Developing and validating a scale to measure tertiary students' sustainability competence in extracurricular settings in Asia

Chong Xiao, Kelvin Wan & Theresa Kwong

View supplementary material

Published online: 05 Apr 2024.

Submit your article to this journal

Article views: 907

View related articles

View Crossmark data

cogent

EDUCATIONAL ASSESSMENT & EVALUATION | RESEARCH ARTICLE   👌 OPEN ACCESS   Check for updates

Developing and validating a scale to measure tertiary students' sustainability competence in extracurricular settings in Asia

Chong Xiao[a]* 🔟, Kelvin Wan[b]* 🔟 and Theresa Kwong[c] 🔟

[a]Department of Applied Social Sciences, The Hong Kong Polytechnic University; [b]Centre for Teaching and Learning, The Hang Seng University of Hong Kong; [c]Centre for Holistic Teaching and Learning, Hong Kong Baptist University

ABSTRACT

This study aims to develop and validate a comprehensive scale for assessing students' sustainability competence in education for sustainable development (ESD). Existing ESD assessment tools have faced criticism for lacking rigorous design and validation. Following a three-phase procedure proposed by Boateng et al., the study utilized qualitative and quantitative methods to develop and validate the scale. The scale was administered to 708 students participating in an extracurricular eTournament on Sustainable Development Goals organized by a Hong Kong university. The validation process involved expert panels, cognitive interviews, factor analyses, and assessments of internal consistency. Using Wiek et al.'s five-competence framework, 16 ESD learning outcomes were generated, forming the basis for the scale items. The survey included a mini-scenario introduction to enable students' evaluation of sustainability practices, even with limited experience. Data analysis revealed a four-factor structure that could be both distinguished and integrated under a single upper-level factor. The scale is developed with the aim to better measuring tertiary students' sustainability competence via a holistic and integrated approach, which completes the constructive alignment of current ESD, particularly in extracurricular settings.

Introduction

Higher education is considered the engine of human society's sustainable development, as it nurtures a capable labor force and responsible citizens (Waltner et al., 2019). When universities and colleges explore pathways to sustainable development, emerging courses and programs define ESD as educational initiatives that encourage changes in knowledge, skills, values, and attitudes to help students think and behave in ways fostering a more sustainable and just society for all (Blessinger et al., 2018). Since the 1990s, ESD has been implemented across macro (e.g. national) to micro (e.g. classroom) levels towards a more generic, holistic, and integrated paradigm, which emphasizes competencies cultivation and demands better design and delivery.

Extant ESD literature endeavored to define educational competencies and conceptualize relevant pedagogies. There has been increasing attention on measuring the effectiveness of ESD, despite the tool development is still in its infancy – mostly with vague or ill-defined learning objectives and a lack of connections across studies or measured competencies (Redman et al., 2021). To address these insufficiencies, the present research develops and validates a scaled self-assessment tool to measure students'

sustainability competence, drawing on the strengths and weaknesses of existing ESD assessment instruments. The goal is to provide a holistic and integrated approach to evaluating the effectiveness of ESD, particularly in extracurricular settings. This also aims to enhance the ongoing advocacy for evidence-based evaluation of various ESD programs by strengthening the alignment between learning outcomes, teaching and learning methods, and assessment methods.

This paper starts with a literature review of the competency-based approach and constructive alignment of ESD. The data and methodology section outlines the three-phase process of nine steps proposed for scale development and validation. Finally, the results section discusses the content validity, construct validity, and confirmatory factor analysis of the Sustainability Competence Scale, leading to conclusions and implications for future research.

Literature review

Since the first action plan, *Agenda 21*, identified education as essential for achieving sustainability, various ESD initiatives at the tertiary level have been tremendously boosting. Due to the COVID-19, the global higher education sector has had to radically adjust their strategies in building sustainable institutions, which are typically linked with UN's blueprints (e.g. SDGs), to make them more adaptable to the post-pandemic new normal of tertiary teaching and learning (Crawford & Cifuentes-Faura, 2022). During this transformation, the conventional content-based approach, caring about what sustainability-related knowledge is learned (e.g. Cifuentes-Faura & Noguera-Méndez, 2023; Lafuente-Lechuga et al., 2024; Waltner et al., 2019), and has gradually extended to a paradigm centered around attributes (Blessinger et al., 2018). The focus has shifted towards the growth of students' competencies, which encompass a complex combination of knowledge, skills, values, attitudes, and desire, leading to effective and embodied human action for specific goals (Crick, 2008; Weinert, 2001). This section includes two sub-sections. Literature on competency-based approaches to ESD at the tertiary level was firstly reviewed, emphasizing the five-competence framework proposed by Wiek et al. (2011). It then addresses insufficiencies in constructive alignment of ESD, highlighting the need for improved assessment tools.

A competency-based approach to ESD: from generic and holistic to integrated

Competencies concerned in ESD evolved from isolated generic competencies measured by large-scale benchmarking assessments to those tailored for sustainability sciences (education), with their comprehensiveness and integratedness increasingly emphasized. This section reviews the evolution of competency-based approach to ESD. It later shines the spotlight on a five-competence framework in sustainability by Wiek et al. (2011), underpinning our developing extracurricular ESD's learning outcomes, scale subconstructs, and thereof items.

Competencies from large-scale benchmarking assessments have long been adopted to frame and guide ESD. Taking the Project of *Definition and Selection of Competencies* as an instance, OECD stated that sustainable development, as the common values anchor of human societies, depends critically on the competencies related to 'economic and social benefits', applicable 'in a wide spectrum of contexts', and aspired by 'all individuals' (OECD, 2005, p. 7). Accordingly, transferable and generic competencies were identified as the key to realizing individuals' and communities' sustainability, which embrace three-fold – acting autonomously within a collective and diachronic context, relating to and cooperating with heterogeneous others, and using tools (language, information, technology) interactively. These key generic competencies are rather abstract, allowing distinct interpretations according to different initiatives' objectives and features. Operationally, their cognitive aspects related to knowledge acquisition and information processing are easier to capture and, therefore, widely adopted (e.g. Taimur & Sattar, 2019). An increasing voice of adjusting ESD dominated by cognitive dispositions emerged, advocating including non-cognitive (values, attitudes, and affects) and behavioral dispositions to meet complex and holistic demands of sustainable development (Ridgeway, 2001).

Given many cognitive competencies substantially influence behaviors through emotion or desire (Fischer & Barth, 2014), a sound ESD should roundly address 'What students know' about sustainability, 'what skills they have to put this knowledge to use', and, more importantly, 'what they choose to do with the knowledge and skills' when encountering sustainability issues (Shephard et al., 2015, p. 857).

Accordingly, Fischer and Barth (2014) referred to the *Definition and Selection of Competencies* project's threefold competency categorization and defined a comprehensive list of seven generic competencies enabling sustainable consumption. Specifically, for the first category, acting autonomously, competencies to reflect on individual needs and cultural orientations, and to plan, implement, and evaluate consumption-related activities are essential for individuals. Regarding the second category, interacting in heterogeneous groups, individuals must critically act as active stakeholders with the agency for change, necessitating the ability to communicate sustainability ideas. For the third category of interactively using tools, individuals need to be able to use, edit, and share different forms of knowledge effectively and reliably, which nowadays is significantly facilitated by information technologies. Besides, future-oriented and interrelated thinking play critical roles in shaping future sustainability. In this work, Fischer and Barth (2014) emphasized the generic and holistic nature of the competency-based approach to ESD – transcending the boundaries of meeting domain-specific requirements and facilitating individuals' engagement with more fundamental values and overall opportunities to live a sustainable life. However, it leaves those competencies parallel and disconnected, which did not address an earlier criticism on the "laundry list" of sustainability competencies without conceptually embedded interlinks (Wiek et al., 2011, p. 205). Therefore, more endeavors are imperative to examine competencies' interconnection and convergent effects on sustainability practices.

