Yann Shiou Ong
Timothy Ter Ming Tan
Yew-Jin Lee *Editors*

A Diversity of Pathways Through Science Education

Springer

A Diversity of Pathways Through Science Education

Yann Shiou Ong · Timothy Ter Ming Tan ·
Yew-Jin Lee
Editors

A Diversity of Pathways Through Science Education

Springer

Editors
Yann Shiou Ong
National Institute of Education
Natural Sciences and Science Education
Nanyang Technological University
Singapore, Singapore

Timothy Ter Ming Tan
National Institute of Education
Natural Sciences and Science Education
Nanyang Technological University
Singapore, Singapore

Yew-Jin Lee
National Institute of Education
Natural Sciences and Science Education
Nanyang Technological University
Singapore, Singapore

ISBN 978-981-97-2606-6 ISBN 978-981-97-2607-3 (eBook)
https://doi.org/10.1007/978-981-97-2607-3

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Introduction

ISEC 2020 was intended to be *the* conference for showcasing research and new thinking in science education. Organized once every couple of years by staff from the Natural Sciences and Science Education academic group at the National Institute of Education, Singapore, its opening was initially planned for the middle of 2020 to coincide with the mid-year or summer holidays for local and international audiences. The International Science Education Conference (ISEC) organizing committee expressed high hopes for at least two reasons: (i) STEM education was to assume an equally prominent theme during this conference, and (ii) ISEC-STEM 2020 aspired that its attendees experience fresh insights on current/future trends and needs in these domains arising from 20-20 vision in this auspicious year. As such, the overall conference theme was entitled *The Tango between Science and STEM* to reflect these aforementioned ideas. It was described on the official website as being able to

> reflect a dance the S-T-E-M education researchers are immersed in as they crossover into interdisciplinary research. Tango also encapsulates the synergy between Science and STEM as Science continues to play a prominent role in STEM education. One of the disruptions to science education as a field is the increasing emphasis on integrated STEM education. With science as the discipline that is currently dominant in integrated STEM, it is strategic that we position ISEC 2021 and STEM 2020 as two related conferences. This will encourage scholars from both fields to interact and develop synergies to move the knowledge forward. (from the conference website)

A disruption of immense proportions did indeed occur although not in the way that the organizers had anticipated because the COVID-19 pandemic plunged most of the world including Singapore into disarray soon after the winter of 2019. This global event scuttled our conference planning resulting in ISEC-STEM being delayed for a year to the summer of 2021 as well as being conducted virtually. STEM 2020 was also decoupled from ISEC 2021 due to various considerations by the organizers. These changes were not as bad as was thought, for now the problems of high registration fees as well as long-distance travel woes were overcome at one stroke though not the issue of participating in real-time across very different time zones. The ISEC 2021 conference was nonetheless successful under these very difficult circumstances with 77 papers/symposia presented by researchers and teachers from 17 states/regions. As with previous ISEC, presenters from Singapore occupied the lion's share of presentations. What readers therefore see in this edited book are a sampling of invited authors who had presented at ISEC 2021.

This book is organized into three parts.

1. Part I: Questions and Questioning in Science/STEM Education
2. Part II: Developing Science Teaching and Assessment
3. Part III: History, Philosophy and Sociology of Science/Engineering and Informal Learning

To summarize, Part I features three chapters foregrounding the epistemic practice of student questioning across grade levels. Part II is hugely diverse in its coverage with five chapters describing different aspects of teaching, learning, and assessment from multiple theoretical standpoints while Part III comprises three chapters that also appear to be very diverse but can be seen as takes on the history and/or development of formal and informal learning in science and engineering. The beginning of each part is accompanied by a Commentary written by each of the book editors who were members of the ISEC organizing committee.

We wish to bring to the reader's attention a unique feature of this edited book: At the end of a chapter, each set of authors has written a "Note to Future Colleagues" to describe their aspirations for the state of science/STEM education research in 2050 in the research area reported in their chapter. Collectively, these "Notes" point toward potential directions that science/STEM education research could take to achieve the espoused visions by the middle of the twenty-first century. On this note (pun intended), the editors of this book would like to end this introduction with our own Note to Our Future Colleagues in 2050.

Note to Our Future Colleagues

It is the year 2050. According to authors Christiana Figueres and Tom Rivett-Carnac of the book *The Future We Choose*, if the world worked together and took appropriate actions to avert the climate crisis, we would be on track to warm by no more than 1.5 °C by the year 2100 (Figueres and Rivett-Carnac 2020). Thus, our future looks bright and science/STEM education continues to be an important part of our culture. Here, we present our three predictions made in the 2020s for science/STEM education in 2050.

I. Epistemic Practices of the Future

By 2050, disciplinary boundaries have blurred as most professionals work in interdisciplinary teams to solve the complex real-world problems. Using the language of 2020s, most professionals and experts are trained in one or more of the "traditional disciplines" such as the sciences or humanities and have working knowledge of several other disciplines. Thus, educators and education researchers have progressed from focusing on disciplinary to interdisciplinary epistemic practices. As working in large interdisciplinary teams is the norm of work in 2050, interpersonal or "soft" skills such as collaboration, communication and empathy have become just as important as content knowledge and procedural knowledge. In addition, engagement in epistemic practices has become part of school norm (though schools no longer take the same form as they did in the 2020s). The ability to critique, construct, and discern trustworthy knowledge claims is now essential to everyday living since most people have become content creators as well as content consumers. Since everyone can find a platform to publish their views, some of which are erroneously/intentionally positioned as truths, it has become challenging—yet part of everyday life—for the layperson to discern trustworthy and sound claims from unwarranted ones, including scams. Incidentally, attempts by large-scale social media platforms and governments to curate online information have failed and thus, the onus of fact-checking remains on the individuals.

II. Science Teaching, Learning, and Assessment in the Age of AI

In the coming three decades, science teaching and learning will become both easier and harder. It will seem easier because so much more will be known about the overall principles of how human cognition functions in the service of acquiring valuable knowledge. The main theories or frameworks will have been mapped out regarding how cognition is dependent on one's internal architecture of neurons as well as the body's engagement with its contextual surround. On the other hand, how cognition interacts with other aspects such as the physical body and its observable states known as emotions or affect will have complexified the fine details of how people learn and behave. Besides, schooling will now be augmented by forms of artificial minds or intelligences in a similar manner as how tools of the past (e.g., the abacus, counting rods, log tables, calculators, the Internet) had assisted classroom learning. The nature and target of assessment too will likely change drastically as mentioned in the commentary for Part II. So despite knowing more than ever about how people learn and in possession of unimaginable new technologies, science teachers in the middle of the century will still have their work cut out for them as they facilitate students in much more demanding tasks (i.e., see the above epistemic decision making) compared to previous eras. A teacher's life will probably remain just as demanding rather than easier and the space-age life with robot teachers imagined in the US cartoon *The Jetsons* will not materialize.

