernández, 2019; Valderrama-Hernández et al., 2025). Research participants often acknowledge that they are more likely to engage in research that requires less time (Sandri et al., 2018). Open-ended question formats encourage thoughtful responses, debate, and honest feedback, thereby promoting self-assessment (Giangrande et al., 2019; Kioupi, 2025).

Scaled self-assessment instruments are the most commonly applied method (Redman et al., 2021). These instruments allow researchers to analyze individual items or indicators, or to group them as aggregate factors that describe a specific KSC. (Sánchez-Carracedo, Sabate, et al., 2021). Scaled instruments can measure not only perceived learning but also students' actual learning (Sánchez-Carracedo, Sabate, et al., 2021). They also offer formative self-evaluation opportunities allowing us to deep into students' perspectives. Finally, statistical tests help us to confirm the validity and reliability of the instruments (Annelin, 2026). During the development process, though, researchers need to balance contextual information with the cognitive demands placed on respondents. (Xiao et al., 2024).

A few authors utilize the *vignette survey* method for data collection. Vignettes are seen as a realistic stimuli (e.g., images, graphics, video or text), simulating familiar situations for participants, aiming to elicit responses (Sossé et al., 2021). They are typically followed by a set of questions designed to uncover the reasoning behind participants' responses (Renta-Davids et al., 2020). Vignettes enhance the internal and external validity of an instrument, allowing students to demonstrate SKCs within a realistic, hypothetical context (Renta-Davids et al., 2020). The combination of open-ended and close-ended questions facilitates data triangulation (Sossé et al., 2021). Vignette-based question designs can gather comparable data from large populations (Sandri et al., 2018) and, as such, can be used in large-scale studies.

Students are sometimes asked to maintain *reflective learning journals* to document their KSCs, typically before and after the completion of a sustainability course (Redman et al., 2021). These journals help students understand alternative interpretations of sustainability challenges and ethical encounters. By integrating competences, students gain self-knowledge and confidence in self-directed learning (Vesala-Varttala et al., 2021). They are also specifically designed to assess students' KSCs.

One case study reported the use of role-playing alongside written course assignments as a method to assess students' KSCs, which allowed students to identify the KSCs themselves (Gordon & Thomas, 2018). Additionally, *focus groups and interviews* connect learning outcomes to teaching activities, encouraging further reflection and insights through discussions with facilitators or peers (Redman et al., 2021).

4.1.6. Weaknesses of the reported methods

Participants may fail to accurately respond to the survey questions, regardless of whether these questions pertain to behaviors, attitudes, perceptions, facts, and intentions (Renta-Davids et al., 2020; Sandri et al., 2018). They may not report their actual behaviour but rather what they perceive as ideal behaviour, leading to social desirability bias (Giangrande et al., 2019; Sandri et al., 2018; Sossé et al., 2021). Furthermore, while *vignette surveys* depict fictional scenarios, they do not always position the participant as the protagonist, thereby not ensuring that the participant will articulate their own behaviour (Renta-Davids et al., 2020; Sandri et al., 2018).

Scaled self-assessment instruments necessitate that students fully comprehend the provided statements or scenarios and engage in adequate self-reflection to select an appropriate point on a Likert-type scale (Annelin, 2026; Redman et al., 2021; Xiao et al., 2024). To capture actual behavior, researchers need qualitative research designs which are impractical for large samples (Sandri et al., 2018). A possible solution to this limitation could be to provide space for students to report specific examples (Sossé et al., 2021). Self-assessment instruments also do not allow assessments by peers (Sandri et al., 2018).

While quantitative methods are suitable for larger samples, qualitative methods are appropriate for small samples (Kioupi, 2025; Sandri et al., 2018). Regarding *reflective learning journals*, it is doubtful if the HEI students can accurately express their thoughts, values, and feelings (Vesala-Varttala et al., 2021). *Role-playing*, while effective, requires significant effort and experience from the teacher (Gordon & Thomas, 2018).

Performance observations may be contested due to subjectivity issues (Redman et al., 2021). They also demand significant time when observing a large number of students, and observers need appropriate training (Redman et al., 2021). To mitigate subjectivity in observations, and in analyzing data from course evaluations, a rubric that includes both the KSCs and related behaviours could be beneficial (Redman et al., 2021). Additionally, at least two observers/raters should assess the same sample to ensure interrater reliability (Redman et al., 2021).

4.1.7. Psychometric properties of the reported instruments

In this section, we examine the extent to which the retrieved empirical studies report on the psychometric properties of their instruments (i.e., pilot testing, validity and reliability) and examine them based on the COSMIN criteria (RQ3).

While quite a few of the quantitative studies report on aspects such as pilot-testing, expert review or factor analysis (for checking validity) and Cronbach's alpha coefficients (for examining internal consistency), this is not universally the case. In some reports, the instrument used for data collection is not presented (Valdes-Ramirez et al., 2024). The sample sizes differ considerably across the studies, reflecting on the different research design and objectives. Sometimes, the sample size of undergraduate students is not specified (Azeiteiro et al., 2015; Gordon et al., 2019). All of the reported instruments measure students' self-perceptions about their KSCs. Hyytinen et al. (2023) examined, though, students' perceived interest in KSCs. The majority of the studies employ summative approaches rather than formative ones. Table 2 provides information on the characteristics of the studies: sample, assessment approach, which KSCs are measured, total instrument items and items per measure, and response options. The most measured KSCs include systems-thinking competence, normative competence, strategic competence, interpersonal competence. Table 3 summarizes the COSMIN evaluation of psychometric properties and evidence quality. The

table shows the rating for the risk of bias for the psychometric property of the instruments and their quality rating regarding the statistical results. The studies that make use of the same instruments are group together and a GRADE is assigned for each psychometric property for each instrument. The table depicts that the selected studies report only few psychometric properties of their instruments. In most of the cases, the risk of bias is rated as *very good*, the quality of rating as *sufficient* and the GRADE as high. In what follows, we detail the risk of bias is rated, the quality of rating and the GRADE for each study and instrument.

4.1.8. Reliability

Many of the selected studies reported reliability testing to examine the instruments' internal consistency. Most authors used Cronbach's alpha as the main reliability indicator, with values ≥ 0.70 generally considered indicative of acceptable internal consistency (e.g., Loewenthal & Lewis, 2018). Some studies also reported composite reliability (CR) (Annelin, 2026; Xiao et al., 2024), for which values in the range of 0.60–0.90 are typically regarded as acceptable (Hair et al., 2019).

EDINSOST-project questionnaire. The Spanish team reported the EDINSOST-project questionnaire for education and engineering students. The questionnaire for the education students demonstrated high reliability with a Cronbach's alpha coefficient of 0.935, indicating strong internal consistency (Muñoz-Rodríguez et al., 2020; Sánchez-Carracedo et al., 2021). Accordingly, the risk of bias (RoB) was rated as *very good*, and internal consistency result was rated as *sufficient*. Although the education-student questionnaire was used in additional studies (Fuertes-Camacho et al., 2019; Sánchez-Carracedo et al., 2022; Valderama-Hernández et al., 2020), internal consistency was reported only in the aforementioned publications. In contrast, the version for engineering students was not evaluated for reliability in the included studies (Albareda-Tiana et al., 2024; Sánchez-Carracedo et al., 2022; Sánchez-Carracedo, Sabate, et al., 2021).

PESD-FT scale. Baena-Morales and colleagues reported high reliability for the PESD-FT scale, with a Cronbach's alpha of 0.949, indicating strong internal. Additionally, the three sustainability dimensions displayed high internal consistency: social dimension (0.853), economic dimension (0.869), and environmental dimension (0.928). The risk of bias is considered *very good* and the quality of the internal consistency is rated as *sufficient* (Baena-Morales, García-Taibo, et al., 2023; Baena-Morales, Urrea-Solano, et al., 2023).

