DOI: 10.2478/dcse-2024-0014

sciendo

Discourse and Communication for Sustainable Education,
vol. 15, no. 2, pp. 28-44, 2024

Using Large Scale Summative Tests to Advance the Integration of Sustainable Development Competencies: A Model for Test Construction

Clavia T. Williams-McBean
https://orcid.org/0000-0003-3434-8913
The University of the West Indies, Mona, Jamaica

Abstract

There is an urgent need for global citizens to develop sustainable development competencies (SDCs) if our world is to survive threats to its development and sustainability. These competencies can be developed through education for sustainable development (ESD). However, implementing ESD is challenged by teachers' and students' prioritisation of the assessed curriculum while ESD remains largely unassessed in the education sector. The knowledge, skills, values and attitudes reflected in tests and assessments, especially large-scale, standardised, summative tests, influence classroom practices considerably. This paper proposes that the integration of SDCs can be accelerated by constructing large-scale tests to measure these competencies, thereby capitalising on the influence of the tests. It presents a model for constructing these tests and provides ways in which the format of multiple choice and extended written response items may be optimised to measure SDCs.
Keywords: Education for sustainable development (ESD), sustainable development competencies, teaching to the test, the Caribbean Examinations Council (CXC), multiple choice questions (MCQs), assessment.

Introduction

Education for sustainable development (ESD) needs to be integrated over longer periods and in more subjects across the curricula to advance the development of sustainable development competence (SDCs) (Pierce & Hemstock, 2021; Roure et al., 2018; Weber et al., 2014; Heasly et al, 2024).). There is also a need to focus on assessment in ESD since its content guides teacher and student preparatory actions and the field needs insights on how sustainable development outcomes can be measured (Ayers et al., 2020; Gajparia et al., 2021). This second conceptual paper extends the proposition that the integration of SDCs can be accelerated by capitalising on the practice of teaching to the test. This integration can be achieved by developing large-scale assessment systems to include both traditional tests and school-based assessments (SBA) and using each component to measure SDCs. The first paper established the proliferation of the practice of teaching to the test, described the Caribbean Examinations Council's (CXC) assessment model, and proposed ways in

which the SBA component can be optimised to measure SDCs; thereby capitalising on the practice of teaching to the test (Williams-McBean, 2023). This paper presents a model for constructing assessments for measuring SDCs through each item format (multiple choice, extended written response and SBAs) and provides guidelines on how the item formats frequently used in traditional large-scale tests – multiple choice and extended written responses (e.g., essays) – may be optimised for measuring SDCs. Importantly, teaching to the test, as used in this paper, refers to curriculum testing, where teachers and students use the knowledge and skills represented on the test to design instruction, and not item-teaching, where teachers expose students to similar or cloned items on an upcoming test (Popham, 2001). The premise is that since teaching to the test is widespread, with large-scale standardised tests significantly influencing classroom activities, then test developers could construct these tests to measure important educational outcomes, specifically SDCs, to advance its integration into the curriculum and teachers' and students' development of the competencies.

Assessment of the Sustainable Development Competencies

In 1987, the World Commission on Environment and Development, in the Brundtland Report, defined sustainable development as "meeting the needs of society without compromising the ability of future generations to meet their own needs" (p.43). In meeting the intra and intergenerational needs, the focus is placed on the environment, the economy and the society –including its cultural aspects. The United Nations (UN) (2015) identified 17 Sustainable Development Goals (SDGs), including quality education, to be achieved by 2030 which should ensure prosperity for all now and in the future. The United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2017) also outlined eight key SDCs that support the SDGs of the United Nation's (UN) 2030 Agenda for Sustainable Development: systems thinking competency, anticipatory competency, normative competency, strategic competency, collaboration competency, critical thinking competency, self-awareness competency and integrated problem-solving competency. These SDCs can be developed through ESD.

UNESCO Global Action Programme on Education for Sustainable Development (2018) explains:

> *ESD empowers learners to take informed decisions and responsible actions for environmental integrity, economic viability and a just society, for present and future generations, while respecting cultural diversity. It is about lifelong learning, and is an integral part of quality education. ESD is holistic and transformational education which addresses learning content and outcomes, pedagogy and the learning environment. It achieves its purpose by transforming society*.

ESD is both an outcome and a means by which the SDGs can be achieved. As the Global Education Monitoring Report (GEM report) explains, education has a critical role to play in accomplishing all the SDGs since each goal requires, among other things, training, raising awareness and educating citizens. Education can transform society, but it can also be transformed to meet these developmental goals (Sterling, 2016). This transformation can be evidenced by the extent to which ESD is integrated into four aspects of education: national education policies, curricula, teacher training and