With the efforts to encompass values, cognition, and actions, the European Union has developed a framework of sustainability competence known as GreenComp in 2022. This framework focuses on four key areas: embodying sustainability values, embracing complexity in sustainability, envisioning sustainable futures, and acting for sustainability (European Commission, Joint Research Centre, 2022). Among the twelve competences proposed, while GreenComp includes competencies specifically related to sustainability, such as promoting nature and futures literacy, as well as conventional generic competencies like critical thinking and exploratory thinking. However, this framework was primarily structured under the concept of 'learning for environmental sustainability' (European Commission, Joint Research Centre, 2022, p. 13). This focus limits its applicability to educational programs that target sustainability challenges more related to economic or social aspects.

In order to tackle ill-defined and complicated sustainability issues, emerging sustainability initiatives expect students to grow into future problem solvers, change agents, and transition managers (Loorbach & Rotmans, 2006; McArthur & Sachs, 2009; Orr, 2002; Rowe, 2007; Willard et al., 2010). In response to this call for problem-driven and solution-oriented ESD, Wiek et al. (2011) proposed an overarching competence – Sustainability Research and Problem-solving, which integrates five competencies not focused on by traditional education but specially attended to in ESD (School of Sustainability at Arizona State University, 2018).

The first competence, *Systems Thinking*, refers to the ability or skills to perform problem-solving in complex sustainability systems comprising multiple domains (e.g. economic-environmental-societal, local-global). Second, *Future Thinking/anticipatory* competence enables people to recognize how sustainability problems evolve and craft rich pictures of future sustainability situations in an uncertain and ambiguous world. *Values Thinking/normative* competence means the ability to collectively specify, apply, reconcile, and negotiate sustainability values, principles, and goals. *Strategic Thinking,* as the most action-oriented competence, demonstrates the ability to design and implement systemic interventions and transformational actions towards sustainability. The last competence, cutting across the previous four and linking sustainability competence with other generic competencies, is *Interpersonal/ Collaboration* competence, which refers to the ability to motivate, enable, and facilitate collaborative and participatory problem-solving towards sustainability.

Wiek et al.'s five-competence framework contributes to the field of ESD by covering a comprehensive list of generic competencies and, more significantly, emphasizing the interconnected and co-dependent nature of the sustainability competence system (Fischer & Barth, 2014). Inherently, each competence's cognitive and non-cognitive dispositions co-dependently define and differentiate it from other 'regular' academic and basic competencies. Interrelatedly, the composition of five competencies helps understand and respond to different sustainability challenges via diachronic (past, present, and future), systemic (parts-whole), multidimensional (self-others and individual-collective), and cross-domain (local-global, economic-environmental-societal, and commitment-action) interplays (Lambrechts, 2019; Wiek et al.,

2011). This generic, holistic, and connected framework has been widely applied to design ESD at the tertiary level, making it suitable for guiding our constructing the intended learning outcomes and scale development. While extant literature mainly applies this framework in formal curricula (e.g. School of Sustainability at Arizona State University, 2018; Wiek et al., 2011) or co-curricular programs (e.g. Savage et al., 2015), the present study particularly examines its applicability in an extracurricular setting.

Insufficiencies in the constructive alignment of ESD

While these sustainability competencies were embodied in ESD learning outcomes, referring to different programs' positioning and objectives (e.g. Brandt et al., 2019), and various pedagogical innovations have been conceptualized for sharing and generalization (e.g. Barth & Burandt, 2013), the practice and research about assessment methods are still underdeveloped (Grosseck et al., 2019). According to the principle of 'constructive alignment' by Biggs and Tang (2011), teaching, led by clearly stated outcomes and delivered through well-designed activities, also requires valid assessment tasks to accurately judge how well outcomes have been attained. In this sense, assessment of learning outcomes is perceived as a 'gap' in constructively aligned ESD (Redman et al., 2021, p.118), Valid tools are urgently needed to ensure that ESD is aligned with the desired direction of student development (Fischer & Barth, 2014).

In response to the need for optimized ESD assessment, Redman et al. (2021) critically reviewed 75 studies to examine the instruments in use. They proposed a typology of eight tool types and summarized their key features, strengths, and weaknesses. This review also identified five common flaws in existing ESD assessment tools, namely (1) an 'apparent afterthought' tendency (p. 127), (2) lacking agreement on learning outcomes, (3) little linkage across research efforts, (4) insufficient consideration on the choice of tool per se, and (5) stopping at actual tool development. Rather than being completely separated, these aspects are mutually influenced. These flaws limit the applicability and compatibility of these tools with other instruments used in contemporary sustainability and its education; however, these insufficiencies also direct our research endeavors to develop a valid assessment tool for ESD.

This study aims to develop a valid assessment tool for ESD by creating a self-assessed Sustainability Competence Scale for tertiary students from post-secondary institutions, including universities (undergraduate and postgraduate), colleges, polytechnic institutes, and vocational schools. The self-assessment scale was chosen for its convenience in administration and data collection for statistical analysis (Redman et al., 2021). Nevertheless, it cannot be denied that students may interpret the prompt and concerned competencies differently (Cebrián et al., 2019). They were also suspected to be unable to rate their capability in sustainability activities they seldom practice (Holdsworth et al., 2018). Taking these constraints into account, we aim to answer four research questions in this study of scale development and validation:

1. What are the learning outcomes of ESD corresponding to the educational competencies of sustainability at the tertiary level?
2. How does the self-assessment scale enable tertiary students' explicit sense-making and conscious rating?
3. Whether the items fairly represent and relate to their respective competence and show an integrated five-factor structure?
4. Does the scale show good internal consistency?

Data and methodology

This study obtained ethical approval from the Institutional Ethics Committee of the university. Before engaging in the eTournament, participants were presented with an online informed consent form delineating study objectives, procedures, and rights. Emphasizing voluntariness, participants provided demographic information upon enrollment. The informed consent form was prominently displayed on the initial page of each survey, underscoring the study's dedication to participant welfare, confidentiality, and ethical research conduct. This approach ensured transparency and affirmed participants' autonomy in alignment with ethical standards.
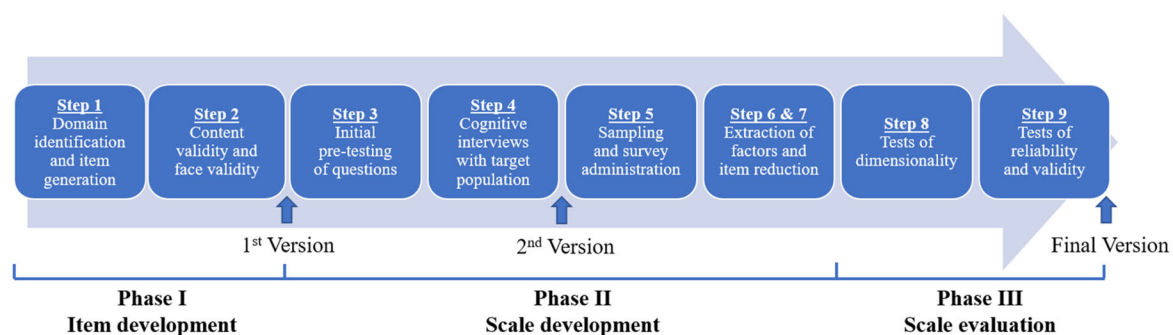
Figure 1. Flow chart of scale creation procedure. Note. Adapted from Boateng et al. (2018).

The scale development and validation generally follows the three-phase process of nine steps proposed by Boateng et al. (2018): (I) item development, (II) scale development, and (III) scale evaluation, with some steps combined and minor order change (Figure 1).

Phase I. Item development

Step 1. Identification of domain and item generation

As the basis of this scale, Wiek et al.'s (2011) five-competence framework was adopted. Learning outcomes were firstly developed based on each competence's definition and concerned variables/concepts. To connect this scale with extant assessment studies, as Redman et al. (2021) suggested, most items were generated from existing validated tools via a deductive method (Hinkin, 1995), with a small portion of self-created ones (Table 1). Then, a preliminary 32-item Sustainability Competence Scale was drafted. As shown in Table 1, the item number of the referenced scales. as well as the self-created items, were included in the preliminary 32-item Sustainability Competence Scale.