Additional instruments. Xiao et al. (2024) reported Cronbach's alpha coefficients ranging from 0.87 to 0.94, indicating high internal consistency for the scales assessed. The composite reliability (CR), exceeding 0.70, affirmed the scale's reliability. The internal consistency then is rated as *sufficient* and the risk of bias *very good*. Internal consistency was also evaluated in Annelin (2026) by using both Cronbach's alpha and Composite Reliability (CR) measures. The results show that both reliability tests for KSC scale yielded acceptable results.

Limited or absent reporting. Renta-Davids et al. (2020) reported $\alpha = 0.62$, which may be acceptable for scales with fewer than 10 items (Loewenthal & Lewis, 2018). Remington-Doucette and Musgrove (2015) assessed inter-rater reliability by Pearson's correlation analysis to ascertain the strength and direction of the ratings. Rater scores for all five competences demonstrated a positive and strong ($r = 0.50$ – 1.0 ; Cohen et al., 2007). Quite a few studies failed to present any evidence of reliability (Chaikovska et al., 2024; Galt et al., 2013; Hyytinen et al., 2023; Kioupi, 2025; Martínez Casanovas et al., 2022; Oh, 2019; Remington-Doucette et al., 2013; Remington-Doucette & Musgrove, 2015; Risopoulos-Pichler et al., 2020; Sandri et al., 2018). None of the included studies reported evidence for other types of reliability such as intra-rater reliability, test-retest reliability or measurement error.

4.1.9. Validity

4.1.9.1. Content validity. Quite a few studies have taken steps to

Table 2
Characteristics of the studies.

Report	Instruments used in the study	Sample	Assessment approach	*Measure(s): Self-perceptions of KSCs	Total items	No of items per measure	Response options
Galt et al. (2013)	Rubric	38 undergraduate students	Summative	Systems-thinking competence Normative competence Strategic competence Interpersonal skills Interpersonal competence			5-point Likert-type scale
Remington-Doucette et al. (2013)	Rubric	105 undergraduate students	Summative	Systems-thinking competence Normative competence Strategic competence	6	3 items 2 items 1 item	5-point Likert-type scale
Remington-Doucette and Musgrove (2015)	Rubric	195 undergraduate students	Summative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence	11	3 items 2 items 2 items 2 items 2 items	5-point Likert-type scale
Azeiteiro et al. (2015)	Semi-structured interviews	undergraduate students with no reference to the total	Summative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence Transdisciplinary competence			
Gordon and Thomas (2018)	Role-playing and written course assignments	undergraduate students with no reference to the total	Summative & formative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence			
Sandri et al. (2018)	Six scenario's/ vignettes with following questions	54 undergraduate students	Summative	Systems-thinking competence Normative competence Strategic competence Interpersonal competence	different for each vignette		A scoring system is similar to a standard test.
Oh (2019)	Rubric	14 undergraduate students.	Summative	Interpersonal competence	5 dimensions		A scoring system is similar to a standard test.
Giangrande et al. (2019)	Rubric		Summative & formative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence Intrapersonal competence	5 dimensions for each		Open-ended questions for (self-) reflection.
Renta-Davids et al. (2020)	Four vignettes	202 first year students	Summative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence			Multiple choices
Risopoulos-Pichler et al. (2020)	Questionnaire with 24 statements	3.200 undergraduate students	Summative	Systems-thinking competence Normative competence Strategic competence			6-point Likert-type scale
Muñoz-Rodríguez et al. (2020)	Questionnaire with 18 statements and a rubric	230 first-year students and 96 fourth-year education students	Summative	Systems-thinking competence Normative competence Strategic competence Interpersonal competence	18		4-point Likert-type scale
Vesala-Varttala et al. (2021)	Reflective learning journals	35 international undergraduates	Summative & formative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence Intrapersonal competence Implementation competence			

(continued on next page)

Table 2 (continued)

Report	Instruments used in the study	Sample	Assessment approach	*Measure(s): Self-perceptions of KSCs	Total items	No of items per measure	Response options
Konrad et al. (2021)	Interviews, focus groups and observations.	36 international undergraduates	Formative	Interpersonal competence			
Valderrama-Hernández et al. (2020)	A questionnaire with 18 statements and a rubric.		Summative	Systems-thinking competence Normative competence Strategic competence Interpersonal competence			4-point Likert-type scale
Sánchez-Carracedo, Sabate, et al., (2021)	A questionnaire with 18 statements and a rubric.	548 first-year students and 394 fourth-year education students.	Summative	the same as previously			4-point Likert-type scale
Fuertes-Camacho et al. (2019)	Rubric	16 education students	Summative	the same as previously			4-point Likert-type scale
Sánchez-Carracedo et al. (2022)	A questionnaire with 18 statements and a rubric	1.705 education students and 919 engineering students	Summative	the same as previously			4-point Likert-type scale
Sabate-Carracedo et al. (2021)	A questionnaire with 34 statements and a rubric	221 first-year students and 136 fourth-year engineering students	Summative	the same as previously			4-point Likert-type scale
Sánchez-Carracedo et al. (2021)	A questionnaire with 34 statements and a rubric	1157 first-year students and 525 fourth-year students.	Summative	the same as previously			4-point Likert-type scale
Albareda-Tiana et al. (2024)	Rubric	322 undergraduates	Summative	the same as previously			
Martínez Casanovas et al. (2022)	Questionnaire with 25 statements	252 undergraduates	Summative		7		No explicit reference to any response scale.
Baena-Morales, García-Taibo, et al. (2023)	Questionnaire with 20 statements	341 physical activity and sport science undergraduate students	Summative	Systems-thinking competence Normative competence Strategic competence Interpersonal competence	20	8 items 6 items 6 items	8-point Likert-type scale
Baena-Morales, Urrea-Solano, et al. (2023)	Questionnaire with 20 statements	341 physical activity and sport science undergraduate students	Summative	Systems-thinking competence Normative competence Strategic competence Interpersonal competence	20	8 items 6 items 6 items	8-point Likert-type scale
Hyytinen et al. (2023)	Questionnaire with 24 statements	797 undergraduates	Summative	**Perceived interest in learning sustainability competences: in Systems-thinking competence Normative competence Strategic competence Interpersonal competence Intrapersonal competence	15	6 items 5 items 7 items 4 items 3 items	7-point Likert-type scale
Xiao et al. (2024)	Questionnaire with 29 statements	381 undergraduates	Summative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence	16		6-point Likert-type scale
Chaikovska et al. (2024)	A questionnaire with 36 statements. Student interviews and classroom observations.	246 education degree students	Summative & formative	the same as previously		6 items for each measure	4-point Likert-type scale
Muhamad Noor et al. (2024)	Semi-structured interviews	21 undergraduates	Summative & formative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence			
Kioui (2025)	A questionnaire with 14 statements. Open question to accompany the closed ones.		Summative & formative	Systems-thinking competence Anticipatory competence Normative competence Strategic competence Interpersonal competence Intrapersonal competence			4-point Likert-type scale
Annelin (2026)	Questionnaire with 21 statements.	691 students	Summative & formative	Systems-thinking competence			5-point Likert-type scale

(continued on next page)

Table 2 (continued)

Report	Instruments used in the study	Sample	Assessment approach	*Measure(s): Self-perceptions of KSCs	Total items	No of items per measure	Response options
				Anticipatory competence Normative competence Strategic competence Interpersonal competence Intrapersonal competence Implementation competence			