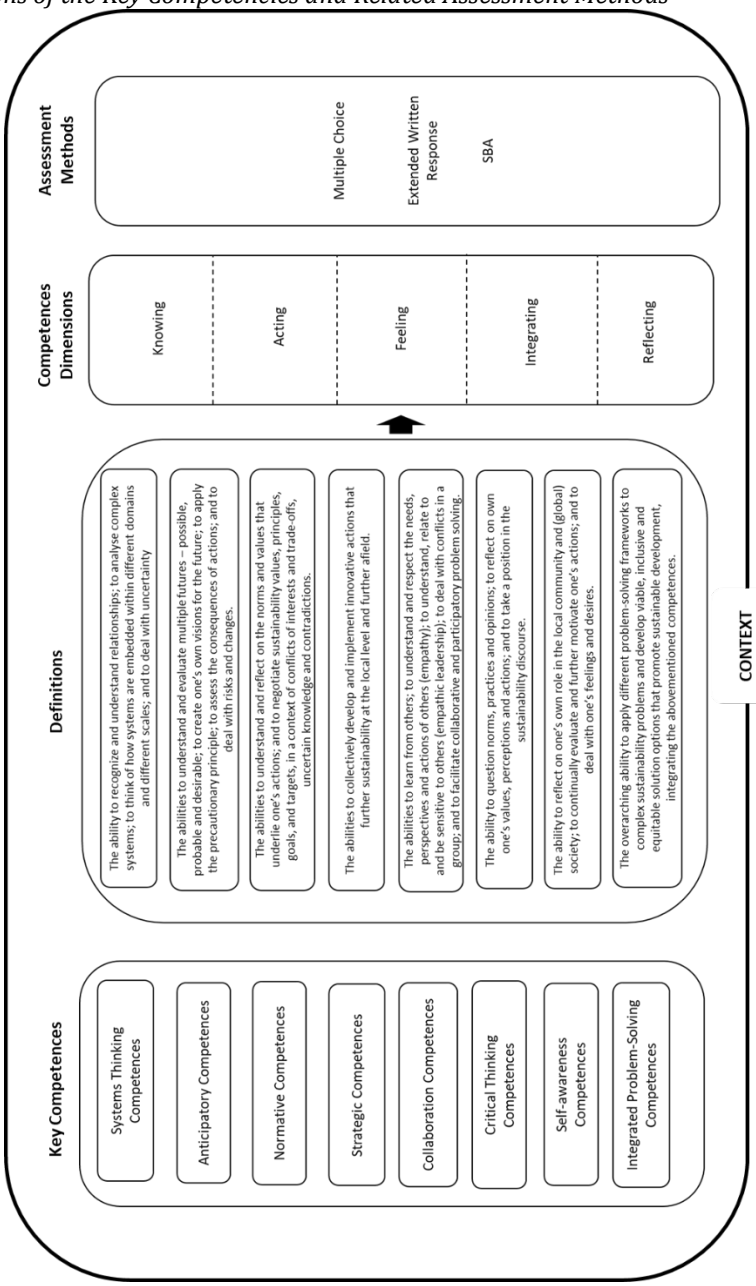
student assessment (Holst et al., 2024; UNESCO, 2019). As Holst et al. (2024) explained, the student assessment aspect requires designing assessment items that promote and measure sustainable development competencies. Therefore, sustainable development should be infused into education for education to serve its transformative function. To achieve the education and other SDGs through education, there is a need to reorient the education system toward the knowledge, skills, values, attitudes and experiences needed to solve complex problems threatening development for present and future generations. However, ESD implementation is challenged by its peripheral position in the education sector (Holst et al., 2024), and the fact that it remained largely 'non-mandated and non-assessed' (Beasy & Gonzalez, 2021, p. 9). One way in which education can be transformed and transformative is by implementing assessment systems and constructing tests and assessments that can validly measure the SDCs. Without these assessments, the efficacy of teaching SD content using SD-supportive pedagogies and in learning environments conducive to SD will remain unknown. Assessment therefore is both a means of assessing the effectiveness of efforts towards ESD and a way to motivate educational actors to integrate and develop the SDCs. While others propose that assessment for learning (formative assessment) and assessment as learning (self-assessment) is to be used instead of assessment of learning (summative assessments) to support ESD (see, for example, Gajparia et al., 2021; Gey et al., 2023), I propose that all three types of assessment are necessary. Additionally, since educational stakeholders place greater attention on summative assessments, especially those that have life-changing effects on students by determining, for example, whether and which colleges they matriculate into, job prospects and social mobility, summative assessments constructed to measure SDCs can be used to advance the integration and development of the SDCs.

Using CXC's Assessment Model to Advance SD Competencies

As previously explained, the CXC's assessment model includes a combination of traditional external, summative examinations and alternative, school-based assessments, which includes both a formative and a summative component. The traditional assessments include multiple choice examinations (Paper 01) and an extended written response paper (Paper 02). The school-based assessment component (Paper 03) takes varying formats, including research reports, multimedia presentations and community projects. This model is ideal for including the knowledge, skills and affective dispositions reflected in the key competencies of sustainable development. As Williams-McBean (2023) explained, the key SDCs comprise knowing, acting, feeling, integrating and reflecting components, which can be measured by the various assessment methods utilised in CXC's examinations (see Figure 1).

Figure 1

Dimensions of the Key Competencies and Related Assessment Methods



Note. Adopted from "Using school-based assessments to advance the integration of sustainable development competences by capitalising on the practice of teaching to the test," by C. T. Williams-McBean, 2023. In *Environmental Education Research*, 29(7) 715-732. Copyright 2023, Taylor & Francis Group.

All definitions are from UNESCO (2017) Education for Sustainable Development Goals; Learning objectives; Paris, France: UNESCO.