III. Schools of the Future

The COVID-19 pandemic was the trigger that kickstarted what will likely be the biggest evolution in schools since the widespread adoption of formal schooling during the first Industrial Revolution, and that 30 years hence, its repercussions continue to shape and redefine what "schools" are. The lockdowns and restrictions imposed certainly advanced the pervasive use of and drove the development of technologies for remote learning and remote work. More significantly, these affordances for and the experience of remote learning normalized it as another mode of "school". The evolution began with home-based learning being instituted on a routine basis, where students stay home one day a month or fortnight, mainly as a way to familiarize students and teachers with remote learning as a contingency against subsequent pandemic-induced disruptions as has already happened in Singapore during the pandemic (Tan and Chua 2021). We see schools increasingly becoming delocalized in subsequent decades, where students can connect to their classes remotely or attend in-person as circumstances or preference dictate. Since the home environment may not always be the most conducive for learning, co-learning spaces will become commonplace. Modeled after coworking spaces where companies lease office space or traveling workers rent a desk for a day or two, students settle into an individual "pod" or as a small group of peers in mini classrooms to attend their lessons on the other side of town, the country, or the world. Students from small remote communities or those from impoverished neighborhoods can receive a quality education at bigger schools without geographical constraints. Inter-school collaborative learning, up to and including those at an international level, is not uncommon and allows for cross-cultural learning that promotes pluralistic understanding and empathy. The co-learning spaces may blend formal and informal learning opportunities, synchronous and asynchronous learning modalities, as well as provide the socialization and interaction among peers that home-based learning does not. As mentioned above, the demands on teachers in 2050 won't be easy. Technologies would certainly have been developed to facilitate and enable more naturalistic remote presence and interactivity between teachers and learners. Of particular relevance to us would be the ways in which science practical work might be conducted in such settings. Perhaps such co-learning spaces would be co-located with community libraries and science centers/museums to jointly form satellite venues for both formal and informal learning. But most importantly for such *delocalized schooling* to have happened successfully, researchers and practitioners must have studied, developed, and refined the pedagogies and management techniques for hybrid classes where some students are physically present while some are remotely connected.

While all these prospects may be futuristic to us in the 2020s, we are confident that you, our colleagues in educational research, have continued to study teaching and learning as a learner-centered and hence human endeavor.

References

Figueres C, Rivett-Carnac T (2020) The future we choose: surviving the climate crisis (1st edn). Alfred A. Knopf

Tan OS, Chua JJE (2021) Science, social responsibility, and education: the experience of Singapore during the COVID-19 pandemic. In: Reimers FM (ed) Primary and secondary education during Covid-19. Springer

Contents

Chapter 7
The Cognitive Demands of Secondary Science Assessment Items: Refinements to a Classification Based on Semantic Gravity and Density

Ning Charlotte Seah, Yew-Jin Lee, and Yann Shiou Ong

7.1 Introduction

All educators, not just in science education, have struggled with evaluating the cognitive demands of assessment items in valid and reliable ways that are also easy to use (Schneider 2014; Porter 2002; Penuel and Shepard 2016; Waddington et al. 2007). These endeavours have led to the widespread reliance on frameworks such as revised Bloom's Taxonomy (RBT), Webb's Depth of Knowledge, and the SOLO Taxonomy among many others (see Anderson et al. 2001). It goes without saying that there have always been disadvantages, criticisms, and trade-offs associated with any method of classification even as there are also vocal advocates to be found for each. Given the interest and practical significance of such categorisation schemes for science teaching, it is expected that research efforts here will be sustained in the near and distant future. Indeed, the stakes are high for governments and international organisations too: The highly influential OECD Programme for International Student Assessment (PISA) in 2018 administered to 15-year-olds from 79 regions and economies adapted Webb's Depth of Knowledge to determine the cognitive demands of their pool of science questions (OECD 2019).

This paper adds to the literature by refining a complimentary coding scheme derived from Legitimation Code Theory (LCT), specifically from its Semantics dimension (Maton 2014; Maton et al. 2017). While developed only within the last two decades, LCT—specifically its Semantics dimension–has recently been utilised

N. C. Seah
College of Humanities, Arts and Social Sciences, Nanyang Technological University, Nanyang, Singapore

Y.-J. Lee (✉) · Y. S. Ong
National Institute of Education, Nanyang Technological University, Nanyang, Singapore
e-mail: yewjin.lee@nie.edu.sg

Y. S. Ong et al. (eds.), *A Diversity of Pathways through Science Education*,
https://doi.org/10.1007/978-981-97-2607-3_7

to assist educators in determining the cognitive demands of science assessment items based on their levels of (i) context-dependency (from semantic gravity [SG] codes), and (ii) complexity (from semantic density [SD] codes) of ideal/expected student responses. Our proposal has expanded in a major way Rootman-le Grange and Blackie's (2018) seminal Semantics coding of undergraduate chemistry examination items by creating a more flexible classification that can also be applied across various science disciplines at secondary school levels. Indeed, we are confident that our revised coding scheme is potentially applicable for similar test formats in science at other educational levels such as in primary, middle-school, senior high, and university courses. Hence, we believe that it is a worthwhile addition to the suite of existing frameworks to gauge the cognitive or intellectual demands of assessment questions in science. In the following sections, we describe the theories and concepts behind LCT and Semantics before justifying with concrete exemplars how we reconceptualised this method with four levels of context-dependency and complexity respectively for categorising secondary science assessment items.

7.2 Theoretical Framework

7.2.1 Legitimation Code Theory and Semantics

Legitimation Code Theory (LCT) is an evolving theoretical framework used to analyse the process of knowledge building within social practices, that is, what kinds of talk, meanings, actions or behaviours (i.e. codes) are regarded as correct, appropriate, or legitimate for a specific social context (Maton 2014). From its roots in sociology and Systemic Functional Linguistics, LCT has now been applied beyond education to enquire how learning and expertise develops in other social fields such as law, architecture, business, and the fine arts (Martin et al. 2020). Research in LCT is moreover motivated by social justice principles due to its fundamental concern with an individual's possession as well as epistemic access to important knowledge in a community of practice (Maton et al. 2017). This access is dependent upon knowing or being able to perform different kinds of legitimation codes, that is, what is deemed to be correct, proper or acceptable within a certain social practice. At present, theory-building with LCT comprises three main dimensions—Specialization, Autonomy, and Semantics [see LCT (n.d.) for more information]—but only the Semantics dimension will be discussed here as it directly involves the classification and organisation of knowledge, and is utilised in this coding scheme for science assessment items.

Within the Semantics dimension of LCT, there are two kinds of domains: Semantic gravity (SG) refers to the degree of context-dependency of meaning in ideas/concepts/practices while semantic density (SD) refers to the degree of condensation or complexity of meanings in the latter (Maton 2014). In a nutshell, meanings that code strongly for semantic gravity would imply that they are very concrete or dependent on the immediate context for sense-making. In our case, a science question with stronger semantic gravity demands a response that is more closely tied to the context presented in the item, such as making use of information described or shown in the question stem (e.g., state the boiling point of a liquid based on a given heating curve), and thus, would not be sensible or answerable beyond this narrowly-defined context. Conversely, a question of weaker semantic gravity requires a more generalised response that would apply to or draw on ideas from many different contexts/domains or sections within the syllabus (e.g., explain the effects of heating a liquid, which can theoretically involve various concepts in chemistry or physics at different degrees of depth). Thus, semantic gravity helps analysts describe the degree of context-dependency or concreteness of meanings. On the other hand, semantic density characterises the strengths or levels of complexity of meanings. A question of stronger semantic density involves more complex meanings, which students are likely to utilise more cognitive effort in unpacking them (e.g., a question that involves multiple logical reasoning phases and/or multiple problem-solving steps to generate an acceptable response). In contrast, a question of weaker semantic density involves less complex meanings; acceptable solutions can be generated through fewer logical reasoning phases or problem-solving steps.