Specifically, *Systems Thinking* competence addresses part-whole and part-part relationships of the complex socio-ecological and technological system (Remington-Doucette et al., 2013). Students with System Thinking competence are supposed to be able to:

LO1. identify the elements, actions, or sub-systems comprising the coupled sustainability system;

LO2. explain the part-whole interdependence and part-part interconnections; and

LO3. analyze the dynamics of the interconnections and interdependence in the sustainability system.

Accordingly, Moore et al.'s Scale of System Thinking (2011) was adopted. Despite being initially developed for continuous quality improvement in healthcare management, Moore et al.'s definition and theoretical dimensions of System Thinking were general enough to be transferred into sustainability. Therefore, seven items were included with modifications to measure students' ability related to concepts of systems, problem ontologies, cause-effect chains, tipping effects, cascading effects, coupled domains and sub-system, inertia, feedback, and so on (School of Sustainability at Arizona State University, 2018).

Future Thinking competence refers to the cognitive process of thinking about the future, which addresses underlying concepts such as temporal terms, consistency and coherence, uncertainty and epistemic status (possibility, probability, plausibility), path dependency and non-intervention features, quality criteria of visions, and risk and precaution (Wiek et al., 2011). Students with Future Thinking are able to:

LO4. explain the current situation of a given sustainability system;

LO5. synthesize the trends from the past to present and plausible pathways to different futures;

LO6. appraise relevant drivers and possible unintended consequences; and

LO7. depict a desired future of sustainability with justification.

Future thinking involves both semantic and episodic memories; for the latter, according to the practices in psychometrics, an essential indicator is specific autobiographical memory (Hallford et al., 2020); therefore, Siew and Rahman (2019)'s scenario-based sustainability-related Future Thinking Written Test

Table 1. Five competencies, learning outcomes, item sources, and samples from the first version.

Competence	LO	Item Generated & Source***		Item No	Sample
System Thinking (ST)	LO_1	3, 6	Systems Thinking Scale (Moore et al., 2011)	7	ST01: I think it is crucial to understand how a situation happens as a result of a series of actions.
	LO_2	11, 14, 15, 18			
	LO_3	20			
Future Thinking (FT)	LO_4	1	Future Thinking Test (Siew & Rahman, 2019)	6	FT05: I can picture a preferable future of sustainability topic.
	LO_5	2, 4			
	LO_6	3			
	LO_7	5			
Values Thinking (VT)	LO_8	Two items by authors*, 14	Sustainable Entrepreneurship Framework (Ploum et al., 2017)	6	VT02: I understand the human-made system should not rule over the natural system.
	LO_9	13, 15			
	LO_10	16			
Strategic Thinking (StT)	LO_11	21, 22	Sustainable Entrepreneurship Framework (Ploum et al., 2017)	6	StT01: I know which stakeholders need to involve.
	LO_12	23			
	LO_13	24, Two items by authors**			
Interpersonal (INT)	LO_14	8, 11, 21	Interpersonal reactivity Index (Davis, 1983)	7	INT03: I try to understand my partners better by imagining how things look from their perspectives.
	LO_15	9, 18, 20	Environnemental Communication Scale (Kassing et al., 2010)		
	LO_16	3	Empowering Leadership Questionnaire (Arnold et al., 2000)		

Note. *Based on the concepts of critical thinking and normative competence in sustainability science education (Komasinski & Ishimura, 2017); **based on the competency model for assessing strategic thinking – concept of Alignment (Goldman and Scott, 2016).

***Please note that the item numbers in Table 1 correspond to the original items from the respective reference scales, and do not reflect the item numbers of our initial survey.

was adopted to create our scale's second subconstruct. To make it compatible with the entire self-assessed questionnaire, the original eleven open-ended tasks were modified and combined into six self-assessed items.

For *Values Thinking* competence, taking relevant normative knowledge as prerequisites, it focuses on the concepts such as trade-offs and 'win-wins' synergies, social-ecological integrity, etc. Accordingly, students equipped with Values Thinking are able to:

LO8. judge the sustainability status of the concerned system with notions of justice, ethics, equity, and socio-ecological integrity;

LO9. apply the norms, values, and principles of sustainability in daily practices; and

LO10. explain sustainability decisions from normative and ethical perspectives.

In this subconstruct, two sources were referred to for item generation: one is the Sustainable Entrepreneurship Framework by Ploum et al. (2017), which was developed using data from higher education and largely aligned with the present study's perception of tertiary students as potential sustainability change agents. Four items addressing normative competence were included, broadening the scope from the commercial sector to more generic ones. In addition, the authors created two items measuring the mastery of relevant normative notions and the relationship between natural and human-made systems.

In the fourth subconstruct of *Strategic Thinking* competence, sustainability is oriented to solving 'wick problems' and 'getting things done' (Wiek et al., 2011, p. 210). Strategic Thinking enables students to:

LO11. devise practical intervention towards sustainability, including milestones and goals, procedural steps, hurdles, and resources;

LO12. execute the strategies by motivating and giving full play to pertinent stakeholders; and

LO13. reflect on the multidimensional, current-future, and sectoral-overall alignments for further optimization.

Accordingly, the four-item subscale of strategic management and action in the Sustainable Entrepreneurship Framework (Ploum et al., 2017) was adopted, concerning concepts like network and

alliance, carriers and barriers, systemic intervention, transformative governance, intentionality, and effectiveness and efficacy. Besides, based on Goldman's (2016) Behavioral Descriptions of Strategic Thinking and the highlighted concept of 'alignment', two items were created to address the latent notions of action programs, success factors, social learning, and adaptation and mitigation.

The last subscale, *Interpersonal* competence, measures the function of negotiation and solidarity and leadership (attributes, styles, and techniques) in sustainability. Students with Interpersonal competence are able to:

LO14. interpret different stakeholders' diverse opinions on concerned sustainability practice;

LO15. facilitate stakeholders' exchange of views, particularly in disagreement or conflict; and

LO16. integrate the plural perspectives in making sustainability decisions.

Since no instruments developed specifically for measuring Interpersonal competence in sustainability, three scales were referred to. First, three items from the Environmental Communication Scale (Kassing et al., 2010), initially validated to address tertiary students' communication function and participatory engagement (negotiation, mediation, and deliberation), were adopted with the applicability broadened to overall sustainability. Second, Interpersonal Reactivity Index (Davis, 1983), validated using undergraduate data, was introduced targeting the concepts of pluralism, empathy, and ethnocentrism in sustainability practices. Three items from its perspective-taking subconstruct (reverse items excluded) were adopted to measure the tendency to spontaneously adopt others' viewpoints; Last, targeting participative decision-making in sustainability, one item from the Empower Leadership Questionnaire (Arnold et al., 2000) was adopted to assess students' behavior and techniques in sustainable leadership.

Step 2 - content validity and face validity

After drafting the items, an expert panel was assembled to examine each item's content and face validity for content relevancy, representativeness, and readability. Four experts, PhDs from humanities, social sciences, and engineering with prior academic experience in competence evaluation and ESD, were invited to assess each item's relevance to Sustainability Research and Problem-solving by rating (1 means least relevant, 5 means most relevant) and suggest modifications.

A total of five-domain 32 items were developed based on the theoretical framework, referenced validated scales, and comments from the expert panel. Table 1 shows the total item numbers of the first version and one sample item from each subscale. To keep coherence across (sub)scales, we asked respondents to report using a Likert scale for all measures. Concerning cultural effects that since East Asians tend to rate the mid-point (Chen et al., 1995) and interpret mixed words (both positive and negative words) differently in earlier cross-cultural psychometrics studies, which resulted in relatively low factor loadings (Wong et al., 2003), the authors decided to remove neutral options and reverse items. A 6-point Likert scale was decided, with 1 representing the least and 6 indicating the most agreement with the items.

Phase II – scale development

Step 3 – initial pre-testing

The first version of the Sustainability Competence Scale was distributed via *Qualtrics* for data collection during the first five days of an SDG eTournament. 708 students mostly studying in Asia responded, as depicted in Table 2. Despite no absolute consensus on a recommended sample size for factor analysis (Boateng et al., 2018), most literature suggested 300 was a minimum acceptable size for validation. Therefore, the 708 respondents of the initial pre-testing were sufficient for this step. An Exploratory Factor Analysis (EFA) was carried out to extract factors.