Each dimension was explained in the first article thus:

> *The knowing dimension refers to the cognitive knowledge or thinking skills needed to demonstrate each SDC and their dimensions and includes both lower order thinking skills (LOTS) – remembering and understanding, and higher order thinking skills (HOTS) – applying, analysing, evaluating and creating. The acting dimension refers to the practical demonstration of competency and includes physically doing or encouraging others to take action towards achieving the SDGs. The feeling dimension refers to the socio-emotional skills related to each key component and the integrating dimension refers to the integration of content, skills, competencies and people to work towards achieving SDGs. It also includes transdisciplinary learning and assessment, group work and collaboration with the community and businesses (internship). Finally, the reflecting dimension involves critically thinking about one's thoughts, actions, feelings and connections or relationships (with self, others, society and the environment) to identify and change them towards SD (pp. 723–724).*

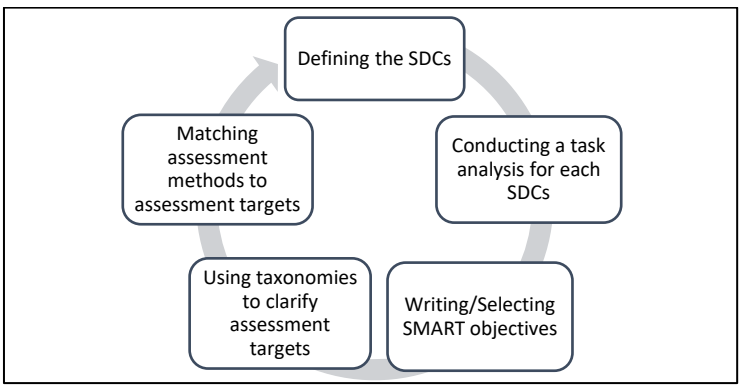
Each dimension can be most appropriately measured by one or more of the item formats utilised in standardised assessments: MCQs, extended written response items and SBA. However, the items should be constructed at appropriate levels of difficulty, following test construction rules and with consideration for and inclusion of the local, regional and international contexts.

A Model for Constructing Varied Assessment to Measure the SDCs

Measuring the SDCs with the appropriate assessment methods requires a systematic approach to ensure the requisite knowledge and skills are aligned with the appropriate method. This alignment can be achieved through five steps (see Figure 2). Standardised and classroom test developers can use these five steps to determine what they want to measure and the most appropriate assessment method for measuring what they want to measure. Aligning what to measure and how to measure will contribute to the validity of the interpretation of the scores produced.

Figure 2

A Process Model for Aligning the Dimensions of the Key SDCs with Appropriate Assessment Methods



Step 1 – Test construction begins with clarifying what is to be measured. This requires test developers to first **define the SDCs.** What therefore does anticipatory or normative competency mean? Here, test developers should use a credible and current competence framework (e.g., UNESCO, 2017) to identify and define the SDCs.

Step 2 – Clarifying what needs to be measured will also require a task analysis because all the competencies are complex (i.e., each competency can be divided into sub-skills). Test developers will need to **conduct a task analysis to identify all the constituent knowledge, skills, values and attitudes that students will need to demonstrate the competency(ies) being measured.** The definitions of each SDC, as provided by UNESCO (2017), provide useful insights. Reading the extant literature on the SDCs can further clarify the sub-skills. Task analysis is needed to break complex skills into sub-skills for easier measurement.

Step 3 – Once all the knowledge, skills, values and attitudes have been identified, test developers will **write/select specific, measurable, attainable, relevant, and timely (SMART) objectives.** Assessments are directly aligned to objectives to increase validity, and SMART objectives make it easier to construct assessments to measure stated objectives. Avoid using unobservable and ambiguous verbs such as recognise, understand and appreciate. Instead use specific verbs, for example, evaluate, defend, compare and propose.

Step 4 – With the objectives in place, **use taxonomies to clarify the assessment target.** The assessment target is what is to be assessed. The targets should be clarified because they determine how the competencies are to be assessed. Taxonomies can be used to clarify the assessment targets because they help to identify the cognitive, psychomotor, and affective levels of each constituent objective and the knowing, acting, feeling, integrating, and reflecting dimensions of the SDCs. For example, the Revised Bloom's Taxonomy (Anderson et al., 2001) can be used to identify the level of each cognitive objective (remembering, understanding, applying, analysing, evaluating and creating) or level of knowing required by the SDCs. Identifying the levels of the objectives makes it easier to clarify the assessment target. According to Chappuis et al. (2012), there are five assessment targets: knowledge, reasoning, skill, product-development and disposition. The knowledge target consists of the lower levels of the cognitive taxonomies while the reasoning target includes the higher levels. Using taxonomy to classify the cognitive objectives as lower- or higher-order enables the identification of the knowledge or the reasoning assessment target. Skill and product-development are usually associated with the psychomotor domain, and the disposition target is associated with the affective domain. Taxonomies for the psychomotor and affective domains can also be used to clarify the required levels of acting, feeling, integrating and reflecting.

Step 5 – The final **step is to match the most appropriate assessment method to the identified assessment target(s).** Chappuis et al. (2012) also identified four assessment methods (selected-response items, extended written response, performance (alternative) assessment and personal communication) that are most suitable for each assessment target. Selected-response items (e.g., multiple choice items) are most suitable for assessing discrete knowledge because they are most efficient while performance assessments are most suitable for assessing skill and product-development targets. Performance assessment would be most suitable for assessing the affective target in large-scale assessment as well because they allow for collaboration, reflection, direct observation of values in action and actively responding to local and global SD challenges.

It is to be noted that the process is cyclical because once the appropriate assessment method is identified, the assessment will be constructed, administered, scored, and evaluated. The results of those processes will be used to define the SDCs further: identify strengths and weaknesses in the current assessment and its coverage of the specified knowledge and skills related to the SDCs and make improvements for subsequent administrations.