Note. All of the reported instruments measure students' self-perceptions about their KSCs. *Most all of reports examines university students' self-perceptions as to their KSCs. We report the KSCs seen through the lenses of the Sustainability Research and Problem-solving Framework (Wiek et al., 2011; 2016). ** Only Hyytinen et al. (2023) examined university students' perceived interest in KSCs.

strengthen the content validity of their instruments. Researchers often submit the instrument to a panel of experts for content evaluation (Baena-Morales, García-Taibo, et al., 2023; Baena-Morales, Urrea-Solano, et al., 2023; Renta-Davids et al., 2020; Sánchez-Carracedo et al., 2021; Sánchez-Carracedo, Sabate, et al., 2021; Xiao et al., 2024). In addition, some authors have conducted cognitive interviews (Chaikovska et al., 2024; Xiao et al., 2024), focus groups with university students (Martínez Casanovas et al., 2022; Sandri et al., 2018) and, in some cases, classroom observations to inform instrument development (Chaikovska et al., 2024). Pilot testing of the initial version is also commonly used as part of the validation process (Annelin, 2026; Galt et al., 2013; Martínez Casanovas et al., 2022; Sánchez-Carracedo, Sabate, et al., 2021, 2021, 2021; Xiao et al., 2024). Annelin (2026) invited teachers to provide feedback after implementing the instrument with their students; these data were then used for triangulation, supporting the validation of findings and assessing the questionnaire's usability in a real-world teaching context.

4.1.9.2. Hypothesis testing for construct validity. Structural validity evidence. Several studies reported using factor-analytic procedures to examine the internal structure of their instruments. Factor analysis is commonly used to evaluate the homogeneity of a measure (Heale & Twycross, 2015), while principal component analysis (PCA) was frequently employed in the included studies (Martínez Casanovas et al., 2022; Risopoulos-Pichler et al., 2020; Sánchez-Carracedo et al., 2021; Sánchez-Carracedo, Sabate, et al., 2021). Before conducting factor analysis or PCA, several authors assessed data suitability using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity (Annelin, 2026; Renta-Davids et al., 2020; Risopoulos-Pichler et al., 2020; Sánchez-Carracedo et al., 2021; Sánchez-Carracedo, Sabate, et al., 2021; Xiao et al., 2024). In general, KMO values > 0.50 are considered acceptable, while a statistically significant Bartlett's test supports the factorability of the correlation matrix (Mokkink et al., 2024).

EDINSOST project questionnaire—education students. For the education-student questionnaire, suitability testing supported PCA: the KMO for the full instrument was 0.939 and Bartlett's test of sphericity was significant ($\chi^2 = 9834.271$, $p < .001$) (Sánchez-Carracedo et al., 2022; Sánchez-Carracedo et al., 2021). Factor loadings were ≥ 0.30 and the solution explained 60% of the variance; therefore, the instrument was rated *very good* for methodological quality and *sufficient* for its results. **EDINSOST project questionnaire—engineering students.** For the engineering-student version, factor loadings were ≥ 0.30 , but the solution explained 47.77% of the variance and was therefore judged *doubtful* (Sánchez-Carracedo et al., 2021; Sánchez-Carracedo, Sureda, et al., 2021). In addition, the authors stated that KMO and Bartlett's test results were acceptable but did not report exact values (Sánchez-Carracedo et al., 2022; Sánchez-Carracedo, Sabate, et al., 2021), limiting transparency and interpretability.

Additional studies. Across the included studies, evidence for structural validity varied in both completeness and strength. Renta-Davids et al. (2020) reported KMO = 0.69 and a significant Bartlett's test ($\chi^2 = 86.47$,

$df = 6$, $p < .001$); however, the extracted solution explained 46.98% of the variance and showed modest loadings (0.473–0.592). Although the study was rated *very good* for methodological quality, its results were judged *doubtful* because the explained variance did not meet the $\geq 50\%$ criterion typically used to support sufficient factor structure, even when loadings meet the ≥ 0.30 threshold (Mokkink et al., 2024). In contrast, Risopoulos-Pichler et al. (2020) reported stronger structural evidence, with loadings ranging from 0.40 to 0.80 and 57% explained variance; accordingly, the study was rated *very good* for methodological quality and *sufficient* for its results. Other studies provided only partial reporting. Hyytinen et al. (2023) reported that the items were factorable (KMO = 0.94) but did not report Bartlett's test of sphericity. Factor loadings ranged from 0.35 to 0.86, while explained variance was not reported. We therefore rated the methodological quality as *adequate* and the statistical outcomes as *sufficient*. Similarly, Xiao et al. (2024) reported factor loadings ranging from 0.40 to 0.91 but did not provide the KMO value and did not report Bartlett's test of sphericity; accordingly, we rated the methodological quality as *adequate* and the statistical outcomes as *sufficient*.

Convergent and discriminant validity evidence. Beyond structural validity, Annelin (2026) reported additional evidence for construct validity and diagnostic checks. Convergent validity was evaluated using the average variance extracted (AVE), which exceeded 0.50, while discriminant validity was assessed using the heterotrait–monotrait ratio (HTMT), with all values below the 0.90 threshold (Hair et al., 2019). Collinearity among indicators was examined via variance inflation factors (VIFs), with values between 1 and 3 indicating no problematic multicollinearity (Hair et al., 2019). Confirmatory factor analysis indicated no standardized loadings below 0.70 (some approached 1.00), and no critical collinearity issues were identified. The study of Annelin (2026) is considered “very good” for its methodological quality and “sufficient” for its statistical results.

Insufficient or limited reporting. In several cases, reporting was insufficient to allow appraisal. Kioupi (2025) conducted an exploratory, small-scale survey but did not report any validity testing. Martínez Casanovas et al. (2022) stated that PCA was performed; however, no methodological or statistical details were provided, preventing appraisal of methodological quality and statistical outcomes. Baena-Morales, García-Taibo, et al. (2023) and Baena-Morales, Merma-Molina, et al. (2023) make no references on validity tests.

Criterion validity. None of the studied reports have reported any correlation with gold standard.

Cross-cultural validity/Measurement invariance. None of the studied reports have reported application in different cultural context.

4.1.9.3. Responsiveness. The majority of the studies adopt a cross-sectional design and therefore do not permit conclusions about changes in effects over time (Annelin, 2026; Baena-Morales, García-Taibo, et al., 2023; Baena-Morales, Merma-Molina, et al., 2023; Chaikovska et al., 2024; Fuertes-Camacho et al., 2019; Galt et al., 2013; I. J. Gordon et al., 2019; Hyytinen et al., 2023; Kioupi, 2025, 2025; Martínez

Table 3
COSMIN evaluation of psychometric properties and evidence quality.