Step 4 – cognitive interviews

Based on the initial pre-testing results, the authors decided to conduct cognitive interviews with four respondents. The technique of verbal probing was adopted to keep the interviewees on the track during

Table 2. Respondents' demographics in the initial pre-testing.

	N	%
Gender		
Female	479	67.7%
Male	229	32.3%
Regions of the institutions		
Asia	690	97.5%
Non-Asia	18	2.5%
Level of study		
Sub-degree	29	4.1%
Undergraduate	587	82.9%
Postgraduate	92	13.0%
Disciplines		
Business	275	38.8%
Health and Social Sciences	167	23.6%
STEM	137	19.4%
Humanities	83	11.7%
Interdisciplinary	32	4.5%
Others	14	2.0%

the interview and dig out sufficient information about their cognitive process (Priede et al., 2013). Respondents were invited to have face-to-face interviews individually. Probing questions on how to interpret the abstract terms, search memories for relevant information, estimate responses, and provide answers in the requested format were asked, corresponding to the four cognitive stages of comprehension, retrieval, judgment, and response (Willis, 2005). Each interview lasted around 1.5 hours. At post-interview design meetings, the authors discussed the interview findings in detail and reached a consensus on item revisions to produce a second version Sustainability Competence Scale.

Step 5 – survey administration and sample size
For the second test, the second version scale was distributed to students who completed the eTournament. Data were collected from 381 respondents through the identical mechanism to the first trial, except ten days were reserved.

Steps 6 to 7 – extraction of factors and item reduction
The IBM SPSS 27 software was used to analyze the data in these two steps. EFA was used to reveal the underlying structure using the two waves of data. The purpose of the first wave EFA was to explore a five-factor structure; however, the results were unexpected. The second wave EFA attempted to assess whether the structure would change after revisions led by cognitive interview findings. It also aimed to determine the optimal number of factors that suited a collection of items. For this purpose, various analytical approaches, including Kaiser–Meyer–Olkin measure of sampling adequacy, extraction methods, and types of rotations, were conducted to inform authors' decisions. Both waves of data used the same configuration.

According to results of the second wave EFA, inter-item correlation table, and Cronbach's alpha (α), items would be removed if it (1) cross-loads to more than one factor with its factor loading $\geq.40$; or (2) only appears in a single factor with its factor loading $<.40$; or (3) demonstrates a weak inter-item correlation.

Phase III – scale evaluation

The IBM SPSS and Amos Graphics 27 software were used to analyze the data in this phase. It included the EFA, CFA, and testing for reliability and validity.

Step 8 – tests of dimensionality
Following item reduction, CFA was conducted using the second wave data to confirm the factor structure. After identifying the underlying relationships between measured variables, it was also applied to determine the dimensionality of a set of estimated parameters. To test the fitness of the data to the hypothesized measurement model, fit indexes, including Root Mean Square Error of Approximation

Table 3. Analysis methods for scale validity and reliability.

Indicator	Analysis method
Face Validity	Expert judges
Content Validity	Expert judges; Cognitive interview with respondents
Construct Validity	CFA; EFA
Internal Consistency Reliability	α; CR
Convergent Validity	CR; AVE
Discriminant Validity	AVE; MSV

(RMSEA), Comparative Fit Index (CFI), and Standardized Root Mean Square Residual (SRMR), were reported.

Step 9 – tests of reliability and validity

Scale scores for the reliability and validity tests were created using finalized items from the dimensionality tests. The average of items in each subscale was used to calculate the scale scores. α and composite reliability (CR) were adopted to assess the scales' internal consistency. To evaluate discriminant and convergent validity, each scale would be subjected to CFA to determine the CR, average variance extracted (AVE), and maximum shared variance (MSV). Table 3 summarizes the indicators and analysis methods for evaluating the scale reliability and validity in this study.

Results

Content validity

Expert judges affirmed that the items were relevant to and representative of the domain of Sustainability Research and Problem-solving. Regarding the rating, four experts gave all 32 items with at least 3/5; therefore, the authors decided to keep all. Ten items rated with 3 were further improved by three means. First, three items had the issue of measuring more than one factor within a statement, asking students' understanding of the notions per se and relation/interaction thereof (e.g. natural and human-made systems and their ruling-over interplay). The original statements were modified by keeping the latter as the variable to be measured, necessitating students' comprehension of the terminologies and higher-order thinking as the prerequisites. The second way to improve the content validity was using explicit and consistent wording. Four items were developed with vague scope/subjects, using phrases like 'concerned system' or 'relevant topic'. This issue was intended to be addressed by adding 'sustainability' to clarify what is concerned or related; however, it was proven insufficient by cognitive interview findings (will discuss more later). The last adjustment was to simplify the statement of three items using complicated wording or redundant semantic phrases with plain language.

The next step was to analyze the data from the initial pre-testing. The EFA results indicated four potential factor structures, one less than the five proposed in the original framework. Items of Values Thinking and Strategic Thinking were loaded into a single factor (seen the section of Construct Validity). The authors considered students' incapability to understand and differentiate terminologies in the statement as possible reasons for the two factors' convergence. Therefore, we decided to have cognitive interviews with the target population.

According to the responses from four invited interviewees, six aspects of the original items were modified. The first was to add a scenario-based lead-in to help respondents contextualize sustainability challenges. All four interviewees asked questions such as 'what topic/project/issue' even though 'sustainability' was added as suggested by the expert judges. The context of sustainability problem-solving is quite vague given students did not have much experience, hindering their judgment. Since this scale is designed to measure tertiary students' competence, taking their cognition and experience into account, the scenario was framed as a university project focusing on SDGs aimed to positively impact the locality. The second adjustment was to make the illustrative prompt consistent for each subscale. During the interviews, students asked for clarification on whether they were supposed to evaluate their behaviors (i.e. actually consider/recognize/keep in mind or not) or the arguments about sustainability. Therefore, each subscale was reviewed to confirm the subject of measurement, to which the same question prompt was applied to keep respondents focusing on the distinct statement of each item as

the concern. The third modification was to provide examples/explanations to those abstract notions/terms. The fourth change was to adjust the word order within some statements or reorder some items according to respondents' normal cognitive logic (e.g. causal sequence, affirmation rather than double negation); The fifth adjustment was to specify the scope, aspect, means, or quantity to facilitate respondents to make judgments. The last improvement was to change the first version scale, with the first and last options labeled, into fully-labeled ones. Students reported it was challenging to differentiate the middle four options; as suggested by DeCastellarnau (2018), clarifying the labels made it easier to make choices and positively impacts reliability. With these modifications, a second version scale was produced.

Construct validity

Exploratory factor analysis

Two waves of EFA were conducted by using the Principal Axis method of extraction. Bartlett's test of sphericity was significant (χ^2 (496) = 14954.27, $p < .001$ in wave one; χ^2 (496) = 8811.49, $p < .001$ in wave two). The Kaiser–Meyer–Olkin measure of sampling adequacy indicated that the strength of the relationships among variables in both waves of data was high (.96). These tests indicate that the two waves' data were appropriate for EFA. Initially, four-factor with eigenvalues greater than one were extruded. A Promax rotation was performed since factors were expected to be correlated. Table 4 displays the pattern matrix of waves one and two with only factor loadings greater than .40 presented. As shown in Table 5, with a high eigenvalue of 14.35 in wave one and 15.76 in wave two, the first factor was robust. In two waves, the four factors together account for at least 60% of the total variance, which is an acceptable explained variance (Watson, 2017). Nevertheless, even after the revisions informed by cognitive interviews, the second EFA results showed that items of Values Thinking and Strategic Thinking had a similar pattern that remained into a single factor.