To illustrate the application of the process model, I will use the systems thinking competency. This competency was selected because teachers are critical change agents in the integration of SD in education, and research has highlighted the need to reorient teacher education in Jamaica to emphasise systems thinking. Additionally, Jamaican teachers' perspective on SD mainly exhibited a systems thinking perspective (Ferguson et al., 2021). Systems thinking was also selected because the SDGs are interrelated (UN-DESA, 2018), and systems thinking competency is primarily concerned with interrelatedness within and between systems (UNESCO, 2017). Furthermore, systems thinking competency is necessary for success within and between every subject taught in school. In Sociology, society and its components (e.g., family, religion, education) consist of interrelated parts. Additionally, with the increased focus on STEM and STEAM education, there is a concerted move away from the fragmented presentation of subjects in school to a greater focus on the interrelatedness of these subjects. This means that the assessment of systems thinking is inevitable. Any contribution to how this competency can be assessed will have local and global significance. However, the process used to align the dimensions of systems thinking with appropriate assessment methods applies to any individual or combination of the competencies.

1. **Identifying the SDCs.** Systems thinking competency has been defined by UNESCO (2017) as "the ability to recognize and understand relationships; to analyze complex systems; to think of how systems are embedded within different domains and different scales; and to deal with uncertainty" (p. 10).
2. **Conducting a task analysis to identify all the constituent knowledge and skills that students will need to demonstrate the competency.** Some of the knowledge and skills are included in the definition above. Benninghaus et al., (2019) further explained that,

   > In principle, such competence is understood as the ability to describe, reconstruct, and model a complex realm of reality as a system and, based on the model, provide explanations, make forecasts, and design, assess, and implement possible courses of action. (p. 3).

These complex realms include understanding the relationship among the systems within the economic, social and environmental aspects of society and the connections among the aspects. Consequently, systems thinking has four dimensions and sub-capabilities that increase in complexity: (1) declarative/conceptual system knowledge (knowledge of system hierarchies), (2) systems modeling (reading, understanding and constructing system models or diagrams), (3) problem-solving using system models (giving explanations, making predictions and designing technologies based on system models) and (4) evaluation of system models (determining the validity of the application) (Benninghaus et al., 2019).

3. **Writing specific, measurable, attainable, relevant, and timely (SMART) objectives.** The objectives to be created or selected will then be specific to the syllabus, subject or groups of subjects and context. An example of a specific objective requiring systems thinking competency from the CXC/CAPE Green Engineering Syllabus is "outline the sustainability issues associated with each

stage of the lifecycle framework" (CXC, 2015, p. 8). Successful achievement of this objective requires candidates:

(a) list the components of the lifecycle framework and different SD issues (e.g., energy consumption and waste generation) and outline the stages of the lifecycle framework (knowledge of system hierarchies),
(b) explain and construct models of the lifecycle framework including SD considerations (systems modeling),
(c) explain and predict how various SD issues affect the stages of existing or proposed lifecycle frameworks (problem-solving using system models) and
(d) determine the validity of the application of lifecycle frameworks considering SD issues (evaluation of system models).

4. **Using taxonomies to clarify the assessment target.** Using the Revised Bloom's Taxonomy, the stated objectives in the knowledge of systems dimension of systems thinking are at the remembering (list the components of the lifecycle framework and different SD issues (e.g., energy consumption and waste generation)) and understanding levels (outline the stages of the lifecycle framework). Both objectives indicate lower-order thinking. They indicate the knowledge assessment target and the knowing dimension of systems thinking competency.
5. **and matching the most appropriate assessment method to the identified assessment target(s).** Objectives associated with the knowledge target could be suitably assessed using multiple choice items in large-scale assessments.

This process is suitable for identifying the objectives of each of the SDCs that are most suitably assessed by the different assessment methods (see Table 1).

Table 1.
Methods of Assessment for Measuring the Dimensions of Systems Thinking

Dimensions of Systems Thinking	Methods of Assessment	Possible Assessment Tasks
Declarative/conceptual system knowledge	Selected-response items (e.g., multiple choice items)	List the factors that contribute to poverty in their society. Identify the effects of poverty in their communities. Explain the concepts of absolute and relative poverty.
Systems modeling	Extended written response, projects	Create a concept map reflecting the cycle of poverty in given scenarios (case studies) or within a selected community (project).
Problem-solving using system models	Projects, lab experiments, research reports	As a group, design and implement a systems model to address an identified SD issue in your community.
Evaluation of system models	Projects, Reflective essays	Assess the success of the implemented model focusing on personal, group, and model strengths and weaknesses.

Table 1 also includes concept mapping, which is usually used to assess systems thinking (Benninghaus et al., 2019; Roczen et al., 2021). Examining each SDC in this way allows for the connections within and between SDCs to be identified and measured together. The possible assessment tasks can be included in high-stakes assessments systems that include external examinations and SBAs. Then, their inclusion in high-stakes assessments could influence teachers to include them in their formative classroom practices since teachers usually model standardised assessment tasks (Berry, 2010; McMillan, 2003; Ong, n.d.; Williams-McBean, 2020; 2022).

However, to develop these competencies, it is not enough to put students in situations where the competencies can be developed. Students need to know and understand what these competencies entail, be armed with useful strategies to demonstrate them, use the strategies to demonstrate their competence, and then reflect on and evaluate their demonstration. For example, to develop collaboration competency, it is not enough to incorporate group work in the classroom or assign group projects for students to complete. Students must know, for example, the roles and functions of different types of groups, strategies for dealing with conflicts within groups and ways of demonstrating respect or empathy for others, before they can be expected to demonstrate them. For this reason, each competency requires cognitive knowledge and skills (lower and higher order), psychomotor or practical skills and affective or social and emotional skills. All of which can be reasonably expected to be taught in the classroom and, therefore, measured by large-scale, standardised tests, especially those that include an SBA component. The challenge is to integrate SD contents such as good health and well-being, climate change, quality education, poverty, gender equality, clean water and sanitation, and sustainable consumption, sustainable cities and communities into the tests and optimising the items to measure, not just require, the demonstration of varying levels of accomplishments of the SDCs.