7.2.2 Comparisons with Other Frameworks of Cognitive Demand

Important to note here is that whether an assessment item is considered to have strong or weak SG/SD does depend on which items are being compared, the contexts of comparison (e.g., the grade level), and of course a coder's background (Blackie 2014; Maton 2014). Although SG/SD codes can vary along a continuum of strengths (they are not dichotomous codes i.e. yes/no), we adapted four levels typical of many LCT research studies such as Rootman-le Grange and Blackie (2018) to classify the degrees of context-dependency and complexity respectively of science assessment items. In this respect, research using LCT allows, in fact, requires a principled translation process to adapt abstract theories/concepts to cater for the specific purposes of the research study. By so doing, this process will facilitate iterative movements between data and theory to "simultaneously adjust the theoretical framework to the data in question and, in turn, to read the data through the theoretical lens" (van Heerden 2020, p. 3). While our coding scheme builds on past LCT-based research, we have operationalized the concepts from Rootman-le Grange and Blackie (2018) more concretely and thus refined them in a major way. At this point, we briefly justify our rationales for employing LCT and Semantics as a complimentary way to code for cognitive demands of science assessment items before going into detail about the development of our coding scheme.

Unlike the popular revised Bloom's Taxonomy (RBT), this Semantics coding scheme does not oblige a predetermined range of cognitive processes and knowledge domains involved in answering a question (Anderson et al. 2001; Krathwohl 2002; Martin et al. 2020). Both SG and SD codes can thus be “fit-for-purpose” through disciplinary and/or pedagogical norms or other criteria that are deemed relevant by the researchers. In RBT, there are suggested pairings of cognitive processes and knowledge domains (Remembering: Factual, Understanding: Conceptual, Apply: Procedural) although no such expectations are necessary when using Semantics codes, which is a distinct advantage. While RBT has potentially 24 code pairings from its twin dimensions, our coding scheme is not far behind with 16 potential pairings from two types of code relations.

Another popular assessment framework is the Structure of Observed Learning Outcomes (SOLO) Taxonomy that ascertains how well learners understand a topic (Biggs and Collis 1982). Like RBT, SOLO Taxonomy assumes a linear hierarchy of levels of understanding (i.e. cognitive complexity) from unistructural to extended abstract (prestructural understanding can be discounted as this marks an absence of knowledge). What distinguishes it from RBT is that SOLO underscores the visible, observed expression or enactment of knowledge as demonstrated by learners in tasks similar to what we are also attempting to do. There is much to commend in the SOLO Taxonomy not least because it offers both a theory of teaching and of learning as well as reportedly being able to distinguish true item complexity from difficulty (Hattie and Brown 2004). On the other hand, we argue that the 16 code combinations of context-dependency and complexity when using Semantics can possibly afford a far broader range and scope to evaluate cognitive demands of assessment items than just the four levels in SOLO. There also have been some criticism of the usage of verbs deemed helpful in guiding learning in the SOLO levels; some verbs can belong to more than one level and are thus more ambiguous than they appear (Brabrand and Dahl 2009).

A simplified though more operational version of the SOLO Taxonomy is the Depth of Knowledge (DOK) (OECD 2017) framework developed by Norman Webb (1997). In DOK, cognitive demand is determined by levels of depth that consider both the content and cognitive process to complete a task from start to finish. The four levels here include: recall and reproduction, utilizing skills and/or conceptual knowledge, strategic thinking (i.e. short-term use of higher order thinking processes), and extended thinking (i.e. extended use of higher order thinking processes) (Webb’s Depth of Knowledge Guide 2009). However, it is apparent that similar issues concerning utilisation of the SOLO Taxonomy applies to DOK as well. As mentioned earlier, it is true that every existing classification method of cognitive demand has their own promotors and detractors. We are therefore not claiming that our coding scheme can supplant or is far superior to other frameworks. Instead, we are suggesting that based on ideas about context-dependency and complexity, it can offer a complimentary and relatively straightforward method to assess the cognitive demands of assessment items in science education.

7.2.3 Deriving the SG/SD Coding Scheme

Educational research studies that have used Semantics have typically reported how learning can be enhanced with instruction spanning a spectrum of context-dependency/abstraction as well as complexity in a course (e.g., Georgiou 2016; Mouton and Archer 2018; Rootman-le Grange and Blackie 2020). With respect to using LCT and Semantics in examining educational assessment and cognitive demands, research here has been limited. Georgiou, Maton, and Sharma (2014) attempted to use Semantics (only semantic gravity codes) to explore the range of context-dependency in students' responses in a university physics course that Steenkamp et al. (2019) later followed. More recently, Lee and Wan (2022) have used Semantics as a means of classifying how abstract or concrete (i.e. its cognitive demands) are science learning outcomes in the intended curriculum in schools. To assess the cognitive demands of science assessment items, this present study adapted and extended earlier work by Rootman-le Grange and Blackie (2018), which was a seminal assessment-focused study using LCT and Semantics.

In parallel with Rootman-le Grange and Blackie (2018) and Lee and Wan (2022), semantic gravity (SG) and semantic density (SD) are differentiated into four strengths/levels ranging from SG−− and SD−− being the lowest to SG++ and SD++ being the highest strengths respectively. Thus, an item coded as SG++ is more context-dependent and demands a response that is closely tied to the context presented in the item whereas an SG−− item is far less dependent on the question context. An SD++ item involves unpacking of complex meanings and demands more logical reasoning phase and/or multiple steps to generate a suitable explanation or solution as an appropriate response whereas an SD−− item does not even require relevant domain knowledge to answer it well.

Again, it is critical to note that the actual coding of items with Semantics may vary according to the context such as the grade level of the assessment items. For instance, the term "energy" may elicit more concrete or less complex acceptable responses for lower grades while graduate courses in physics may expect the same term to encompass many nuanced meanings that link multiple concepts in/across the discipline. As such, all SG and SD codes are "neither definitional nor definitive" and depend on the study context (Maton 2019, p. 3). While it may seem frustrating in that there appears to be no fixed points of reference, it may be seen as providing room to manoeuvre in the iterative process of specifying the various levels of coding for teachers.

Our proposed classification scheme targets a specific format of science assessment item common at the secondary level (from Grades 7 to 9/10) around the world, namely, those that are variously called restricted, structured, short-answer completion/supply type items. Belonging to the larger family of constructed-response test formats, we focused on items that only required brief answers (words, phrases, few sentences) that rarely exceeded four marks for each question (Frey 2018). We analysed the viability of our coding scheme based on published science questions drawn from past Singapore-Cambridge General Certificate of Education (GCE) O-level examinations across the traditional disciplines of physics, chemistry and biology. The O-level examinations are taken by 15–16 year olds at the end of Grade 10, which is typically the terminal year of secondary education in Singapore. Note that even though a question in the GCE O-level examination will have multiple sub-questions within it (e.g., Question 1a, 1b, 1c), we coded all these sub-questions as individual items. This allowed us to analyse the specific cognitive demand of each sub-question as the entire question may code for diverse and contrasting Semantics levels. Based on this approach, we therefore sampled a total of 55 items (18 physics, 23 chemistry, 14 biology) published across the years 2014–2017.