Table 4. Pattern matrix for SC scale by using data and items from two waves.

| | Wave one | | | | | Wave two | | | |
| | Factor loading | | | | | Factor loading | | | |
Item	1	2	3	4	Item	1	2	3	4
StT02	.88				StT03	.91			
StT05	.79				StT02	.90			
StT06	.78				StT01	.78			
StT04	.77				StT04	.73			
StT03	.75				VT05	.57			
StT01	.75				VT06	.56			
VT05	.69				VT03	.51			
VT06	.66				StT05	.50			
VT03	.61				StT06	.50			
VT04	.50				VT04	.45			
VT01	.40				INT05	.40			
INT02		.81			INT03		.86		
INT05		.79			INT07		.78		
INT03		.77			INT01		.77		
INT04		.75			INT02		.77		
INT07		.75			INT04		.73		
INT06		.71			INT06		.63		
INT01		.69			VT02				
VT02					ST02			.79	
ST05			.77		ST03			.78	
ST03			.75		ST06			.72	
ST02			.75		ST05			.71	
ST01			.74		ST01			.60	
ST06			.64		ST04			.60	
ST04			.62		ST07			.55	
ST07			.62		FT02				.94
FT03				.73	FT01				.80
FT02				.69	FT04				.76
FT01				.68	FT03				.75
FT04				.48	FT06				.61
FT05				.43	FT05				.60
FT06				.40	VT01				

Table 5. Eigenvalues, percent of variance, and α of two waves' data.

	Wave one				Wave two			
	1	2	3	4	1	2	3	4
Eigenvalues	14.35	2.76	1.55	1.04	15.76	1.99	1.51	1.20
Variance (%)	44.86	8.63	4.85	3.24	49.26	6.21	4.72	3.74
α	.93	.91	.87	.89	.94	.91	.87	.91

Overall, the EFA revealed that most items had loaded to the respective hypothesized factor, with only three items (VT01, VT02, and INT05 in Table 4) failing to load in any factor or having a small factor loading, and, therefore, were removed. The authors combined the remaining items from Values Thinking (VT03-06) and Strategic Thinking (StT01-06) and renamed the new subscale as 'Value-driven Strategic Thinking competence' (VdStT). For following CFA, these four factors were considered subscales of the Sustainability Competence Scale.

Confirmatory factor analysis

To confirm the four-factor structure suggested by two-wave EFA results, a CFA for First-Order model was conducted by using the second wave data. The current study did not have any missing data and used the Standard Maximum Likelihood Estimation for CFA. Figure 2 illustrates the final diagram with significant coefficients. Upon fitting the study data, the results of the CFA exhibited an excellent fit – all the regression models were significant ($p < .01$, $t > 2.33$) followed by the criteria proposed by Kline (1998).

The First-Order CFA results inform an interdependent relationship among these factors, with a possibility that each factor is partially weighing a higher-level latent factor – Sustainability Research and Problem-solving competence. Therefore, another CFA for Second-Order model was carried out. Figure 3 depicts the Second-Order CFA results, whereas Table 6 compares two model fit-indices. Most of the model fit-indices did not show differences, and the χ^2 value for the first model is only slightly smaller than the second one. A Chi-square difference test (CDT) then verified that this difference is not significant ($\chi^2 (2) = 6.02$, $p = .05$) in terms of their fit.

Final version of the scale

After the steps of factors extraction, item reduction, and tests of dimensionality, a 29-item scale was finalized (Supplementary Appendix A1).

Internal consistency reliability, convergent and discriminant validity

Since no difference observed between the First and Second-Order models, the following validity and reliability testing were conducted by using the First-Order model for convenience. The CR and α indicated that the scale was reliable ($>.70$). The convergent and divergent validity testing results show acceptable CR and AVE. Convergent validity presented in the study as the CR for the construct was greater than .70 and the AVE was greater than .50 (Fornell & Larcker, 1981). However, the AVE values of the constructs were lower than the MSV, raising concerns about discriminant validity (Table 7).

While earlier work recommended multiple methods must be employed in the (discriminant) validation process (Rönkkö & Cho, 2022), this study adopted two alternatives to further test the discriminant validity. One was inter-factor correlation (IFC). As presented in Table 8, the final version scale showed that the factors were positively correlated ($r = .58$ to $.78$; $p < .0001$) in acceptable range ($r < .85$). The other angle was to perform a CDT (Zait & Bertea, 2011). We examined two models (Table 9) – one indicates zero factor correlation while another indicates the free correlation between factors. The difference in test results was significant ($\chi^2 (6) = 920.06$, $p < .001$), indicating these factors present discriminant validity. Therefore, based on the IFC and CDT results, we argue the final version scale demonstrates satisfactory discriminant validity, despite the concerns raised according to Fornell and Larcker's criteria (1981).
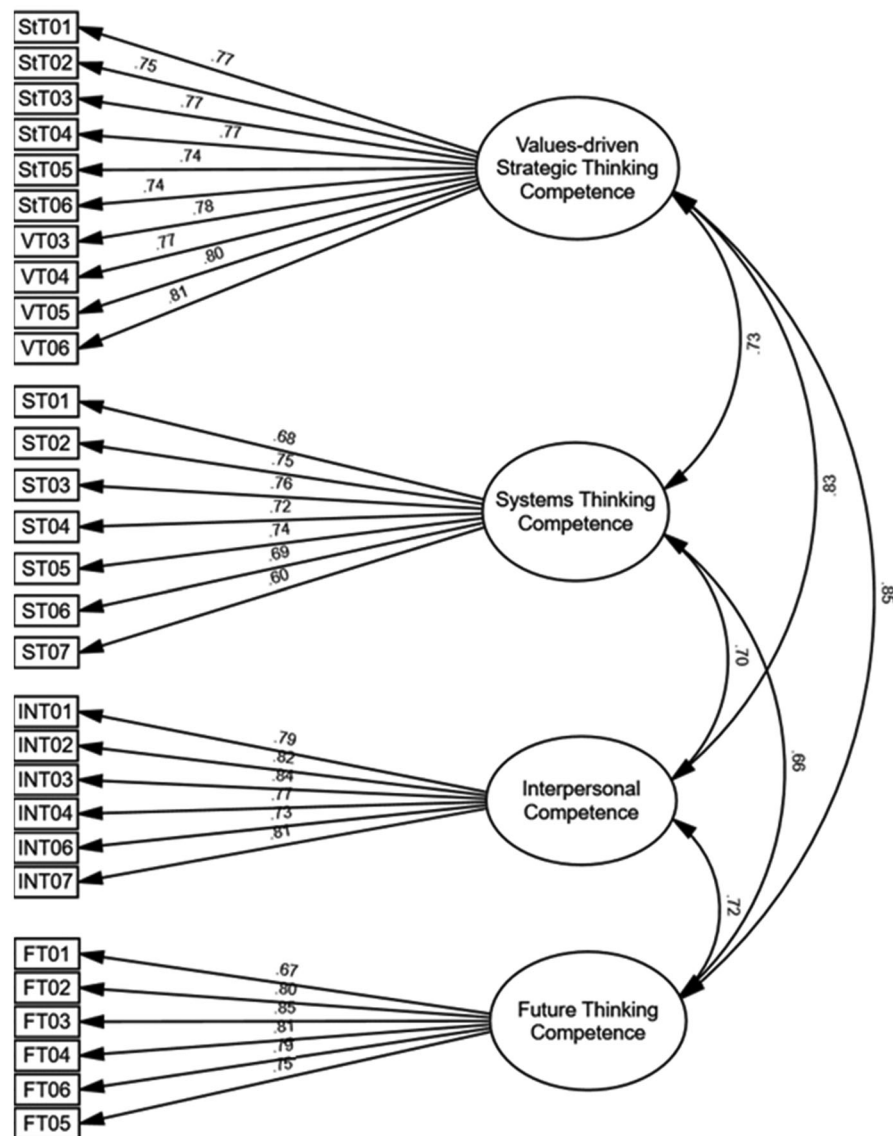
Figure 2. CFA for the four-factor structure of the Sustainability Competence Scale.