Integrating the SD Content and Optimising the Assessment Methods in Large-Scale Assessments for SDC

This section does not aim to provide a detailed outline of SD content in general. Many useful resources provide that information (see, UNESCO 2017). The aim is to suggest ways SD content can be integrated into large-scale, high-stakes, summative assessments. This integration can be achieved by offering standalone examinations of SD and integrating SD content and competencies in the existing subject examinations.

Offering Standalone SD Examinations

Sustainable development can be taught as a standalone course at every level of the education system, with the content being age and context-appropriate. Offering these examinations could lead to changes in policy and practice that are needed to allow education to support SD knowledge and practice. It could make it easier for the government to implement policy changes that mandate achievement in sustainable development at all levels. It would influence universities and teacher training institutions to develop and revise current programmes to adequately prepare teachers for their subsequent preparation of students for said examinations. It could also motivate teachers to become specialists in SD since there would be increased employment opportunities if SD became mandatory. It could influence textbook developers to revise or create textbooks to support the development of SDCs. This

revision of textbooks to reflect changes in large-scale assessment was evident after CXC implemented SBAs in English and Mathematics. The textbooks were revised or created to include SBA components in response to the assessment change. Furthermore, universities and non-educational organisations could use achievement in SDC as a requirement for scholarships since a valid and reliable assessment is readily available to give them the data they would need to award scholarships. A standalone examination of SD could influence the integration of SD content and competencies across every aspect of education.

Integrating SD Content, Knowledge, and Skills in Current Examinations

Sustainable development content and competencies are inter- and transdisciplinary. They can be easily integrated into subjects such as CAPE Agricultural Science, Entrepreneurship, Environmental Science, Tourism and Economics, CSEC Social Studies, Religious Education and Principles of Business. In fact, there is already evidence of this integration in the examinations and syllabuses developed by the CXC. In CAPE Tourism, one of three modules in UNIT 1 is "Sustainable Tourism" which requires students to, among other things, "understand concepts (pillars) of sustainability as they relate to tourism [and] demonstrate how sustainable tourism benefits communities and destinations" (CXC, 2013, p. 19). Perhaps the most notable integration of SD in the CXC's examinations is the addition of Green Engineering at the CAPE level in 2017. According to the CXC,

> *the study of CAPE Green Engineering will enable students to acquire the knowledge, skills, values and attitudes needed to sustain the natural environment … apply Scientific, Technological, Engineering and Mathematical (STEM) principles to improve their environment at the local, regional and global levels [and] enhance quality of life for present and future generations, while providing wealth creation through new and innovative job opportunities and other economic possibilities including entrepreneurship*. (para. 1 – 2).

This subject pointedly eschews the pillars of SD. However, further integration is needed in mandatory subjects such as Mathematics and English, which have the largest uptake of student, considered core subjects at the primary and secondary levels and required to access higher education and employment. This integration can be achieved in English, for example, by ensuring that at least one of the stimuli on each examination paper is related to SD or using statistics related to sustainability issues to demonstrate Mathematical reasoning or calculations. If SD is integrated into these high-uptake core subjects, more students and teachers would feel compelled to learn and teach SD because they would view it as necessary for doing well on these high-stakes examinations. In this way, the culture of teaching to the test would serve the advancement of ESD.

Optimising the Assessment Methods in Large-Scale Assessments for Measuring SDC

As previously discussed, examinations that develop and offer large-scale standardised examinations usually use selected-response items (primarily multiple-choice questions) and constructed response items (primarily essays). Some of these examinations include an internal SBA component which usually utilises alternative assessment formats such as projects, research reports and laboratory experiments. CXC utilises all three assessment methods, and all three can be optimised to measure SDC and thereby influence teachers to include and develop students' competency in SD in the classroom. This paper focuses on the traditional methods: MCQs and extended written responses since the SBA component was discussed in the first paper.

Multiple Choice Questions (MCQs).

MCQs contribute to the efficiency of large-scale testing, such as the CXC examinations. Despite the common perception that MCQs are limited to measuring lower order thinking skills (LOTS) – remembering and understanding, they can and have been used to validly measure higher order thinking skills (HOTS) – applying, analysing, evaluating and creating – in many fields and most countries across the globe (Imeson, 2017; Kolomitro, 2020; Schilling, 2019; Scully, 2017). There is also evidence that valid and reliable MCQ tests can be developed to measure SDC (Carteron et al., 2014; Roczen et al., 2021). Since it is generally accepted that MCQs can effectively measure LOTS, the focus here is on how MCQs can be optimised to measure HOTS. Notwithstanding, even in measuring LOTS, MCQs should be well-constructed – including a clear stem, plausible distractors and a correct answer, while adhering to established item-writing rules that are widely available. To optimise traditional MCQs (with a stem and four or five options from which students select a single correct answer) to measure higher-order thinking, examination boards can utilise interpretive exercises, multiple choice with justifications and multiple response format multiple choice questions.