7.2.4 Deriving the Four Levels of SG Codes for Context-Dependency

We used Rootman-le Grange and Blackie's (2018) coding scheme as a basis of our work, but it is important to highlight some major differences between them. As mentioned, the earlier authors set to assess the quality of assessment items to engage meaningful learning of university chemistry whereas we wished to evaluate the cognitive demands of items across the sciences for secondary school students. They sampled 44 assessment items from one examination paper while we based our coding scheme on 55 GCE O-level items from 2013 to 2017 (4 years). Lastly, our coding scheme attempted more detailed and elaborated criteria that considered the potential awarded marks for questions coded in the different levels as well as associated keywords/command verbs that may appear in the questions for the SD coding scheme. We thus sought to be more precise in the descriptions for the expected responses detailing the various code strengths to improve decision-making. For instance, the description for SD+ in Rootman-le Grange and Blackie's (2018) coding scheme reads, "The given information needs to be manipulated—unpacked before it can be interpreted." Such a description is likely to be open to the interpretation of the coder as it is not clearly defined what the manipulation of the given information entails. In contrast, while our coding scheme for SD+ (explained later) also requires the processing or manipulation of given information, we expanded on this idea by stating that students are expected to "recall formulae/equation/knowledge and interpret/explain values/structures/information from question in multiple steps" and that "it is easy to derive the requirements of the question straightaway".

Table 7.1 below shows the four strengths of SG codes for context-dependency that we refined based on ideal/expected responses to secondary science assessment items (some of which are hypothetical), which we now explain in turn. Questions coded as SG−− are the most abstract and require responses that embody knowledge from multiple disciplines not solely limited to science in order to give a full, comprehensive answer. For example, if asked to explain the effects of heat on a metal, a complete response should include ideas about chemical and physical effects with considerations for various states of the metal that will draw on a large spectrum of concepts in the physical sciences. Next, questions coded as SG− require responses from only one major scientific discipline, but from more than one sub-section of the official syllabus document. For example, a question on comparing how heat versus electric current is transferred through a metal would require drawing on more than one sub-section of the physics syllabus because responses should include descriptions of thermal conduction utilising the particulate model of matter and the movement of charged particles under the influence of electric fields.

We acknowledge that it may be difficult to distinguish topics in the Physical Sciences as belonging strictly to chemistry or physics, but most contemporary educational systems would have written documents that state the topical boundaries within a discipline. We further recognise that different educational systems might categorise topics in very different ways, but this approach of relying on a standard/official syllabus document can offer some degree of coding consistency at least within one educational jurisdiction/district/school regardless how its science topics have been categorised. An issue might also arise when dealing with integrated science; the latter is widespread at primary and middle-school levels and consists of a few science disciplines mixed within one school subject. As such, it has sections and sub-sections that might follow arbitrary themes or grade and discipline divisions defined by respective educational institutions, which means that the SG coding for integrated science will look different from single discipline science tests. The coding scheme might also differ from country to country because of how their curriculum has been organised. Nonetheless, in all situations the SG coding will remain internally consistent for that particular context with our coding scheme.

Both SG+ and SG++ code for question responses that involve conceptual knowledge coming from a single sub-section of the syllabus. What distinguishes SG+ from SG++ questions is the extent to which contextual information in the question informs an acceptable response. Responses to SG+ questions typically require less reliance of contextual information from the question and instead, expects students to recall and apply previously learnt conceptual knowledge. For instance, the question "explain how heat is transferred through a metal rod" situates the explanation within the section of thermal physics. The context in this question ("a metal rod") provides little information (apart from the rod being made of metal) that students need to incorporate in their response. Thus, students could respond using the particulate model of matter and/or the role of delocalised electrons in the metal rod in the process of heat energy transfer. On the other hand, questions coded as SG++ strongly emphasise contextual details whereby it is mandatory for students to make use of and evaluate the information provided in the question stem. An example of such a question would be "Using the particulate model of matter, describe how does heat transfer from point A of the rod to its other end, point B." The question includes a diagram of a metal rod with a heat source at one end of the rod at point A in the same question. Consequently, students have to utilize the particulate model of matter in their answer, as requested in the question stem, and make explicit reference to points A and B of the rod in the given diagram. Thus, the expected response to the SG++ question is reliant on provided information. It should also be noted that no keywords or command verbs are highlighted for SG codes as the verbs used in

Table 7.1 SG codes for context-dependency with SG code descriptors from Rootman-le Grange and Blackie (2018) as comparison

SG code	From Rootman-le Grange and Blackie (2018)	SG code descriptors	Examples of questions
SG−−	Concepts situated in the curriculum are integrated with general everyday knowledge to create meaning that is applicable in any type of context	Scientific knowledge in the response can/must be drawn from **more than one discipline** (i.e. other disciplines of science or from other subjects)	Explain the effects of heat on a metal Explain how plants photosynthesize to create chemical energy
SG−	The question requires concepts from different sections in the curriculum to be integrated to create a unified theory that is applicable to a broader context	Scientific knowledge in the response is derived from **more than one sub-section** in the official **syllabus** document but still within one major scientific discipline	Compare how heat versus electric current is conducted through a metal rod Calculate the molecular mass of propane Explain how plants make food and how the energy generated from plants moves along a food chain
SG+	The question requires application of Chemical concept(s) from one section of the curriculum to a specific example	Scientific knowledge in the response is derived from a **specific sub-section of the official syllabus**	Explain how heat is conducted through a metal rod State one difference between alkanes and alkenes State one possible adaptation a plant may have to increase its rate of photosynthesis
SG++	The question is located in a specific section of the curriculum and only requires recall of the concepts, definitions or rules	Scientific knowledge in the response is derived from a **specific sub-section of the official syllabus** Response is highly **context dependent** i.e., requires specific information from the question stem to a significant extent	Using the particulate model of matter, describe how heat is transferred from point A to point B of the metal rod in the given diagram State the functional group of the organic compound shown in the molecular structure presented From the plant cell diagram, label which part(s) are responsible for photosynthesis and explain their functions

the question are not helpful in determining the level of context-dependency based on specificity of disciplinary knowledge drawn upon and the context-dependency of expected response. Thus, keywords are not considered for coding of context-dependency, but will be significant for the coding of complexity that we describe next.

7.2.5 Deriving the Four Levels of SD Codes for Complexity

Table 7.2 shows the descriptors for the four SD codes in our coding scheme for complexity based on ideal/expected response types by students. Science questions coded as SD−− are the least complex as they do not require any specialised domain knowledge to answer the questions. For example, a student specializing in biology or chemistry would still be able to answer a SD−− physics question adequately. If a question provides a physics formula for braking force, along with the values of the mass and acceleration of an e-scooter, a student should be able to substitute these values into the given formula in the question stem to obtain the correct braking force of the e-scooter. Hence, an average student at that grade level regardless of disciplinary background would be able to derive the answer easily given a supplied equation.

The difference in complexity between SD− and SD+/SD++ is that the latter group requires intermediate steps or more logical phases of reasoning i.e. unpacking of more condensed meanings to arrive at the final answer. However, answers for questions coded as SD− can be derived more directly since they do not require students to make any further interpretation or manipulation of the questions unlike SD+/SD++ questions. This is because there is only one scientific term/formula/structure/diagram/graph that needs to be interpreted/processed to answer the question. Another clue is that usually 1–2 marks are awarded in SD− questions and with certain command verbs such as using identify, define, or predict that are lower on the revised Bloom's Taxonomy. An example of a SD− question would be to calculate the braking force of an e-scooter given the e-scooter's mass and its acceleration. The braking force is directly derived by using Newton's Second Law of motion, which is a simple equation found in the physics syllabus.