Reports	Reliability			Measurement Error	Validity			
	Internal Consistency	Reliability (test-retest)	Inter-rater reliability		Content validity	Criterion Validity	Construct validity Structural Validity	Responsiveness Cross-cultural validity/ Measurement invariance
Galt et al. (2013)								
Risk of bias								
Quality rating								
GRADE								
Remington-Doucette et al. (2013)								
Risk of bias								V
Quality rating								+
Remington-Doucette and Musgrove (2015)								
Risk of bias		V						V
Quality rating		+						+
GRADE		H						H
Sandri et al. (2018)								
Risk of bias					V			
Quality rating					+			
GRADE					H			
Oh (2019)								
Risk of bias								
Quality rating								
GRADE								
Renta-Davids et al. (2020)								
Risk of bias	V				V		V	
Quality rating	+				+		?	
GRADE	H				H		M	
Risopoulos-Pichler et al. (2020)								
Risk of bias							V	
Quality rating							+	
GRADE							H	
Muñoz-Rodríguez et al. (2020)								
Risk of bias	V							
Quality rating	+							
Valderrama-Hernández et al. (2020)								
Risk of bias								
Quality rating								
Sánchez-Carracedo, Sabate, et al., 2021)								
Risk of bias	V				V		V	
Quality rating	+				+		+	
Fuertes-Camacho et al. (2019)								
Risk of bias								
Quality rating								
Sánchez-Carracedo et al. (2022)								
Risk of bias							V	
Quality rating							+	
GRADE	H				H		H	
Sánchez-Carracedo, et al. (2021)								
Risk of bias					V		V	
Quality rating					+		?	
Sánchez-Carracedo et al. (2021)								
Risk of bias					V		V	
Quality rating					+		+	
Albareda-Tiana et al. (2024)								
Risk of bias								
Quality rating								
GRADE					H		M	
Martínez Casanovas et al. (2022)								
Risk of bias					A			
Quality rating					?			
GRADE					M			

(continued on next page)

Table 3 (continued)

Reports	Reliability	Validity		
Baena-Morales, García-Taibo, et al. (2023)				
Risk of bias	V	V		
Quality rating	+	+		
Baena-Morales, Urrea-Solano, et al. (2023)				
Risk of bias	V	V		
Quality rating	+	+		
GRADE	H	H		
Hyytinen et al. (2023)				
Risk of bias		V		A
Quality rating		+		?
GRADE		H		L
Xiao et al. (2024)				
Risk of bias	V			A
Quality rating	+			+
GRADE	H			M
Chaikovska et al. (2024)				
Risk of bias		V		
Quality rating		+		
GRADE		H		
Kioupi (2025)				
Risk of bias				
Quality rating				
GRADE				
Annelin (2026)				
Risk of bias	V	V		V
Quality rating	+	+		+
GRADE	H	H		H

Note. Reports with the same instruments are grouped together. Risk of bias: (V) very good, (A) adequate, (D) doubtful, (I) inadequate. Quality rating: (+) sufficient, (?) indeterminate, (-) insufficient. GRADE: (H) high, (M) moderate, (L) low, (VL) very low.

Casanovas et al., 2022; Muñoz-Rodríguez et al., 2020; Renta-Davids et al., 2020; Risopoulos-Pichler et al., 2020; Sánchez-Carracedo et al., 2022; Sánchez-Carracedo et al., 2021; Sánchez-Carracedo, Sabate, et al., 2021; Sabate, et al., 2021, 2021; Valderrama-Hernández et al., 2020). In contrast, Remington and colleagues investigated the effects of a sustainability course on students' key sustainability competences (KSCs) while explicitly considering the responsiveness of their instrument—that is, its ability to detect change over time in the construct being measured (Mokkink et al., 2021). Responsiveness is regarded as an aspect of an instrument's longitudinal validity (Mokkink et al., 2021). In Remington-Doucette et al. (2013), students' (a) systems-thinking competence and (b) normative and strategic competence scores (treated collectively in that study) were assessed at the beginning and end of the semester. In Remington-Doucette and Musgrove (2015), the researchers examined systems-thinking, anticipatory, normative, strategic, and interpersonal competences over time. Across both studies, paired-sample *t*-tests indicated increases in competence scores, with mean gains ranging from 0.14 to 0.76 and moderate-to-large effect sizes ($h^2 = 0.10\text{--}0.44$) (Cohen, 2013). The methodological quality of these studies was rated as *very good*, and their designs were considered *sufficient* to support the reported outcomes.

5. Discussion

In this study, we made use of the Sustainability Research and Problem-Solving Framework to explore the methodologies, methods, and instruments HESD scholars use to assess the Key Sustainability Competences of university students. This framework functions as an interpretation lens in this study, not a passive backdrop. The framework supports the formulation of research questions around problem definition, informs the development of KSCs assessment instruments, and establishes normative benchmarks that determine evidence as legitimate progress toward Sustainable Development solutions. Finally, it imposes constraints on data-collection decisions, thereby directing the empirical trajectory in accordance with the theoretical framework.

We found that HESD research employs a broad spectrum of

methodologies, methods, and instruments for that purpose. The majority of the researchers report on quantitative approaches (i.e., scaled assessment instruments often in conjunction with a rubric) to measure and assess the KSCs of university students. Qualitative approaches, i.e., semi-structured interviews, role-playing, written course assignments, and reflective learning journals, or mixed-method approaches, such as course evaluations, vignette surveys, or self-assessment rubrics combined with a reflective essay, are only sporadically used.

The methodologies, methods, and instruments used have strengths as well as weaknesses. The selection of a particular methodology, method or instrument depends on several factors, e.g., the purpose of the assessment. For instance, scaled assessment instruments are considered convenient for summative assessments. As Cebrian, and Segalas, and Hernandez (2019) also found, despite many studies incorporating aspects of formative assessments, most research measures students' KSCs for summative purposes. Even when students assess their own KSCs at the beginning of a course, they are not actively involved in the development of the course. As Valdes-Ramirez et al. (2024) note, these initial assessments indicate the non-presence of KSCs. This suggests that tutors should help their students cultivate them through their ESD teaching (Valdes-Ramirez et al. (2024)). However, students need to reflect on their own learning experience and assessment (Holt & Willard-Holt, 2000).

Performance-based assessment is more closely aligned with competence frameworks because it operationalises competence as demonstrated, context-sensitive capability—namely, the observable mobilisation of knowledge, skills, and values in authentic tasks—whereas self-report instruments primarily capture perceived competence and related dispositions, which are vulnerable to systematic bias and provide only indirect evidence of enacted performance. To this regard, we need to design instruments that prompt students to actively demonstrate their KSCs while providing supporting evidence. This process requires students to holistically mobilize their knowledge, skills, and attitudes to deliver optimal solutions for particular scenarios (Renta-Davids et al., 2020). Authentic assessment of KSCs involves students themselves, who indeed are uniquely positioned to articulate

their own competences (Kioupi, 2025). The KSCs, as framed by (Brundiens et al., 2021; Redman et al., 2021; Wiek et al., 2011, 2016) emphasise problem solving, higher-order thinking, and reasoning on sustainability real-life problems. Performance can be assessed either through the final product generated by students or through observation of the performance process, which requires implementation and the application of task-relevant knowledge and skills (Yilmaz, 2024). Students' performance can be evaluated against predetermined criteria (Daas et al., 2016). In a holistic rubric, the evaluator provides an overall judgment of performance quality, whereas in an analytic rubric, evaluators score each task criterion independently, enabling a clearer identification of pupils' strengths and learning needs (Yilmaz, 2024).

The scaled assessment instruments, as proven by this study but also by Redman et al. (2021), are the most frequently instruments for KSC assessment. Nevertheless, we found that many of the applied instruments do not report psychometrics properties or a clear procedure for determining them. Quite often, the instruments are not pilot-tested, and details about the sample or the instrument itself are not given, either. In most cases, the authors report only a small subset of psychometric features of the instruments, leaving many key properties unreported. When such information was provided, it was typically supported by high-certainty evidence (GRADE). This limited evidence bases of the instruments the HESD research community without reliable and valid instruments. Moreover, the published instruments are used in a small number of studies (typically one or two). This does not help the validation and use of the instrument.