Discussion and conclusions

This study developed and validated a scale measuring the sustainability competence of tertiary students, based on the widely-adopted framework by Wiek et al. (2011). The five competencies of concern, namely Systems Thinking, Future Thinking, Values Thinking, Strategic Thinking, and Interpersonal Competence, were refined and canvassed by referring to the extant sustainability literature and contextualized in the field of ESD. Adopting a competency-based approach, 16 intended learning outcomes of ESD were developed to operationalize the assessment of tertiary students' sustainability competence in an extracurricular setting. Accordingly, most items of this scale were extracted from existing validated tools with modifications; other items were self-drafted if no suitable scales were available to address specific learning outcomes. Expert judges confirmed the relevance of these items in addressing sustainability challenges.

This Sustainability Competence Scale was created with a particular intention to enable explicit sensemaking and conscious rating of tertiary students on sustainability practices, even if they have limited experience in this area. The scale development procedure followed the rationale of scenario-based assessment, where students are presented with a hypothetical situation and required to apply their knowledge accordingly (Din & Jabeen, 2014). Throughout the development procedure, the authors aimed to balance the context details and the cognitive efforts from respondents. Two trials were
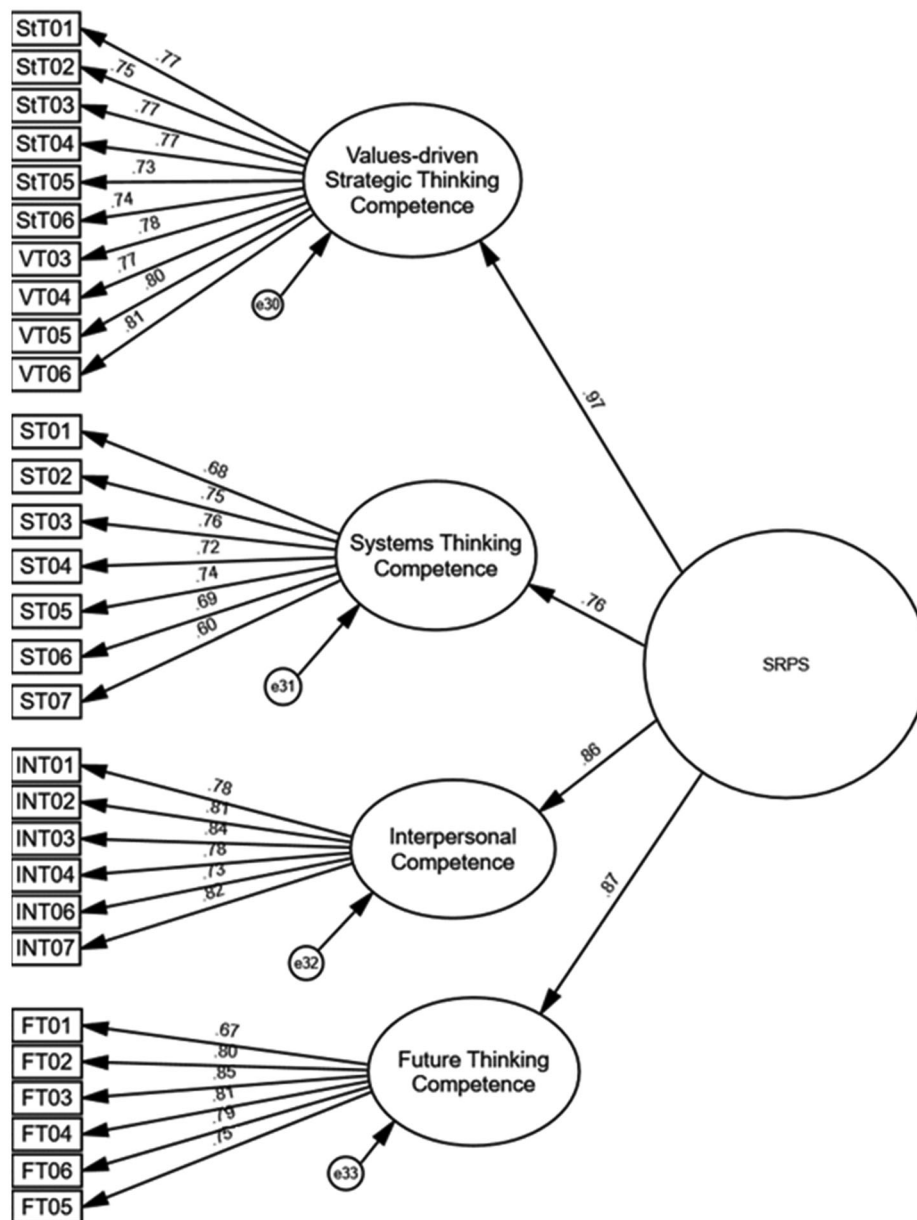
Figure 3. Second-order CFA for the Sustainability Competence Scale. Note. E30, 31, 32 and 33: Error Terms.

Table 6. The fit-indices of the proposed two models.

Model	χ^2	df	χ^2/df	TLI	NFI	CFI	IFI	RMSEA	SRMR
First-order	730.98	361	2.03	.95	.91	.95	.95	.05	.04
Second-order	737.00	363	2.03	.95	.91	.95	.95	.05	.04

Table 7. Reliability and validity for the final version scale.

Construct	Items No	$\bar{x}$	s	α	CR	AVE	MSV
ST	7	5.06	.70	.87	.88	.50	.53
FT	6	4.83	.79	.91	.90	.61	.73
VdStT	10	4.92	.71	.94	.94	.59	.73
INT	6	5.10	.73	.91	.91	.63	.69

conducted to refine the scale. In the first trial, the technique of mini-scenario was applied as a lead-in of the questionnaire, to transfer those abstract comprehension questions into a concrete situation with a problem to solve, a goal to fulfil, and a scope to image (Tucker, 2019). Cognitive interview results confirmed the effectiveness of this trial in enhancing the clarity of the items. The other trial focused on

Table 8. IFC of the final version scale.

Construct	1	2	3	4
1. ST	1	.58	.64	.62
2. FT		1	.77	.67
3. VdStT			1	.78
4. INT				1

Note. All factors correlated at the 0.01 level (2-tailed).

Table 9. CDT for testing discriminant validity.

Construct	χ	df	p
Model 1	1651.04	367	.00
Model 2	730.98	361	.00
Model 1 – Model 2	920.06	6	.00

Note. Model 1 set correlation to 0, and Model 2 allowed free correlation.

developing the subscale of Future Thinking. This involved transforming a typical scenario-based assessment with complex branching and stimulations into self-rating ones. Redesigning those open-ended questions was based on a clear understanding of the factor and its concerned variables and underlying concepts. Compared to the subscales of Systems Thinking and Interpersonal Competence, the original scales of which are self-rating with proven excellent internal consistency, the redesigned subscale of Future Thinking demonstrated similar promising internal consistency while also distinguishing itself from other subscales.

The aim of this scale, which measures relevant generic competencies roundly, was realized to a large extent through careful factor selection and item development. More importantly, it was designed to transfer the 'laundry list' of various competencies into an integrative entity, verify the integration, and measure it. Intriguing findings corroborate and prove our claim that the selected competencies are able to be integrated as a whole but still allowing for heterogeneity. To validate this scale's structure, factor analysis, a data-driven method, was used to identify and test complex interconnections among items and group items that are part of integrated concepts (Boateng et al., 2018). First, the authors selected Promax rotation in EFA, one of the oblique rotations which generally hypothesis factors were correlated. By selecting this method, two-wave data show a similar structure despite the revisions. Second, the correlations between the factors are moderately high and positive. A very strong correlation often implied redundant variables, but the results ($r = .58$ to .78) showed that the scale's factors were in a moderate relationship. Third, the CDT on second and first-order CFA showed no significant difference in their fit. It concluded that the factors could be integrated under the umbrella of a single upper-level factor – Sustainability Research and Problem-solving; therefore, the scale can be interpreted as multi-level and multidimensional (Gould, 2015).