Compared to SD+, SD++ coded questions are less straightforward and do not provide any hints on how to approach answering the question. For example, a SD+ version of the e-scooter question would require students to find the acceleration (given mass of the e-scooter) using a velocity–time graph before calculating the braking force. This question is relatively straightforward as the students can supply the answer assuming that the method has been already mastered. On the contrary, a SD++ level question would require students to first decipher the implicit demands and conceptual map of a particular question before answering it. Referring to the e-scooter example, a SD++ question could test students on the practicality of a specific speed limit on an e-scooter. The cognitive demands are now greater; students would first need to suggest a feasible mass of an e-scooter prior to calculating its acceleration using a

Table 7.2 SD codes for complexity with SD code descriptors from Rootman-le Grange and Blackie (2018) as comparison

SD code	From Rootman-le Grange and Blackie (2018)	SD code descriptors	Examples of questions
SD−−	No chemical terminology or concepts are required to answer the question	**Does not require content domain knowledge** (based on targeted "Subject Content"); students are only required to process or reason from the data/information presented Usually **1–2 mark question** Associated keywords: List, State, Plot, Recognize, Name	Using the expression: braking force = mass × deceleration, calculate the braking force acting on a 15 kg e-scooter as it decelerates at 0.9 m/s^2 Consider that the concentration of a solution is derived as the number of moles of solute divided by the total volume of solution. Calculate the concentration of aqueous sodium chloride if 0.5 mol of sodium chloride (solute) is dissolved in 100 dm^3 of water
SD−	Only one term/structure/formula is given and needs to be interpreted in order to answer the question	Using domain knowledge, only **one scientific term/ formula/structure/diagram/graph** needs to be **interpreted/processed** to answer the question It is **easy to derive the requirements** of the question Usually **1–2 mark question** Associated keywords: Identify, Define, Predict	Calculate the braking force acting on a 15 kg e-scooter as it decelerates at 0.9 m/s^2 Calculate the concentration of aqueous sodium chloride if 0.5 mol of sodium chloride is dissolved in 100 dm^3 of water Write out the equation for photosynthesis in plants in symbols

(continued)

Table 7.2 (continued)

SD code	From Rootman-le Grange and Blackie (2018)	SD code descriptors	Examples of questions
SD+	The given information needs to be manipulated—unpacked before it can be interpreted	**Manipulation** of **two or more terms/structures/ formulae/diagrams/graphs** based on domain knowledge required before interpretation It is **easy to derive the requirements** of the question, but multiple steps (defined as two or more phases) are required to derive the answer Involves **recalling** formula/equation/knowledge and **interpreting/explaining** values/structures/information from question in **multiple steps/**logical reasoning phases Usually **2–3 mark question** Associated keywords: Explain, Describe, Outline	Determine the braking force acting on a 15 kg e-scooter using its velocity–time graph (i.e., need to determine acceleration from graph) Calculate the concentration of aqueous sodium chloride if 50 g of sodium chloride is dissolved in 100 cm^3 of water. Express your answer in mol/dm^3 From the light intensity graph, describe how the rate of photosynthesis changes when light intensity increases?
SD++	The chemical problem must first be identified before any interpretation or manipulation can be done in order to get to a solution/answer to the question (multiple steps required)	A clear understanding of the ill-defined problem involving domain knowledge **must first be established before any interpretation** to formulate appropriate responses It is **not easy to derive the requirements** of the question, and multiple steps (defined as two or more phases) are required to derive the answer Involves **recalling** formula/equation/knowledge and **interpreting/explaining** values/structures/information from question in **multiple steps/**logical reasoning phases Usually question is **4 marks and above** Associated keywords: Calculate, Design, Analyse	Determine whether the speed limit of 25 km/h for e-scooter is scientifically reasonable Sodium chloride serves many uses in the real world, such as in nasal sprays and well drilling. Suggest why it is so versatile Explain photosynthesis and its importance in an ecosystem

suitable braking time and thus, determine its braking force. Finally, students would need to reason whether their calculations are realistic in the physical world, taking into account the safety of the rider and the surrounding road conditions. Therefore, such a question would be coded as SD++ due to its high complexity or condensation of the various scientific concepts and thinking processes involved. While SD+ and SD++ codes are close, additional clues come in the form of the number of marks that can be awarded based on the norms of the Singapore Cambridge O-level examination board as well as possible command verbs in the question stem as shown in Table 7.2.

Common to both SG and SD code descriptors, we wanted to code for and foreground features of the task in the assessment item rather than relying on inferred learner cognitive characteristics. Hence, guided by the Assessment Triangle (Pellegrino et al. 2001) and following Rootman-le Grange and Blackie (2018), our SG/SD coding avoided using the language of psychological constructs (in the Cognition vertex) that must be inferred or modelled. Instead, our codes were located at the Observation vertex to look for the performance of relevant and observable tasks/actions that learners must do to answer questions acceptably. While there is still an interaction between the learner and tasks, focusing on the nature of tasks (i.e. what must the learner be able to do) at the Observation vertex was felt to add greater objectivity and reliability in our coding scheme.

7.2.6 Illustrative Applications of the Refined SG and SD Coding Scheme

In this section, we provide worked examples of our coding scheme adapted from Rootman-le Grange and Blackie (2018) and articulate our decisions and rationales in classification. Three typical questions drawn from the major science disciplines for secondary students will be presented here. Due to copyright restrictions, these examples closely resemble official GCE O-level examination items from Singapore, but are not themselves the official questions.

a. **Physics**: An iron ring hangs from a string. A bar magnet is positioned close to the ring. The iron ring is then attracted to the bar magnet. Explain why.

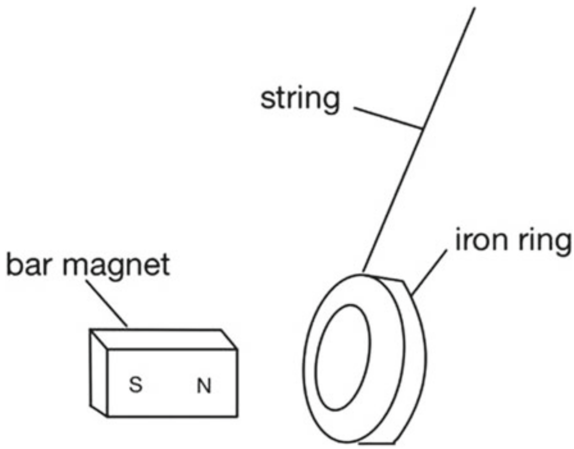
The domain knowledge of this question is taken from the sub-section of "Electricity and Magnetism" in the official physics syllabus document from Singapore. In terms of context-dependency, we have coded this question as SG++ indicating that it is concrete and very context-dependent. A good response needs a student to pay close attention to the given information above the diagram as well as within the diagram itself. When answering, a student needs to notice the material of the ring as well as which pole of the magnet is facing the ring. As the question includes Fig. 7.1, whereby the bar magnet is specifically labelled with north–south (N–S) polarity, the student is expected to deduce and mention the induced poles in the ring (i.e. induced S-pole on side closer to bar magnet) based on the polarity of the side of the bar magnet closer to the ring (i.e. N-pole). Thus, a response that only mentions "ring is attracted to the bar magnet as it is made of a magnetic material" is inadequate. Hence, this question is coded as SG++ as the required response is highly dependent on the context provided in the question i.e. the ring's material and the polarity of the magnet is crucial to the formulation of a complete and correct response. For complexity, we have categorised the question as SD+ as students have to "explain" the phenomenon of magnetic attraction (c.f. associated keywords for SD+, Table 7.2) in this 2-mark question. The answer involves more than one logical reasoning phase. First, students need to state that the ring is made of a magnetic material. Second, students need to state that the side of the ring closer to the north pole of the magnet is induced as a south pole and it is attracted to the north pole of the magnet given the property that opposite magnetic poles attract.