Among the measurement properties proposed by COSMIN—reliability (internal consistency, inter-/intra-rater reliability, test–retest reliability, and measurement error), validity (content validity, criterion validity, and construct validity, including structural validity and cross-cultural validity/measurement invariance), and responsiveness—authors most frequently reported internal consistency, content validity, and structural validity. These properties are indeed central to establishing an instrument's reliability and validity when quantifying theoretical constructs (Gray et al., 2019; Hajjar, 2018). However, content validity—i.e., the extent to which items adequately represent the construct—typically relies on expert judgement rather than statistical evidence (Kline, 2023). Moreover, most scales used a single reliability indicator, usually Cronbach's alpha; yet alpha is a lower-bound estimate and may underestimate true score reliability (Sijtsma, 2009). Reliability should therefore be examined using two or more complementary indices, such as composite reliability, McDonald's omega (ω), inter-item correlations, and item–total correlations. Finally, Rasch analysis was employed less frequently than factor-analytic approaches across the reviewed scales. Because Rasch analysis and factor analysis are complementary, their combined use can strengthen instrument development and enable a more nuanced interpretation of psychometric evidence (Richardson, 2005). We, therefore, argue for the careful development, validation, and reporting of instruments used to measure KSCs. The authors should also justify their methodological choices (Redman et al., 2021).

KSCs are seen as integrated, multidimensional, and performance-oriented capabilities implemented within specific contexts (Wiek et al., 2011, 2016; Brundiens et al., 2021). To this sense, vignette-based designs that incorporate realistic scenarios and stimulate students' reasoning (and, when feasible, their intended actions) may offer a more accurate reflection of KSC assessment compared to decontextualised self-report measures. We advocate for the increased implementation of authentic performance tasks—such as projects, simulations, portfolios, or role-playing—coupled with rubric-based assessment to effectively capture integrated sustainability problem-solving. Performance observation allows an expert to evaluate the students' KSCs in context while also capture unexpected occurrences.

While performance-based approaches can be efficient measurement method, policy institutions often favor quantitative data to be able to evaluate potential advancements (Giangrande et al., 2019).

Nonetheless, Likert-type self-report instruments are vulnerable to social desirability bias and may reflect perceptions rather than actual KSCs. The overwhelming majority of the identified instruments examines students' self-perceptions as to their KSCs. Only one study (Hyytinen et al., 2023) focused on students' perceived interest in KSCs. Matthias Barth (2011) underscores that university students' perceptions of sustainability-related tasks do not necessarily align with their actual performance. This suggests a potential gap between intended and enacted behaviors of students. Context-specific constraints, such as interdisciplinary complexities and institutional barriers, may impede the enactment of students' perceptions (Barth, 2011; Braßler & Sprenger, 2021).

Also, none of the reported instruments explicitly measure the meta-competence or integrated problem-solving competence. This gap has also been identified by Annelin and Boström (2022), who note that limited research has addressed how these competences can be assessed. This may be due to the fact that integrated competences cannot be effectively captured through exclusively quantitative methodologies, which dominate KSC measurement; instead, qualitative approaches can better reveal the complex interrelations among competences in open and context-sensitive ways (Macchi Silva & Ribeiro, 2021).

The implementation of a single method to measure the KSCs of university students may introduce bias. Employing a wide range of methods helps us gain a more holistic overview of KSCs cultivation among students (Cebrian & Segalas, & Hernandez, 2019; Ssossé et al., 2021; Valderrama-Hernández et al., 2020). Mixed-method methodologies allow us to gather data from different aspects of the same phenomenon, to use one method to build the next one, and finally, to triangulate the results (Annelin, 2026; Cebrian & Cebrián, 2019; Kioupi, 2025; Ssossé et al., 2021).

The development, application, and validation of appropriate assessment instruments to measure students' KSCs remain an enormous challenge due to both the complex nature of sustainability and the intricate notion of competence (Mulà et al., 2022; Sandri et al., 2018). To overcome these difficulties, instruments should be easily understandable for the respondents. The terms used should be unequivocal to the reader and the statements or scenarios contextualized (Sandri et al., 2018). Non-response represents another major threat for data collection (Lozano & Barreiro-Gen, 2021). Making instruments accessible to the respondents will help minimize this treat. Annelin and Boström (2022) advocate for an instrument (probably a scale) to measure the KSCs of university students across disciplines. Consequently, we need to develop and validate an instrument that will be user-friendly for university students and facilitate the data analysis through multiple methodologies.

Yet, the choice of research methodologies, methods, and instruments should be tailored to the purpose of the assessment. When assessment is intended to inform quality assurance and strategic planning, scalable instruments can support comparisons across disciplines and cohorts. In such cases, it is useful to capture not only students' perceptions of their thinking and actions, but also relevant aspects of teaching and learning activities; scaled instruments may therefore be complemented with performance-based approaches (e.g., reflective journals and rubric-guided evaluation). For curriculum and programme evaluation, rubric-based assessments, vignette-based instruments, and other performance tasks—applied, for example, at the beginning and end of a study programme—can be used to examine learning outcomes, including students' KSC development, and to enable comparisons across programmes and disciplines. At the classroom level, where assessment is typically formative, vignette-based prompts, reflective journals, peer assessment, and rubric-based feedback can provide richer insights into students' reasoning and actions in relation to sustainable development issues.

5.1. Limitations and future research

In this review, we have specifically focused on the following terms:

“competence” and “competency”. While HESD literature considers a range of similar terms such as “capabilities”, “skills”, “literacy”, “abilities”, “attributes”, and “learning outcomes” (e.g., Sterling et al., 2017), we have chosen to adhere to “Key Sustainability Competences” (KSCs). This lies on the fact that there is a growing body of HESD research focusing explicitly on KSCs. Sterling et al. (2017) observed that “competence” or “competency” was used in the majority of the reports they studied. We acknowledge that this specific focus may impose certain limitations such as potentially overlooking studies that use similar research instruments but different terminology. Additionally, we did not include methodologies and instruments designed to measure each KSC individually. Future research could also consider these studies.

As this review demonstrates, there is an increasing body of research in the field of HESD that empirically investigates university students’ KSCs. The most frequently examined KSCs—systems-thinking, normative, strategic, and interpersonal competence—are typically measured using multi-item scales, with separate item sets for each competence. In contrast, the competences more recently added to the framework—intrapersonal competence, transdisciplinary competence, meta-competence and implementation competence—are assessed only rarely (e.g., intrapersonal competence) or not at all. Given that the Sustainability Research and Problem-Solving Framework sees KSCs as integrated and thus, instruments should assess them not in isolated scales but in an integrated manner, as well. Further research will allow us to examine whether the Sustainability Research and Problem-solving Framework can be effectively used to improve university students’ KSCs. Empirical evidence will help us validate or potentially refine the Sustainability Research and Problem-solving Framework. Future research should examine the validity of vignette-based and performance-based assessments in comparison to self-report measures. Moreover, HESD research should test the measurement invariance and cross-cultural validity of these instruments across disciplines and educational contexts. Finally, intervention designs are needed to test the potential of the instruments to capture KSCs development over time.

6. Conclusion

This systematic literature review offers a comprehensive synthesis of the methodologies, methods, and instruments used to measure the Key Sustainability Competences (KSCs) of university students. The review reveals that HESD researchers employ a broad range of methodologies, methods, and instruments for that purpose. Quantitative approaches (particularly scaled instruments) are favored over qualitative or mixed-method ones. These scaled instruments are primarily used for summative assessment purposes. The psychometric properties of these instruments are sometimes not reported, hindering readers from evaluating their quality. While the number of assessment instruments is growing, more empirical research is critically needed to establish which of them are reliable, valid, convenient to use, and useful for drawing conclusions.

CRedit authorship contribution statement

Maria Daskolia: Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Eleni Sinakou:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.stueduc.2026.101609.