Interpretive Exercises. Interpretive exercises are test items in which students are presented with a common stimulus (text, pictures, tables, maps) which is accompanied by a series of MCQs based on the stimulus. They are most suited for measuring HOTS such as identifying relationships, comparing and contrasting (methods, perspectives, actions, authors), evaluating assumptions, arguments, and inferences, designing experimental procedures, formulating conclusions or hypotheses, applying and assessing the application of principles and methods and integrating knowledge from related areas (Miller et al., 2013). These higher-order thinking skills are reflective of SDCs. Systems thinking requires students to identify relationships and anticipatory competence requires students to assess causes and effects, and all the competencies require the application of principles and methods. There is evidence that these exercises are being used to measure SDCs. The MCQ test developed by Roczen et al. (2021) measured systems thinking without being subject-specific, and the items were designed and weighted to measure the varying levels of systems thinking using interpretive exercises.

Multiple choice with justification. These MCQs require students to (a) select an answer to a MCQ (b) justify their selected answer. The premise of MCQs is that students select correct responses because they possess the underlying knowledge and skills to identify the correct answer. However, students may guess the correct answer because the options are presented to them. In fact, for four-option MCQs, the most popular format utilised in standardised testing, students have a 25% chance of guessing the correct answer. The possibility of guessing the correct answer increases if there are undemanding distractors. Additionally, research has shown marked differences in students' selection of correct responses and justifications for those responses, with students who selected the correct responses providing incorrect or partially correct justifications and vice versa (Dulger & Deniz, 2017; Tamir, 1991). The obvious advantage of utilising MCQs with justification is that they clearly indicate students' understanding and allow assessors to determine whether students guessed or possessed the requisite knowledge and skills (i.e., increased validity). This item format also increases the possibility of detecting cheating. It may also have a more positive washback effect as students may attempt to learn the content on a deeper level to ensure they can provide accurate justifications. It would also require teachers to prepare students to provide reasoned justification rather than merely selecting correct answers. Incorporating these items would come at the expense of the efficiency and objectivity of the traditional format of MCQ. However, since measurement experts have agreed that validity is a primary concept in educational measurement (Griffith, 2015; Williams-McBean, 2018) and the addition of justifications would allow for increased accuracy of measurement and interpretations of test scores (i.e., validity), the tradeoff may be worth it.

Multiple response format (MRF) multiple choice questions. These MCQs require students to select multiple correct answers or rate each option as true or false. MRFs were developed in response to the limitation of traditional MCQs in detecting students with partial conceptual understanding whereby students select the correct answer but still think another or other options are correct (Brassil & Couch, 2019; Nehm & Schonfeld, 2008). MRFs allow for that detection and encourage deeper thinking by requiring students to evaluate each response option (Brassil & Couch, 2019; Kalas et al., 2013; Newman et al., 2016). It can suitably measure higher-order thinking, such as critical thinking, which is an essential SDC and twenty-first-century skill. Other advantages of using MRFs include the identification of attractive distractors, the reduction of guessing, the maintenance of objective scoring and increased validity (Brassil & Couch, 2019). Additionally, since only a moderate increase in time has been reported by the use MRFs (Brassil & Couch, 2019), large-scale testing agencies, including CXC, could include these item formats and increase the level of learning required for passing its examinations while maintaining efficient and objective scoring and wide coverage of the required content. In requiring students to select multiple correct answers, it would be best not to indicate how many correct answers should be selected or how many options are true or false. This omission would reduce the possibility of guessing the correct answer while promoting a greater assessment of the options to identify all the correct responses.

Multiple choice items have been used in large-scale assessments across the globe because they are efficient, can be easily administered to large groups of students, are objective, reliable and allow for greater content coverage. Notwithstanding they have been heavily criticised as measuring predominantly lower-order thinking skills.

Consequently, there may be some reservations about the suitability of MCQs to assess the higher-order thinking skills necessary to demonstrate SDCs. However, utilising interpretive exercises, MCQs with justification and MRFs can optimise MCQs for SD assessment. CXC already includes interpretative exercises. However, there is no evidence of MCQs with justification or MRFs in its examinations. Additionally, to improve the prospect of advancing SDCs through summative, traditional assessments, examination boards should ensure interpretive exercises include SD content and the questions based on the stimuli provided require higher-order thinking within and across subjects. Consideration would also be given to including MCQs with justification and MRF, especially in e-testing, where the possibilities for cheating and the presentation of multimodal stimuli are significantly increased.

Extended Written Response Items

Extended written response items are constructed-response items that require students to create their answers instead of selecting them. They include short answer questions, essays and other extended written response formats (concept mapping, drawing). It is generally accepted that these items are suitable for measuring HOTS and inappropriate for measuring LOTS only (Miller et al., 2013). They have also been credited for measuring a range of skills in a single item. A single essay can measure students' content knowledge, application of that knowledge and evaluation of their application. It is also suited for assessing students' ability to generate ideas and organise and synthesise information to create cogent responses. These items' limitations include limited content coverage as students can only write a few essays in an examination, and they are time-consuming and difficult to score objectively, even when rubrics are used to guide scoring (Miller et al., 2013). They may reduce the reliability and validity of the inferences from the test scores produced by these examinations (Schilling, 2019).

Constructed-response items are usually included on Paper 02 and Paper 03/2 (Alternative Paper) on the CXC's CAPE and CSEC examinations. Most of these items require students to write essays, reports, stories and draw diagrams. Since the constructed-response items have been generally accepted as measuring HOTS, the emphasis is on optimising these examinations to include those HOTS related to SD. This can be accomplished in the following ways:

1. including SD content in questions,
2. making greater use of concept mapping and case studies,
3. including more items that require forecasting and backcasting,
4. writing items that require interdisciplinary skills and content and reflection.