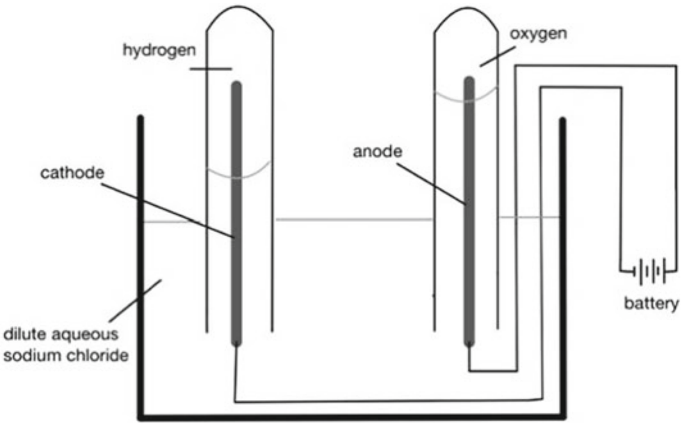
b. **Chemistry**: The diagram below depicts an electrolysis set-up. Dilute aqueous sodium chloride forms hydrogen and oxygen through electrolysis.

This question tests content from the section of "Chemistry of Reactions" in the official chemistry syllabus document in Singapore. Prior to this sub-question, students were required to state the ionic equations for reactions occurring at the anode and cathode, essentially demonstrating the electrolysis of water. With respect to context-dependency, this sub-question is coded as SG++ as students may need to refer to the ionic equations that they have written in the previous part to show the theoretical volume ratio of hydrogen and oxygen produced in the set-up. As students are required to refer to a previous part of the question, context is important in the formulation of their answers. In terms of complexity, we have coded it as SD+ as multiple steps, including making explicit reference to their answers in the previous question part are required to answer this 2-mark question. Through correctly interpreting the ionic equations for the reactions in the anode and cathode, students can then deduce the theoretical volume ratio of hydrogen to oxygen as 2:1 (Figs. 7.2 and 7.3).

c. **Biology**: The following diagram shows a germinated pollen grain.

Fig. 7.1 Example of a physics question involving bar magnet and iron ring





The two gases are collected and their volumes are then measured. Theoretically, the ratio of hydrogen to oxygen should be 2:1. Oxygen has a higher solubility in water than hydrogen. This causes a change in volume of the respective gas collected. Why is the theoretical ratio of hydrogen to oxygen 2:1?

Fig. 7.2 Example of a chemistry question on electrolysis of water

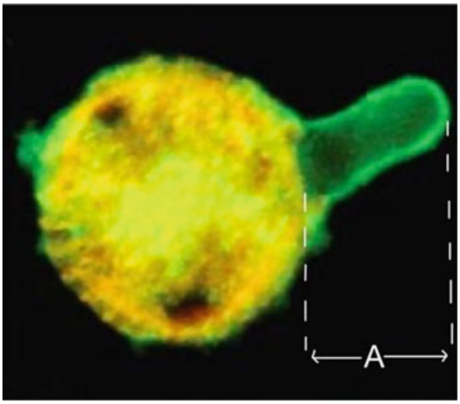


Fig. 7.3 Example of biology question on a pollen grain. *Note* From '*Germination pollen grain*' [Photograph], by AJC1 (2013), Flickr (https://www.flickr.com/photos/47353092@N00/10170683834). CC BY-SA 2.0

ci Define sexual reproduction

This sub-question references the biology syllabus section on "Maintenance and Regulation of Life Processes". For SG, this sub-question was coded as SG+ given that it is less context-dependent and only requires students to recall the definition of sexual reproduction without the need to make reference to any other information in the question. In terms of SD, we have coded the sub-question as SD+; while it is a simple 1-mark definition question (which would be SD− based on associated keyword alone), it requires several key phrases and concepts, warranting multiple logical reasoning phases.

cii. Name the part labelled A.

This sub-question was coded as SG++ because of its high context-dependency; only through examining the given diagram would students be able to identify the unknown part labelled A. It was categorised as SD− since students are simply required to identify the part labelled A without any manipulation of the information presented in this 1-mark question.

ciii. Describe **and** explain how the tip of part A reaches the female nucleus in a flower.

Similarly, we have coded the question as SG++ since students are still required to make reference to the part labelled A. It was coded as SD+ as there are multiple steps to the answer and students need to provide description along with their explanation to this 2-mark question. Both command verbs: "describe" and "explain" are associated with SD+. The requirements of the question are easily derived, however, as students have already identified the unknown part labelled A in the previous sub-question.

7.3 Empirical Analysis with the Coding Scheme

7.3.1 The Semantic Plane of SG and SD

Both SG and SD codes can be presented orthogonally on a Cartesian plane where they move independently along a continuum on a semantic plane. We can divide this plane into four quadrants to display four combinations of SG and SD code modalities as shown above in Fig. 7.4. This enables us to better visualise the meanings of coded items in terms of their context-dependency and complexity in a graphical form. Based on Fig. 7.4, Quadrant 1 depicts questions that are less context-dependent as well as more complex in nature. Quadrant 2 represents items that are also low in context-dependency (i.e., more abstract), but low in complexity. Question responses depicted in Quadrant 3 expect students' responses to be highly context-dependent (i.e., more concrete), yet not requiring any specific domain knowledge (i.e., low complexity). Finally, Quadrant 4 codes for responses that are the most context-dependent, and requires multiple logical thinking phases or problem-solving steps (i.e. high complexity) to derive appropriate responses.

Putting our coding scheme to the test, Fig. 7.5 depicts the distribution of SG/SD codes from our sample of 55 questions. Our data was therefore restricted to only Quadrants 3 and 4 with SG+ (14 items or 26%) and SG++ (39 items or 74%) codes for context-dependency, and SD− (27 items or 51%) and SD+ (26 items or 49%) codes for complexity. The majority of the questions (n = 21) were coded as SG++/SD+ followed closely by SG++/SD− codes (n = 18) while only five questions were coded as SG+/SD+. Being short-answer structured or supply type question items requiring specific answers to a question (usually based on a given stimulus), it stands to reason that this overall semantic profile was observed. These types of questions (mainly