The combination of Values Thinking and Strategic Thinking subscales, which was informed by EFA and confirmed by CFA, induced a revisiting of the original five-domain framework of sustainability competence. Strategic Thinking is a mental process used to achieve a goal through diagnosis and synthesis (Beaufre, 1965), while Values Thinking addresses how this goal is set by mapping with certain values. Although Values Thinking, as a non-cognitive disposition, and Strategic Thinking, as an action-oriented competence, can be discussed distinctively by definition, they cannot be measured separately. This is particularly true in the field of sustainability, given relevant practices often involve compromise of short-term and sectional interests and, therefore, necessitates new assessments and breaking of established action patterns (Barth et al., 2007). In other words, sustainability competence is acquired not only by knowledge restructuring but also by values interiorization – students must be enabled to discover and analyze their own values system, and revise it concerning its adequacy to the reality and context (Barth, 2014). This process of deconstruction and reconstruction is critical to explaining the value-driven and action-oriented nature of sustainability competence, which can be mutually negotiated and should be integratedly assessed. In this sense, the way integrating competence into this scale perfectly echoes the intrinsic connections and interactions of cognitive and non-cognitive competencies in the transformative process of Sustainability Research and Problem-solving.

In conclusion, the development and validation of this Sustainability Competence Scale verified Wiek et al.'s framework, which has been widely applied in formal and co-curricular settings, and is also

applicable in the extracurricular setting of higher education. This study supplements the constructive alignment of ESD by defining a clear set of learning outcomes and providing a valid assessment instrument. The mining process demonstrated how these generic competencies could be linked to an umbrella concept without fading distinct characteristics as a comprehensive integration. The applicability of the assessment tool to both formal curriculum and co/extra-curricular settings allows academic and non-academic units to measure the growth of students through various activities held on campus. From the perspective of university governance, the five-competence framework and the defined learning outcomes provides a common language that can be used across the institution, enabling more consistent educational design, delivery, and quality assurance. The findings of our study have important implications for implementing sustainability in the classroom. By assessing students' sustainability competence, higher education institutions can gain insights into the effectiveness of their sustainability education initiatives and promote critical thinking and problem-solving skills. This can empower students to become environmentally and socially conscious individuals capable of addressing sustainability challenges. Universities can apply the scale to conduct institution-wide measurement of students' sustainability competence, identify areas of improvement, and rectify the siloed manner observed in current sustainability efforts by bridging the gap between facilities, academics, and student affair (Hansen et al., 2021).

The limitation of this study is a lack of hard science perspectives when verifying the content and face validity, given the significant roles hard science professionals play in sustainable development. While statistical testing on this scale yielded promising results, there are some directions worth investigating. First, we employed a 6-point Likert scale, which requires respondents to carefully understand the context before making a positive or negative decision. Despite the impact of scale polarity on data quality being less mentioned in existing research (DeCastellarnau, 2018), it would be interesting to see whether the neutral option derived the psychometrics meaning within this scale. Additionally, the mini-scenario technique was novel in measuring sustainability competence, and its effectiveness in facilitating self-rating may vary depending on the administration and circumstances specific to different programs (e.g. the way to frame the mini-scenario as the lead-in). The possible alternations are worth examining.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

About the authors

Dr Xiao is an instructor at the Department of Applied Social Sciences and an associate fellow of the Institution of Higher Education Research and Development at The Hong Kong Polytechnic University. She has diverse experiences in instruction and assessment at both undergraduate and postgraduate levels. As an experiential educator with a strong commitment to the scholarship of teaching and learning, Dr. Xiao's research interests lie in curriculum and instruction, education for sustainable development, and education policy studies.

Kelvin Wan works at the Centre for Teaching and Learning at The Hang Seng University of Hong Kong, focusing on initiatives aimed at augmenting the pedagogical efficacy of faculty members and enhancing students' learning experiences through diverse digital solutions. He has contributed to various e-learning project, along with formulating policy pertinent to digital learning within the university. His main research interests include scale development, digital pedagogy and learning methodologies, game-based learning, and learning analytics in educational contexts.

Theresa Kwong, Ph.D., SFHEA, is the Director of the Centre for Holistic Teaching and Learning at Hong Kong Baptist University. She leads efforts to foster a culture of quality and advance the scholarship of teaching and learning within the University and beyond. She strategically promotes the development of educational technology and innovative pedagogies. Her research focuses on academic integrity, generic competencies, and technology-enhanced learning and assessment.

ORCID

Chong Xiao 🔟 http://orcid.org/0000-0001-5068-5544
Kelvin Wan 🔟 http://orcid.org/0000-0002-7860-9401
Theresa Kwong 🔟 http://orcid.org/0000-0001-8294-8580

References

Arnold, J. A., Arad, S., Rhoades, J. A., & Drasgow, F. (2000). The empowering leadership questionnaire: The construction and validation of a new scale for measuring leader behaviors. *Journal of Organizational Behavior*, *21*(3), 249–269. https://doi.org/10.1002/(SICI)1099-1379(200005)21:3 < 249::AID-JOB10 > 3.0.CO;2-%23

Barth, M. (2014). *Implementing sustainability in higher education: Learning in an age of transformation*. Routledge. https://books.google.com.hk/books?id=IXjZBAAAQBAJ

Barth, M., & Burandt, S. (2013). Adding the "e-" to learning for sustainable development: Challenges and innovation. *Sustainability*, *5*(6), 2609–2622. https://doi.org/10.3390/su5062609

Barth, M., Godemann, J., Rieckmann, M., & Stoltenberg, U. (2007). Developing key competencies for sustainable development in higher education. *International Journal of Sustainability in Higher Education*, *8*(4), 416–430. https://doi.org/10.1108/14676370710823582

Beaufre, A. (1965). *An introduction to strategy: With particular reference to problems of defense, politics, economics, and diplomacy in the nuclear age*. Praeger. https://books.google.com.hk/books?id=EJ4rAAAAYAAJ

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at University*. McGraw-Hill Education (UK). https://books.google.com.hk/books?id=XhjRBrDAESkC

Blessinger, P., Sengupta, E., & Makhanya, M. (2018). *Higher education's key role in sustainable development*. University World News. https://www.universityworldnews.com/post.php?story=20180905082834986

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149. https://doi.org/10.3389/fpubh.2018.00149

Brandt, J., Bürgener, L., Barth, M., & Redman, A. (2019). Becoming a competent teacher in education for sustainable development. *International Journal of Sustainability in Higher Education*, *20*(4), 630–653. https://doi.org/10.1108/IJSHE-10-2018-0183

Cebrián, G., Pascual, D., & Moraleda, Á. (2019). Perception of sustainability competencies amongst Spanish pre-service secondary school teachers. *International Journal of Sustainability in Higher Education*, *20*(7), 1171–1190. https://doi.org/10.1108/IJSHE-10-2018-0168

Chen, C., Lee, S., & Stevenson, H. W. (1995). Response style and cross-cultural comparisons of rating scales among east Asian and North American students. *Psychological Science*, *6*(3), 170–175. https://doi.org/10.1111/j.1467-9280.1995.tb00327.x

Cifuentes-Faura, J., & Noguera-Méndez, P. (2023). What is the role of economics and business studies in the development of attitudes in favour of sustainability? *International Journal of Sustainability in Higher Education*, *24*(7), 1430–1451. https://doi.org/10.1108/IJSHE-10-2022-0324

Crawford, J., & Cifuentes-Faura, J. (2022). Sustainability in higher education during the COVID-19 pandemic: A systematic review. *Sustainability*, *14*(3), 1879. https://doi.org/10.3390/su14031879

Crick, R. D. (2008). Key competencies for education in a European context: Narratives of accountability or care. *European Educational Research Journal*, *7*(3), 311–318. https://doi.org/10.2304/eerj.2008.7.3.311

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

DeCastellarnau, A. (2018). A classification of response scale characteristics that affect data quality: A literature review. *Quality & Quantity*, *52*(4), 1523–1559. https://doi.org/10.1007/s11135-017-0533-4

Din, A. M., & Jabeen, S. (2014). Scenario-based assessment exercises and the perceived learning of mass communication students. *Asian Association of Open Universities Journal*, *9*(1), 93–103. https://doi.org/10.1108/AAOUJ-09-01-2014-B009

European Commission, Joint Research Centre. (2022). *GreenComp, the European sustainability competence framework*. Publications Office of the European Union. https://doi.org/10.2760/13286

Fischer, D., & Barth, M. (2014). Key competencies for and beyond sustainable consumption an educational contribution to the debate. *GAIA - Ecological Perspectives for Science and Society*, *23*(3), 193–200. https://doi.org/10.14512/gaia.23.S1.7

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/10.2307/3151312

Goldman, E., & Scott, A. R. (2016). Competency models for assessing strategic thinking. *Journal of Strategy and Management*, *9*(3), 258–280. https://doi.org/10.1108/JSMA-07-2015-0059

Gould, S. J. (2015). *Second Order Confirmatory Factor Analysis: An Example*. In *Proceedings of the 1987 Academy of marketing science (AMS) annual conference*. (pp. 488–490). Springer.