Conclusion

Teaching to the test is inevitable, especially when standardised tests determine students' future. However, it has a negative reputation because it focuses on item teaching and encourages drilling instead of learning. The main criticism has been that standardised tests, especially those that include only MCQs, promote rote and surface learning. However, if tests are constructed to require deeper learning, for example, to interpret stimuli, provide justifications for selected responses or determine which answers are correct, then traditional MCQs can be optimised to require and measure SDCs. Additionally, including SD content and competencies on large-scale summative tests can lead to curriculum teaching, thereby advancing the integration and

development of SDCs. Importantly, including and measuring SDCs in large-scale tests do not negate the need for policy and curricula changes, teacher training and administrative support for SD to be more fully integrated into education and for ESD to contribute to accomplishing the other SDGs.

References

Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Ayers, J., Bryant, J., & Missimer, M. (2020). The use of reflective pedagogies in sustainability leadership education: A case study. *Sustainability*, *12*, 6726. https://doi.org/10.3390/su12176726

Beasy, K., & Gonzalez, L. R. (2021). Exploring changes in perceptions and practices of sustainability in ESD communities in Australia during the COVID-19 pandemic. *Journal of Education for Sustainable Development*, *15*(1), 5-24. https://doi.org/10.1177/09734082211012081

Benninghaus, J. C., Mühling, A., Kremer, K., & Sprenger, S. (2019). The mystery method reconsidered—A tool for assessing systems thinking in Education for Sustainable Development, *Education Sciences, 9*(4), 260. https://doi.org/10.3390/educsci9040260

Berry, R. (2010). Teachers' orientations towards selecting assessment strategies. *New Horizons in Education*, *58*(1), 96–107.

Brassil, C., & Couch, B. (2019). Multiple-true-false questions reveal more thoroughly the complexity of student thinking than multiple-choice questions: a Bayesian item response model comparison. *International Journal of STEM Education* 6,16 https://doi.org/10.1186/s40594-019-0169-0

Caribbean Examination Council (2013). *Caribbean Advanced Proficiency Examinations Tourism Syllabus*. St. Michael, Barbados: Caribbean Examination Council.

Caribbean Examination Council (2015). *Caribbean advanced proficiency examinations green engineering syllabus, specimen paper & Mark Scheme*. St. Michael, Barbados: Caribbean Examination Council.

Caribbean Examination Council (2017). *Caribbean Advanced Proficiency Examinations Green Engineering Syllabus, specimen paper, mark scheme, subject reports*. St. Michael, Barbados: Caribbean Examination Council.

Carteron, J., Haynes, K., & Murray, A. (2014). Education for sustainable development, the UNGC PRME initiative, and the sustainability literacy test: Measuring and assessing success: Quarterly journal. *S.A.M. Advanced Management Journal, 79*(4), 51–58.

Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right – Using it well*. (2nd ed.), Pearson.

Dulger, M., & Deniz, H. (2017). Assessing the validity of multiple-choice questions in measuring fourth graders' ability to interpret graphs about motion and temperature. *International Journal of Environmental and Science Education*, *12*(2),177-193.

Ferguson, T., Roofe, C., & Cook, L.D. (2021). Teachers' perspectives on sustainable development: the implications for education for sustainable development, *Environmental Education Research*, *27*(9), 1343–1359, doi: 10.1080/13504622.2021.1921113

Gajparia, J., Strachan, G., Vare, P., & Ferguson, T. (2021). Identifying assessment opportunities in postgraduate learning for sustainability. *Discourse and Communication for Sustainable Education*, *12*(1), 151–175. https://doi.org/10.2478/dcse-2021-0011

Gey, N., Pellaud, F., Blandenier, G., Lepareur, C., Massiot, P., Shankland, R., & Gay, P. (2023). Assessment of cross-cutting competences in education for sustainable development. *Environmental Education Research, 29*(8), 1132. https://doi.org/10.1080/13504622.2023.2252676

Griffith, S. A. (2015*). School-based assessment in a Caribbean public examination*. Kingston, Jamaica: The University of the West Indies Press.

Heasly, B., Iliško, D., & Lindner, J. (2024). Co-creating transformations towards sustainability: Embracing innovations, honouring traditions. *Discourse and Communication for Sustainable Education, 15*(1), 1-4. https://doi.org/10.2478/dcse-2024-0001

Holst, J., Singer-Brodowski, M., Brock, A., & de Haan, G. (2024). Monitoring SDG 4.7: Assessing education for sustainable development in policies, curricula, training of educators and student assessment(input-indicator). *Sustainable Development*, *32*(4), 3908–3923. https://doi.org/10.1002/sd.2865

Imeson, P. (2017). Improving student learning through the use of multiple-choice questions. *Teaching Business & Economics, 21*(2), 18-19.

Kalas, P., O'Neill, A., Pollock, C., & Birol, G. (2013). Development of a meiosis concept inventory. *CBE Life Sciences Education*, *12*(4), 655–664. https://doi.org/10.1187/cbe.12-10-0174

Kolomitro, K., MacKenzie, L. W., Lockridge, M., & Clohosey, D. (2020). Problem-solving strategies used in anatomical multiple-choice questions. *Health Science Reports 3*(4), 209. https://doi.org/10.1002/hsr2.209

McMillan, J.H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34–43. https://doi.org/10.1111/j.1745-3992.2003.tb00142.x

Miller, D., Linn, R., & Gronlund, N. (2013). *Measurement and assessment in teaching* (11th ed.). Pearson Education.