References

- Albareda Tiana, S., & Alf  rez Villarreal, A. (2016). A collaborative programme in sustainability and social responsibility. *International Journal of Sustainability in Higher Education*, 17(5), 719–736. <https://doi.org/10.1108/IJSHE-07-2016-0134>
- Albareda-Tiana, S., Fernandez-Borsot, G., Berbegal-Mirabent, J., Regadera G  nzalez, E., Mas-Machuca, M., Graell, M., Manresa, A., Fern  ndez-Morilla, M., Fuertes-Camacho, M. T., Guti  rrez-Sierra, A., & Guardiola, J. M. (2024). Enhancing curricular integration of the SDGs: Fostering active methodologies through cross-departmental collaboration in a Spanish university. *International Journal of Sustainability in Higher Education*, 25(5), 1024–1047. <https://doi.org/10.1108/IJSHE-07-2023-0299>
- Annelin, A. (2026). Sustainability competence mindset: Building capacity with students. *International Journal of Sustainability in Higher Education*, 1–23. <https://doi.org/10.1108/IJSHE-04-2025-0369>
- Annelin, A., & Bostr  m, G. O. (2022). An assessment of key sustainability competencies: A review of scales and propositions for validation. *International Journal of Sustainability in Higher Education*, 24(9), 53–69. <https://doi.org/10.1108/IJSHE-05-2022-0166>
- Arene. (2022). Recommendation on the Shared Competences of Universities of Applied Sciences and their Application. (<https://www.arene.fi/wp-content/uploads/Raportit/2022/Kompetenssit/RECOMMENDATION%20ON%20THE%20SHARED%20COMPETENCES%20OF%20UNIVERSITIES%20OF%20APPLIED%20SCIENCES%20AND%20THEIR%20APPLICATION.pdf>).
- Azeiteiro, U. M., Bacelar-Nicolau, P., Caetano, F. J. P., & Caeiro, S. (2015). Education for sustainable development through e-learning in higher education: Experiences from Portugal. *Journal of Cleaner Production*, 106, 308–319. <https://doi.org/10.1016/j.jclepro.2014.11.056>
- Baena-Morales, S., Garc  a-Taibo, O., Merma-Molina, G., & Ferriz-Valero, A. (2023). Analysing the sustainability competencies of preservice teachers in Spain. *Journal of Applied Research in Higher Education*, 15(3), 731–744. <https://doi.org/10.1108/JARHE-02-2022-0040>
- Baena-Morales, S., Merma-Molina, G., & Ferriz-Valero, A. (2023). Integrating education for sustainable development in physical education: Fostering critical and systemic thinking. *International Journal of Sustainability in Higher Education*. <https://doi.org/10.1108/IJSHE-10-2022-0343>
- Baena-Morales, S., Urrea-Solano, M., Gavil  n-Martin, D., & Ferriz-Valero, A. (2023). Development and validation of an instrument to assess the level of sustainable competencies in future physical education teachers. Questionnaire. *Journal of Applied Research in Higher Education*, 15(1), 1–19. <https://doi.org/10.1108/JARHE-09-2021-0330>
- Barth. (2011). Higher education for sustainable development: Students’ perspectives on an innovative approach to educational change. *Journal of Social Sciences*, 7(1), 13–23. <https://doi.org/10.3844/jssp.2011.13.23>
- Bearman, M., Smith, C. D., Carbone, A., Slade, S., Baik, C., Hughes-Warrington, M., & Neumann, D. L. (2012). Systematic review methodology in higher education. *Higher Education Research & Development*, 31(5), 625–640. <https://doi.org/10.1080/07294360.2012.702735>
- Bianchi, G., Pisiotis, U., & Cabrera Giraldez, M. (2022). GreenComp The European sustainability competence framework. In Y. Punie, & M. Bacigalupo (Eds.), *EUR 30955 EN. Luxembourg*: Publications Office of the European Union. <https://doi.org/10.2760/821058. JRC128040>
- Biggs, J. (1996). Enhancing Teaching through Constructive Alignment (Pt 347-364). *Higher Education*, 32(3).
- Black, P., & William, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Bra  ler, M., & Sprenger, S. (2021). Fostering Sustainability Knowledge, Attitudes, and Behaviours through a Tutor-Supported Interdisciplinary Course in Education for Sustainable Development. *Sustainability*, 13(6), 3494. <https://doi.org/10.3390/su13063494>
- Brundiers, K., Barth, M., Cebri  n, G., Cohen, M., Diaz, L., Doucette-Remington, S., Dripps, W., Habron, G., Harr  , N., Jarchow, M., Losch, K., Michel, J., Mochizuki, Y., Rieckmann, M., Parnell, R., Walker, P., & Zint, M. (2021). Key competencies in sustainability in higher education—Toward an agreed-upon reference framework. *Sustainability Science*, 16(1), 13–29. <https://doi.org/10.1007/s11625-020-00838-2>
- Cebrian, G., & Cebri  n, G. (2019). Assessment of sustainability competencies: A literature review and pathways for future research and practice Standard-Nutzungsbedingungen: CENTRAL european review of economics and management assessment of sustainability competencies: A Literature Review and Future pathways for ESD Research and Practice33194410.29015/cerem.664.
- Cebrian, G., Segalas, J., & Hernandez, A. (2019). Assessment of sustainability competencies: A literature review and future pathways for ESD research and practice. *Central European Review of Economics and Management*, 3(3), 19–44. <https://doi.org/10.29015/cerem.664>
- Chaikovska, H., Levchuk, I., Adamska, Z., & Yankovych, O. (2024). Formation of sustainable development competencies in Ukrainian English for specific purposes students. *International Journal of Sustainability in Higher Education*, 25(4), 744–766. <https://doi.org/10.1108/IJSHE-07-2023-0306>
- Coe, R., Waring, M., Hedges, L., & Ashley, L. (2025). *Research Methods and Methodologies in Education* (4th ed.). SAGE Publications.
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*, 0 edn. Routledge. <https://doi.org/10.4324/9780203771587>
- Cohen, L., Manion, L., & Morrison, K. (2007). Research Methods in Education. *In Education*, 55. https://doi.org/10.1111/j.1467-8527.2007.00388_4.x