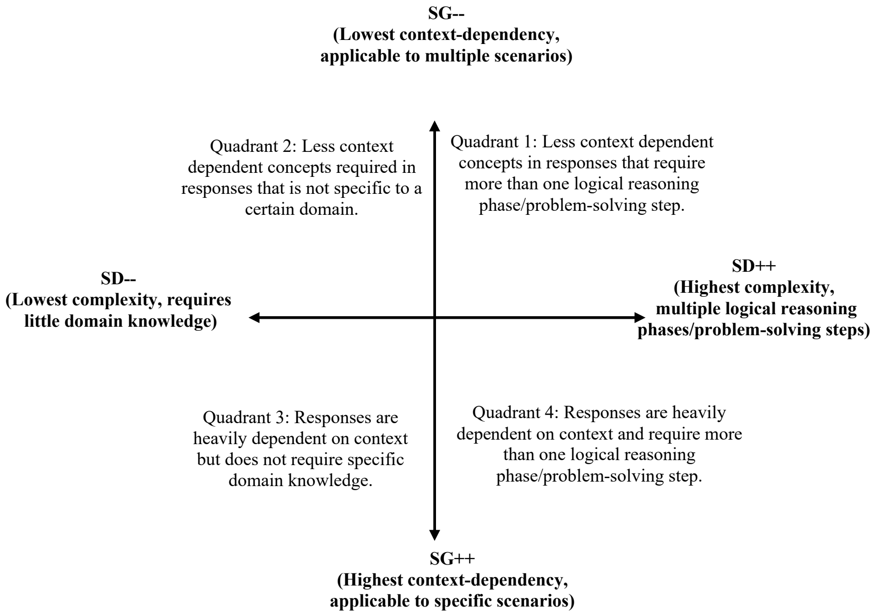


Fig. 7.4 The Semantic plane showing the four combinations of SG and SD codes in each quadrant

SG++ codes) therefore required a high degree of contextual information based on data in the question stem, which meant that students had to pay careful attention to the given “stem of information” (MOE, p. 5). Also, responses were drawn from just one specific sub-section of the syllabus with respect to these SG codes, that is, expected responses were narrowly focused in its conceptual spread. In terms of SD codes, it was also tightly clustered into two quadrants albeit achieving more balance in frequency between SD− and SD+ codes. In terms of complexity, it was therefore relatively easy to derive the requirements of the question through simple interpretation or a few steps of logical thinking/problem-solving that corresponded to a range of one to three marks for each question.

7.3.2 Inter-Rater Reliability with Weighted Kappa

To check the reliability of our SG/SD coding scheme, we utilized weighted kappa as a measure for inter-rater agreement for the 55 items coded. Kappa measures the proportion of agreement across multiple coders corrected for chance (Fleiss and Cohen 1973). The kappa scale ranges from −1 to +1 where a negative value implies a “poorer than chance agreement”, 0 implies an agreement that is “exactly chance” and a positive value implies a “better than chance agreement” (Fleiss and Cohen 1973, p. 613). Moreover, there is an implicit assumption that all disagreements are equally

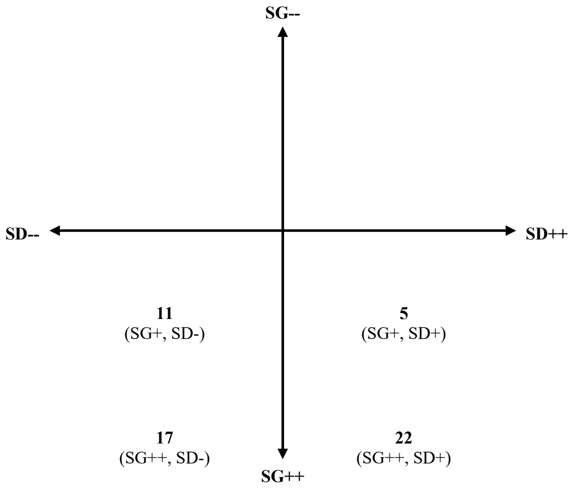


Fig. 7.5 Frequency distribution of coded GCE O-level items (n = 55) from Singapore in the three sciences in the semantic plane

significant and crucial. Since the four levels in our SG/SD coding schemes form an ordinal rating scale, we used weighted kappa so disagreements can be weighted by their magnitudes. In other words, further disagreements (e.g. question parts where one rater coded as SG++; the other as SG−−) were weighted less or indicates lesser agreement than near disagreements (e.g. question parts where one rater coded as SG+; the other as SG−). In this study, for SG codes the weighted kappa (linear weight) was 0.587 ($p < 0.001$) while for SD codes it was 0.608 ($p < 0.001$) among two coders (first and third authors). Hence, the weighted kappa values indicate moderate agreement for the SG codes and substantial agreement for SD codes (Fleiss and Cohen 1973). Our reported weighted kappa values reflect the reliability extent that can be reasonably achieved with approximately 20 h of training time.

7.4 Conclusion and Discussion

This study was primarily motivated by the longstanding search among educators for better tools to determine the cognitive demands of assessment items. We have therefore improved a complimentary coding scheme in Tables 7.1 and 7.2 from the Semantics dimension of LCT to assess science assessment items at the secondary school level. By adapting and expanding earlier work by Rootman-le Grange and Blackie (2018), we refined in a major way four levels of context-dependency and complexity based on semantic gravity and density respectively that yielded a total of 16 possible code combinations of cognitive demand based on Semantics. We are encouraged that the reliability of the coding scheme from its weighted kappa values showed moderate to substantial agreement based on our test sample of 55 test items from Singapore. The empirical analysis of these sample items in Fig. 7.5 showed that they were restricted to Quadrants 3 and 4, with the majority of items located at SG++ (highly content-dependent) while they seemed equally distributed between SD− and SD+ codes (medium complexity). As we have explained, this semantic profile seems reasonable given the nature of supply type item formats that required short answers by students.

Looking at the overall changes to the coding scheme first devised by Rootman-le Grange and Blackie (2018) in Tables 7.1 and 7.2, the differences are more pronounced with respect to SD than to the SG code descriptors. Our SD codes are more elaborate and show the various criteria in each level based the number of hypothetical steps in order to solve the problem, deployment (or not) of scientific knowledge, the use of keywords (which relies on Bloom's cognitive processes), and the allocated marks for that question. For SG codes, we paid attention now to the number of disciplines, location of the student answer in the (sub-) sections of the syllabus, and how context-dependent the answer is on the question. We believe that these criteria would provide classroom practitioners sufficient practical guidance to classify the cognitive demands of their assessment items according to the two dimensions of context-dependency and complexity in Semantics.

Our coding scheme has some potential drawbacks, which revolve around the uneven spread of codes displayed from our sample. As seen in Fig. 7.5, SG−− and SG- codes were absent in our trial coding, likewise for SD−− and SD++ codes. Nonetheless, because structured questions have very specific learning outcomes and are often based on a given stimulus, this might account for the restricted semantic profile that was observed here. Indeed, it is unlikely that these short answer supply questions would require answers that involve concepts distributed across multiple sections of the syllabus or even have multi-disciplinary answers at the O-level although such events are catered for by our coding scheme. It was also believed to be a chance phenomenon that SD−− codes were not observed during our trial coding for structured questions; these item formats were present in other O-level science questions outside of our study sample. While our trials only contained GCE O-level questions, we believe that this coding scheme is potentially applicable across other education levels as well as being relatively straightforward to implement given there are only two distinct dimensions (SG/SD) with four levels/strengths. We thus invite researchers interested in assessment to utilise and evaluate our coding scheme so as to further improve upon it. Other researchers interested in establishing relationships between item features and item difficulty could also extend the presented work in this direction.