Nehm, R.H., & Schonfeld, I.S. (2008). Measuring knowledge of natural selection: A comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching, 45*(10), 1131-1160. https://doi.org/10.1002/tea.20251

Newman, D. L., Snyder, C. W., Fisk, J. N., & Wright, L. K. (2016). Development of the Central Dogma Concept Inventory (CDCI) assessment tool. *CBE-Life Sciences Education*, *15*(2), 9. https://doi.org/10.1187/cbe.15-06-0124

Ong, S. L. (n.d.). *Profiling classroom teachers' assessment practice*. www.iaea.info/documents/paper_4d32f2cd.pdf

Pierce, C., & Hemstock, S. (2021). Resilience in formal school education in Vanuatu: A mismatch with national, regional and international policies. *Journal of Education for Sustainable Development*, *15*(2), 206–233. https://doi.org/10.1177/0973408 2211031350

Popham, W. J. (2001). Teaching to the test? *Educational Leadership, 58*(6), 16-20

Roczen, N., Fischer, F., Fögele, J., Hartig, J., & Mehren, R. (2021). Measuring system competence in Education for Sustainable Development. *Sustainability*, *13*, 4932. https://doi.org/10.3390/su13094932

Roure, B., Anand, C., Bisaillon, V. & Amor, B. (2018). Systematic curriculum integration of sustainable development using life cycle approaches: The case of the civil engineering department at the Université de Sherbrooke. *International Journal of Sustainability in Higher Education*, *19*(3), 589-607. https://doi.org.rproxy.uwimona.edu.jm/10.1108/IJSHE-07-2017-0111

Salīte, I., et al. (2024). Editorial: Kaleidoscope of sustainable practices: Developing minds and shifting patterns, *Journal of Teacher Education for Sustainability, 26*(1), 1-4. https://doi.org/10.2478/jtes-2024-0001

Schilling, J. F. (2019). Assessing the application of knowledge in athletic training: extended-matching questions. *Athletic Training Education Journal*, *14*(1), 80– 89.

Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation*, *22*(4), 1-13. doi: https://doi.org/10.7275/swgt-rj52

Sterling, S. (2016). A commentary on education and sustainable development goals. *Journal of Education for Sustainable Development*, *10*(2), 208–213. https://doi.org/10.1177/0973408216661886

Tamir, P., (1991). Practical work in school science: An analysis of current practice. In Woolnough, B.E., (ed.). *Practical science.* Milton Keynes: Open University Press.

United Nations. (2015). Transforming our world: The 2030 agenda for sustainable development. Resolution Adopted by the General Assembly on 25 September 2015, 42809, 1–13. https://doi.org/10.1007/s13398-014-0173-7.2

United Nations Department of Economic and Social Affairs. (2018*). The Sustainable Development Goals Report 2018. United Nations.* https://unstats.un.org/sdgs/files/report/2018/TheSustainableDevelopmentGoalsReport2018-EN.pdf

United Nations Educational, Scientific, and Cultural Organization Global Action Programme on Education for Sustainable Development (2018). *Education for sustainable development and the SDGs: Learning to Act, Learning to achieve. Policy Brief Advancing ESD Policy.* https://en.unesco.org/sites/default/files/gap_pn1_-_esd_and_the_sdgs_policy_brief_6_page_version.pdf

United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2017). *Education for sustainable development goals; Learning objectives*. Paris, France: UNESCO.

United Nations Educational, Scientific, and Cultural Organization (UNESCO) (2019). Proposal for monitoring of SDG indicators 4.7.1, 12.8.1and 13.3.1. TCG6/REF/14. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2019/08/TCG6-REF-14-Proposal-for-monitoring-of-SDG-indicators-4.7.1-12.8.1-and-13.3.1.pdf

Weber, N.R., Strobel, J., Dyehouse, M.A., Harris, C., David, R., Fang, J., & Hua, I. (2014). First-year students' environmental awareness and understanding of environmental sustainability through a life cycle assessment module. *Journal of Engineering Education*, *103*(1), 154–181. doi: 10.1002/jee.20032.

Williams-McBean, C. T. (2018). Implementing the school-based assessment in English in Jamaica: Teachers' perception, preparedness and challenges. *Journal of Development and Education in the Caribbean, 17*(2), 67-101

Williams-McBean, C. T. (2020). Formative assessment in Jamaican classrooms: Towards a model for more effective implementation. [Abstract]. *The Caribbean Journal of Mixed Methods, 1*(1), 193–194.

Williams-McBean, C.T. (2022). Assessment tools and strategies used by Jamaican teachers. *International Journal of Assessment Tools in Education, 9*(4), 883–905 https://doi.org/10.21449/ijate.980870

Williams-McBean, C. T. (2023). Using school-based assessments to advance the integration of sustainable development competences by capitalising on the practice of teaching to the test, *Environmental Education Research*, *29*(5), 715–732. doi: 10.1080/13504622.2022.2107616

World Commission on Environment and Development. (1987). *Our common future. Report of the World Commission on Environment and Development*; Also known as the Brundtland Report, Oxford University Press, New York, http://www.un-documents.net/ocf-ov.htm

Correspondence concerning this article should be addressed to the corresponding author, Clavia Williams-McBean, The School of Education, The University of the West Indies, Mona Campus, Jamaica. Email: claviawilliams@yahoo.com