- Critical Appraisal Skills Programme (Systematic Review) Checklist. (2023). (<https://cas-p.uk.net/casp-tools-checklists/>).
- Daas, R., Ten Dam, G., & Dijkstra, A. B. (2016). Contemplating modes of assessing citizenship competences. *Studies in Educational Evaluation*, 51, 88–95. <https://doi.org/10.1016/j.stueduc.2016.10.003>
- De Kraker, J., Lansu, A., & van Dam-Mieras, R. (2007). Competences and competence-based learning for sustainable development. In J. De Kraker, A. Lansu, & R. Van Dam-Mieras (Eds.), *In Crossing Boundaries. Innovative Learning for Sustainable Development in Higher Education* (pp. 103–114). UNU Press.
- Dlouhá, J., Heras, R., Mulà, I., Salgado, F. P., & Henderson, L. (2019). Competences to address SDGs in higher education—A reflection on the equilibrium between systemic and personal approaches to achieve transformative action. *Sustainability*, 11(13), 3664.
- Erlinger, A. (2018). Outcomes Assessment in Undergraduate Information Literacy Instruction: A Systematic Review. *College & Research Libraries*, 79(4), 442–479. <https://doi.org/10.5860/crl.79.4.442>
- Frey, A., Hartig, J., & Moosbrugger, H. (2009). Effekte des adaptiven Testens auf die Motivation zur Testbearbeitung am Beispiel des Frankfurter Adaptiven Konzentrationsleistungs-Tests. *Diagnostica*, 55(1), 20–28. <https://doi.org/10.1026/0012-1924.55.1.20>
- Fuertes-Camacho, M. T., Graell-Martín, M., Fuentes-Loss, M., & Balaguer-Fàbregas, M. C. (2019). Integrating sustainability into higher education curricula through the project method, a global learning strategy. *Sustainability (Switzerland)*, 11(3). <https://doi.org/10.3390/su11030767>
- Galt, R. E., Parr, D., & Jagannath, J. (2013). Facilitating competency development in sustainable agriculture and food systems education: A self-assessment approach. *International Journal of Agricultural Sustainability*, 11(1), 69–88. <https://doi.org/10.1080/14735903.2012.683569>
- Giangrande, N., White, R. M., East, M., Jackson, R., Clarke, T., Coste, M. S., & Penha-Lopes, G. (2019). A competency framework to assess and activate education for sustainable development: Addressing the UN sustainable development goals 4.7 challenge. *Sustainability (Switzerland)*, 11(10). <https://doi.org/10.3390/su1102832>
- Glasser, H. (2016). Towards the development of robust learning for sustainability core competencies. *Sustainability: The Journal of Record*, 9(3), 121–134. <https://doi.org/10.1089/SUS.2016.29054.hg>
- Gora, A. A., Ștefan, S. C., Popa, Ștefan C., & Albu, C. F. (2019). Students' perspective on quality assurance in higher education in the context of sustainability: A PLS-SEM approach. *Sustainability*, 11(17), 4793.
- Gordon, I. J., Bawa, K., Bammer, G., Boone, C., Dunne, J., Hart, D., Hellmann, J., Miller, A., New, M., Ometto, J., Pickett, S., Wendorf, G., Agrawal, A., Bertsch, P., Campbell, C. D., Dodd, P., Janetos, A., Mallee, H., & Taylor, K. (2019). Forging future organizational leaders for sustainability science. *Nature Sustainability*, 2(8), 647–649. <https://doi.org/10.1038/s41893-019-0357-4>
- Gordon, S., & Thomas, I. (2018). The learning sticks: Reflections on a case study of role-playing for sustainability. *Environmental Education Research*, 24(2), 172–190. <https://doi.org/10.1080/13504622.2016.1190959>
- Gray, J. S., Connolly, J. P., & Brown, M. A. (2019). Measuring intercultural knowledge and competence in college essays: Does a performance-based rubric have construct validity? *Studies in Educational Evaluation*, 62, 142–148. <https://doi.org/10.1016/j.stueduc.2019.05.007>
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2–24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Hajjari, S. (2018). Statistical analysis: Internal-consistency reliability and construct validity. *International Journal of Quantitative and Qualitative Research Methods*, 27–38.
- Hammer, T., & Lewis, A. L. (2023). Which competencies should be fostered in education for sustainable development at higher education institutions? Findings from the evaluation of the study programs at the University of Bern, Switzerland. *Discover Sustainability*, 4(1), 19. <https://doi.org/10.1007/s43621-023-00134-w>
- Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence Based Nursing*, 18(3), 66–67. <https://doi.org/10.1136/eb-2015-102129>
- Holt, D. G., & Willard-Holt, C. (2000). Let's Get Real™: Students Solving Authentic Corporate Problems. *Phi Delta Kappan*, 82(3), 243–246. <https://doi.org/10.1177/003172170008200315>
- Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O' Cathain, A., Rousseau, M.-C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information*, 34(4), 285–291. <https://doi.org/10.3233/EFI-180221>
- Hyttinen, H., Laakso, S., Pietikäinen, J., Ratvio, R., Ruippo, L., Tuononen, T., & Vainio, A. (2023). Perceived interest in learning sustainability competencies among higher education students. *International Journal of Sustainability in Higher Education*, 24(9), 118–137. <https://doi.org/10.1108/IJSHE-06-2022-0198>
- Kiouti, V. (2025). Curriculum Design to Foster Sustainability Competence Development, a Case Study on Surfacing Sustainability Knowledge and Skills at the University of Leeds (Competence Building in Sustainable Development). In A. Annelin, & G.-O. Boström (Eds.), *W. Leal Filho, B. Gaspardo Rebelatto* (pp. 85–103). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-80380-2_5 (Competence Building in Sustainable Development).
- Kline, R. (2023). *Principles and Practice of Structural Equation Modeling*. Guilford Publications.
- Konrad, T., Wiek, A., & Barth, M. (2021). Learning processes for interpersonal competence development in project-based sustainability courses – insights from a comparative international study. *International Journal of Sustainability in Higher Education*, 22(3), 535–560. <https://doi.org/10.1108/IJSHE-07-2020-0231>
- Loewenthal, K. M., & Lewis, C. A. (2018). *An Introduction to Psychological Tests and Scales*. Psychology Press. <https://doi.org/10.4324/9781315782980>
- Looney, J. (2011). Integrating Formative and Summative Assessment: Progress Toward a Seamless System? (OECD Education Working Papers No. 58; OECD Education Working Papers, Vol. 58). <https://doi.org/10.1787/5kgx3kbl734-en>
- Lozano, R., & Barreiro-Gen, M. (2021). *Developing Sustainability Competences Through Pedagogical Approaches: Experiences from International Case Studies*. Springer International Publishing. (<https://books.google.gr/books?id=X4okEAAQBAJ>)
- Macchi Silva, V. V., & Ribeiro, J. L. D. (2021). A discussion on using quantitative or qualitative data for assessment of individual competencies. *Personnel Review*, 50(6), 1460–1478. <https://doi.org/10.1108/PR-08-2019-0444>
- Martínez Casanovas, M., Ruiz-Munzón, N., & Buil-Fabregá, M. (2022). Higher education: The best practices for fostering competencies for sustainable development through the use of active learning methodologies. *International Journal of Sustainability in Higher Education*, 23(3), 703–727. <https://doi.org/10.1108/IJSHE-03-2021-0082>
- Michel, J. O., Siciliano, P., Zint, M., & Collins, S. (2024). Toward diversifying higher education sustainability competency scholarship: Findings and implications from a bibliometric analysis. *International Journal of Sustainability in Higher Education*, 25(2), 221–237. <https://doi.org/10.1108/IJSHE-08-2022-0250>
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9), S63–S67. <https://doi.org/10.1097/00001888-199009000-00045>
- Mokkink, L. B., Elsmann, E. B. M., & Terwee, C. B. (2024). COSMIN guideline for systematic reviews of patient-reported outcome measures version 2.0. *Quality of Life Research*, 33(11), 2929–2939. <https://doi.org/10.1007/s11136-024-03761-6>
- Mokkink, L., Terwee, C., & De Vet, H. (2021). Key concepts in clinical epidemiology: Responsiveness, the longitudinal aspect of validity. *Journal of Clinical Epidemiology*, 140, 159–162. <https://doi.org/10.1016/j.jclinepi.2021.06.002>
- Muhamad Noor, M. F., Mamat, M. Z., & Mohamad, Z. F. (2024). Impact of engagement in campus sustainability activities to competency development: Change agents' experiences and perspectives. *Sustainability*, 16(5), 1780. <https://doi.org/10.3390/su16051780>
- Mulà, I., Cebrían, G., & Junyent, M. (2022). Lessons Learned and Future Research Directions in Educating for Sustainability Competencies. In P. Vare, N. Lausset, & M. Rieckmann (Eds.), *Competences in Education for Sustainable Development* (pp. 185–194). Springer International Publishing. https://doi.org/10.1007/978-3-030-91055-6_22
- Muñoz-Rodríguez, J. M., Sánchez-Carracedo, F., Barrón-Ruiz, Á., & Serrate-González, S. (2020). Are we training in sustainability in higher education? Case study: Education degrees at the university of Salamanca. *Sustainability (Switzerland)*, 12(11). <https://doi.org/10.3390/su12114421>
- Oh, D. S. (2019). Building inter-personal competence in architecture and urban design students through smart cities at a higher education institution. *Sustainability*, 11(24), 7179. <https://doi.org/10.3390/su11247179>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Redman, A., Wiek, A., & Barth, M. (2021). Current practice of assessing students' sustainability competencies: A review of tools. *Sustainability Science*, 16(1), 117–135. <https://doi.org/10.1007/s11625-020-00855-1>
- Remington-Doucette, S., & Musgrove, S. (2015). Variation in sustainability competency development according to age, gender, and disciplinary affiliation: Implications for teaching practice and overall program structure. *International Journal of Sustainability in Higher Education*, 16(4), 537–575. <https://doi.org/10.1108/IJSHE-01-2013-0005>
- Remington-Doucette, S. M., Hiller Connell, K. Y., Armstrong, C. M., & Musgrove, S. L. (2013). Assessing sustainability education in a transdisciplinary undergraduate course focused on real-world problem solving. *International Journal of Sustainability in Higher Education*, 14(4), 404–433. <https://doi.org/10.1108/IJSHE-01-2012-0001>
- Renta-Davids, A.-I., Camarero-Figuerola, M., & Tierno-García, J.-M. (2020). Assessment of the quality education awareness competence of pre-service educators using vignettes. *Sustainability (Switzerland)*, 12(23), 1–18. <https://doi.org/10.3390/su122310203>
- Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387–415. <https://doi.org/10.1080/02602930500099193>
- Risopoulos-Pichler, F., Daghofer, F., & Steiner, G. (2020). Competences for solving complex problems: A cross-sectional survey on higher education for sustainability learning and transdisciplinarity. *Sustainability (Switzerland)*, 12(15). <https://doi.org/10.3390/su12156016>
- Sánchez-Carracedo, F., Moreno-Pino, F. M., Romero-Portillo, D., & Sureda, B. (2021). Education for Sustainable Development in Spanish University Education Degrees. *Sustainability*, 13(3), 1467. <https://doi.org/10.3390/su13031467>
- Sánchez-Carracedo, F., Romero-Portillo, D., Sureda Carbonell, B., & Moreno-Pino, F. M. (2022). Education for sustainable development in Spanish higher education: An assessment of sustainability competencies in engineering and education degrees. *International Journal of Sustainability in Higher Education*, 23(4), 940–959. <https://doi.org/10.1108/IJSHE-02-2021-0060>
- Sánchez-Carracedo, F., Sabate, F., & Gibert, K. (2021). A methodology to assess the sustainability competencies in engineering undergraduate programs. *International Journal of Engineering Education*, 37(5), 1231–1243.
- Sánchez-Carracedo, F., Sureda, B., Moreno-Pino, F. M., & Romero-Portillo, D. (2021). Education for Sustainable Development in Spanish engineering degrees. Case study.