Grosseck, G., Țîru, L. G., & Bran, R. A. (2019). Education for sustainable development: Evolution and perspectives: A bibliometric review of research, 1992–2018. *Sustainability*, *11*(21), 6136. https://doi.org/10.3390/su11216136

Hallford, D. J., Takano, K., Raes, F., & Austin, D. W. (2020). Psychometric evaluation of an episodic future thinking variant of the autobiographical memory test – Episodic future thinking-test (EFT-T). *European Journal of Psychological Assessment*, *36*(4), 658–669. https://doi.org/10.1027/1015-5759/a000536

Hansen, B., Stiling, P., & Uy, W. F. (2021). Innovations and challenges in SDG integration and reporting in higher education: A case study from the University of South Florida. *International Journal of Sustainability in Higher Education*, *22*(5), 1002–1021. https://doi.org/10.1108/IJSHE-08-2020-0310

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967–988. https://doi.org/10.1177/014920639502100509

Holdsworth, S., Thomas, I., & Sandri, O. (2018). Assessing graduate sustainability attributes using a vignette/Scenario approach. *Journal of Education for Sustainable Development*, *12*(2), 120–139. https://doi.org/10.1177/0973408218792127

Kassing, J. W., Johnson, H. S., Kloeber, D. N., & Wentzel, B. R. (2010). Development and validation of the environmental communication scale. *Environmental Communication*, *4*(1), 1–21. https://doi.org/10.1080/17524030903509725

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. Guilford Publications.

Komasinski, A., & Ishimura, G. (2017). Critical thinking and normative competencies for sustainability science education. *Journal of Higher Education and Lifelong Learning*, *24*, 21–37. https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/65041/1/2403.pdf

Lafuente-Lechuga, M., Cifuentes-Faura, J., & Faura-Martínez, Ú. (2024). Teaching sustainability in higher education by integrating mathematical concepts. *International Journal of Sustainability in Higher Education*, *25*(1), 62–77. https://doi.org/10.1108/IJSHE-07-2022-0221

Lambrechts, W. (2019). 21st century skills, individual competences, personal capabilities and mind-sets related to sustainability: A management and education perspective. *The Central European Review of Economics and Management*, *3*(3), 7–17. https://doi.org/10.29015/cerem.855

Loorbach, D., & Rotmans, J. (2006). Managing transitions for sustainable development. In X. Olsthoorn & A. J. Wieczorek (Eds.), *Understanding industrial transformation: Views from different disciplines.* (pp. 187–206). Springer Science & Business Media. https://doi.org/10.1007/1-4020-4418-6_10

McArthur, J. W., & Sachs, J. (2009, June 18) *Needed: A new generation of problem solvers*. The Chronicle of Higher Education. https://www.chronicle.com/article/needed-a-new-generation-of-problem-solvers/

Moore, S., Singh, M., Plamieri, P., & Alemi, F. (2011). *The Systems Thinking Scale: A measure of systems thinking*. Case Western Reserve University. https://case.edu/nursing/sites/case.edu.nursing/files/2018-04/STS_Manual.pdf

OECD. (2005). *Definition and selection of key competencies: executive summary*. Organisation for Economic Co-operation and Development. https://www.oecd.org/pisa/35070367.pdf

Orr, D. W. (2002). Four challenges of sustainability. *Conservation Biology*, *16*(6), 1457–1460. https://doi.org/10.1046/j.1523-1739.2002.01668.x

Ploum, L., Blok, V., Lans, T., & Omta, O. (2017). Toward a validated competence framework for sustainable entrepreneurship. *Organization & Environment*, *31*(2), 113–132. https://doi.org/10.1177/1086026617697039

Priede, C., Jokinen, A., Ruuskanen, E., & Farrall, S. (2013). Which probes are most useful when undertaking cognitive interviews? *International Journal of Social Research Methodology*, *17*(5), 559–568. https://doi.org/10.1080/13645579.2013.799795

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. 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

Ridgeway, C. (2001). Joining and functioning in groups, self concept and emotion management. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies.* (pp. 205–212). Hogrefe & Huber.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. *Organizational Research Methods*, *25*(1), 6–14. https://doi.org/10.1177/1094428120968614

Rowe, D. (2007). Education for a sustainable future. *Science (New York, N.Y.)*, *317*(5836), 323–324. https://doi.org/10.1126/science.1143552

Savage, E., Tapics, T., Evarts, J., Wilson, J., & Tirone, S. (2015). Experiential learning for sustainability leadership in higher education. *International Journal of Sustainability in Higher Education*, *16*(5), 692–705. https://doi.org/10.1108/IJSHE-10-2013-0132

School of Sustainability, Arizona State University. (2018). *Key competencies in sustainability*. https://static.sustainability.asu.edu/schoolMS/sites/4/2018/04/Key_Competencies_Overview_Final.pdf

Shephard, K., Harraway, J., Lovelock, B., Mirosa, M., Skeaff, S., Slooten, L., Strack, M., Furnari, M., Jowett, T., & Deaker, L. (2015). Seeking learning outcomes appropriate for 'education for sustainable development' and for higher education. *Assessment & Evaluation in Higher Education*, *40*(6), 855–866. https://doi.org/10.1080/02602938.2015.1009871

Siew, N., & Rahman, M. (2019). Assessing the Validity and Reliability of the Future Thinking Test using Rasch Measurement Model. *International Journal of Environmental and Science Education*, *14*(4), 139–149. http://www.ijese.net/makale_indir/IJESE_2110_article_5d360a01d139a.pdf

Taimur, S., & Sattar, H. (2019). Education for sustainable development and critical thinking competency. In W. Leal Filho, A. Azul, L. Brandli, P. Özuyar, & T. Wall (Eds.), *Quality education.* (pp. 1–11). Springer International Publishing. https://doi.org/10.1007/978-3-319-69902-8_64-1

Tucker, C. (2019, November 26). *Mini-scenarios for assessment*. Experiencing eLearning. https://www.christytucker-learning.com/mini-scenarios-for-assessment

Waltner, E., Rieß, W., & Mischo, C. (2019). Development and validation of an instrument for measuring student sustainability competencies. *Sustainability*, *11*(6), 1717. https://doi.org/10.3390/su11061717

Watson, J. C. (2017). Establishing evidence for internal structure using exploratory factor analysis. *Measurement and Evaluation in Counseling and Development*, *50*(4), 232–238. https://doi.org/10.1080/07481756.2017.1336931

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies.* (pp. 45–66). Hogrefe & Huber.

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

Willard, M., Wiedmeyer, C., Warren Flint, R., Weedon, J. S., Woodward, R., Feldman, I., & Edwards, M. (2010). The sustainability professional: 2010 competency survey report. *Environmental Quality Management*, *20*(1), 49–83. https://doi.org/10.1002/tqem.20271

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. SAGE Publications. https://books.google.com.hk/books?id=g9cdAQAAIAAJ

Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research*, *30*(1), 72–91. https://doi.org/10.1086/374697

Zait, A., & Bertea, P. E. (2011). Methods for testing discriminant validity. *Management & Marketing*, *9*(2), 217–224. http://www.mnmk.ro/documents/2011-2/4_Zait_Bertea_FFF.pdf