- Journal of Cleaner Production*, 294, Article 126322. <https://doi.org/10.1016/j.jclepro.2021.126322>
- Sandri, O., Holdsworth, S., & Thomas, I. (2018). Vignette question design for the assessment of graduate sustainability learning outcomes. *Environmental Education Research*, 24(3), 406–426. <https://doi.org/10.1080/13504622.2016.1263280>
- Schmidt-Ross, I., & Hensel-Börner, S. (2025). Competence Building Through Off-Campus Service Learning—Exploring the Impact of a ‘Sustainability Challenge’. In A. Annelin, & G.-O. Bostrom (Eds.), *Competence Building in Sustainable Development*. W. Leal Filho, B. Gasparetto Rebelatto (pp. 173–199). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-80380-2_9
- Segalàs, G. C. J., & Hernández, Á. (2019). Assessment of sustainability competencies: A literature review and future pathways for ESD research and practice. *Central European Review of Economics and Management*, 3(3), 19–44.
- Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>
- Sossé, Q., Wagner, J., & Hopper, C. (2021). Assessing the Impact of ESD: Methods, Challenges, Results. *Sustainability*, 13(5), 2854. <https://doi.org/10.3390/su13052854>
- Sterling, S., Glasser, H., Rieckmann, M., & Warwick, P. (2017). “More than scaling up”: A critical and practical inquiry into operationalizing sustainability competencies. In P. B. Corcoran, J. P. Weakland, & A. E. J. Wals (Eds.), *Envisioning futures for environmental and sustainability education* (pp. 153–168). Brill | Wageningen Academic. https://doi.org/10.3920/978-90-8686-846-9_10
- Teddle, C., & Tashakkori, A. (2011). Mixed methods research. *The Sage Handbook of Qualitative Research*, 4, 285–300.
- Thompson, J. (2022). A Guide to Abductive Thematic Analysis. *The Qualitative Report*. <https://doi.org/10.46743/2160-3715/2022.5340>
- Tójar-Hurtado, J.-C., Lechuga-Jiménez, C., & Esteban-Ibáñez, M. (2024). Evaluating and Developing Transversal and Sustainability Competencies in University Classrooms to Empower New Generations. *Education Sciences*, 14(8), 877. <https://doi.org/10.3390/educsci14080877>
- UNESCO. (2018). *Issues and trends in education for sustainable development*. UNESCO. <https://doi.org/10.54675/YELO2332>
- Valderrama-Hernández, R., Barrón Ruíz, A., Solís-Espallargas, C., & Limón Domínguez, D. (2025). An instrument to analyse the sustainability competence of university students in education sciences. *International Journal of Sustainability in Higher Education*. <https://doi.org/10.1108/ijshe-01-2024-0041>
- Valderrama-Hernández, R., Sánchez-Carracedo, F., Alcántara Rubio, L., & Limón Domínguez, D. (2020). Methodology to Analyze the Effectiveness of ESD in a Higher Degree in Education. A Case Study. *Sustainability*, 12(1), 222. <https://doi.org/10.3390/su12010222>
- Valdes-Ramirez, D., De Armas Jacomino, L., Monroy, R., & Zavala, G. (2024). Assessing sustainability competencies in contemporary STEM higher education: A data-driven analysis at Tecnológico de Monterrey. *Frontiers in Education*, 9, Article 1415755. <https://doi.org/10.3389/educ.2024.1415755>
- Vesala-Varttala, T., Humala, I., Isacson, A., Salonen, A.O., & Nyberg, C. (2021). Fostering Sustainability Competencies and Ethical Thinking in Higher Education: Case Sustainable Chocolate. In *Research*. (<http://urn.fi/URN:NBN:fi-fe2021120759474>).
- Wiek, A., Bernstein, M., Foley, R., Cohen, M., Forrest, N., Kuzdas, C., Kay, & Keeler, L. W. (2016). operationalising competencies in higher education for sustainable development. In M. Barth, G. Michelsen, I. Thomas, & M. Rieckmann (Eds.), *Routledge Handbook of Higher Education for Sustainable Development* (pp. 241–260). Routledge.
- Wiek, A., Withycombe, L., & Redman, C. L. (2011). Key competencies in sustainability: A reference framework for academic program development. *Sustainability Science*, 6(2), 203–218. <https://doi.org/10.1007/s11625-011-0132-6>
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-29044-2>
- Xiao, C., Wan, K., & Kwong, T. (2024). Developing and validating a scale to measure tertiary students' sustainability competence in extracurricular settings in Asia. *Cogent Education*, 11(1), Article 2334107. <https://doi.org/10.1080/2331186X.2024.2334107>
- Yılmaz, F. N. (2024). Comparing the reliability of performance task scores obtained from rating scale and analytic rubric using the generalizability theory. *Studies in Educational Evaluation*, 83, Article 101413. <https://doi.org/10.1016/j.stueduc.2024.101413>

Further reading

- Sapsford, R. (1999). *Survey design*. London: Sage.
- Sinakou, E., & Daskolia, M. (n.d.). Key Sustainability Competences of University Graduates: A Scoping Literature Review. *International Journal of Sustainability in Higher Education*, under review.