7.5 Notes to Our Future Colleagues

The urge to categorize or rank what we humans deem as important, valuable, and meaningful has always been part of our social existence. Since time immemorial, questions have asked about who is the most beautiful, bravest, cunning, evil as well as practical questions of life regarding what is the strongest, the tastiest, the most durable and so on. Similar aspirations to classify and measure also exist within education domains with the desired attributes that we most typically value located in the cognitive domain. Certainly, we don't foresee research in this area to cease to you working in 2050! The reasons we think are simple: The work of education, for the most part, has been all about deliberating over issues of what is ability/competency and who has it. These twin, inseparable aspects of schooling have not diminished since the invention of schools writ large across world cultures. Indeed, our book chapter is no exception as we explore a different way of classifying intellectual challenges of examination test items by which one can then make valid inferences concerning "ability" in science education. We therefore feel that it is very unlikely the situation 30 years down will find anything drastic to displace this search for better definitions and measurements of competency/ability among learners.

Our message to our future colleagues is not, however, an unconditional surrender to doom because it is possible to get good education in an age of measurement (Beista 2011). We are optimistic that there will be a recognition for more 'authentic' or realistic measures of competency and knowing arising from schooling. This we hope can be coupled with greater acceptance that skill, knowledge or expertise are multi-faceted in nature and that schooling has been rewarding a narrow spectrum of human ability for too long. In the next decades, we are hopeful to dream of a broader range of desired outcomes from education such as empathy, care etc. that are valued, if not assessed in meaningful ways. In other words, there will be a wider appreciation of what constitutes intelligence and human thriving, a wider set of important skills and knowledge of educational purposes.

Moreover, we anticipate that the nature of assessment itself will be forced to undergo dramatic changes if what we read in the newspaper today (Chia 2023) is true concerning the rise of intelligent chatbots. The latter, it is said, can instantaneously compose texts and solve equations in human-like fashion thereby threatening the benefits that arise from "independent" work by students. Commentators, however, speculate that these AI-algorithms will exist alongside the process of learning (and testing), and teachers will instead be setting more assignments that test human creativity and application of knowledge in novel ways. These are the distinguishing features that separate machines from humans while others believe these AI applications are over-hyped. It now leaves you, dear colleagues, to decide if these brief thoughts concerning the end purposes of education, human learning, and assessment have been prescient or badly off-tangent, and write a book chapter about it!

Ethics Approval No ethics approval was sought as no human subjects were involved in the study.

References

AJC1 (2013) Germination pollen grain [Photograph]. Flickr. https://www.flickr.com/photos/47353092@N00/10170683834

Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE et al (eds) (2001) A taxonomy for learning, teaching and assessing. A revision of Bloom's taxonomy of educational objectives. Addison-Wesley Longman, White Plains, NY

Beista G (2011) Good education in an age of measurement: ethics, politics, democracy. Routledge, New York

Biggs JB, Collis KF (1982) Evaluating the quality of learning. Academic Press, New York

Blackie ALM (2014) Creating semantic waves: using legitimation code theory as a tool to aid the teaching of chemistry. Chem Educ Res Pract 15(4):462–469. https://doi.org/10.1039/C4RP00147H

Brabrand C, Dahl B (2009) Using the SOLO taxonomy to analyze competence progression of university science curricula. Higher Educ 58:531–549

Chia O (2023) Teachers v ChatGPT: schools face new challenge in fight against plagiarism. The Straits Times. https://www.straitstimes.com/tech/teachers-v-chatgpt-schools-face-new-challenge-in-fight-against-plagiarism

Fleiss JL, Cohen J (1973) The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. Educ Psychol Measur 33(3):613–619

Frey B (2018) Supply items. https://methods.sagepub.com/reference/the-sage-encyclopedia-of-educational-research-measurement-and-evaluation/i20333.xml

Georgiou H (2016) Putting physics knowledge in the hot seat: the semantics of student understandings of thermodynamics. In: Maton K, Hood S, Shay S (eds) Knowledge-building: educational studies in legitimation code theory. Routledge, Abingdon, pp 176–192

Georgiou H, Maton K, Sharma M (2014) Recovering knowledge for science education research: exploring the "Icarus Effect" in student work. Can J Sci Math Technol Educ 14(3):252–268

Hattie JAC, Brown GTL (2004) Cognitive processes in asTTle: the SOLO taxonomy. asTTle Technical Report #43. University of Auckland/Ministry of Education

Krathwohl DR (2002) A revision of bloom's taxonomy: an overview. Theory Pract 41(4):212–218. https://doi.org/10.1207/s15430421tip4104_2

LCT [Legitimation Code Theory] (n.d.) Theory. https://legitimationcodetheory.com/theory/

Lee Y-J, Wan D (2022) How complex or abstract are science learning outcomes? A novel coding scheme based on semantic density and gravity. Res Sci Educ 52(2):493–509. https://doi.org/10.1007/s11165-020-09955-5

Martin JR, Maton K, Doran YJ (2020) Accessing academic discourse: systemic functional linguistics and legitimation code theory. Routledge, Abingdon

Maton K (2014) Knowledge and knowers: towards a realist sociology of education. Routledge, London

Maton K, Hood S, Shay S (2017) Knowledge-building: educational studies in legitimation code theory. Routledge, London

Maton K (2019) Semantics from legitimation code theory: how context dependence and complexity shape academic discourse. https://legitimationcodetheory.com/wordpress/wp-content/uploads/2018/07/2019Maton-Semantics-intro.pdf

Ministry of Education [MOE] (2019) Physics. Singapore-Cambridge General Certificate of Education Ordinary Level (2021). Syllabus 6091. Singapore Examinations and Assessment Board

Mouton M, Archer E (2018) Legitimation code theory to facilitate transition from high school to first-year biology. J Biol Educ 53:2–20. https://doi.org/10.1080/00219266.2017.1420681

OECD (2019) PISA 2018 science framework. https://www.oecd-ilibrary.org/docserver/f30da688-en.pdf?expires=1625010963&id=id&accname=guest&checksum=994E1F816E11C9A91C3421AC0C48FD37

Organisation for Economic Co-operation and Development [OECD] (2017) PISA 2015 assessment and analytical framework: science, reading, mathematics, financial literacy and collaborative problem solving, revised edition, PISA, OECD Publishing, Paris. https://doi.org/10.1787/9789264281820-en

Pellegrino JW, Chudowsky N, Glaser R (eds) (2001) Knowing what students know: the science and design of educational assessment. National Academies Press, New York

Penuel WR, Shepard LA (2016) Assessment and teaching. In: Gitomer DH, Bell CA (eds) Handbook of research on teaching. American Educational Research Association, New York, pp 787–850

Porter AC (2002) Measuring the content of instruction: uses in research and practice. Educ Res 31:3–14

Rootman-le Grange I, Blackie MA (2018) Assessing assessment: in pursuit of meaningful learning. Chem Educ Res Pract 19:484–490. https://doi.org/10.1039/c7rp00191f

Rootman-le Grange I, Blackie MA (2020) Misalignments in assessments: using semantics to reveal weaknesses. In: Winberg C, McKenna S, Wilmot K (eds) Building knowledge in higher education: enhancing teaching and learning with legitimation code theory. Routledge, New York, pp 72–85

Schneider J (2014) From the ivory tower to the schoolhouse: how scholarship becomes common knowledge in education. Harvard Education Press, Cambridge

Steenkamp CM, Rootman-le Grange I, Müller-Nedebock KK (2019) Analysing assessments in introductory physics using semantic gravity: refocusing on core concepts and context-dependence. Teach Higher Educ 12:1–16. https://doi.org/10.1080/13562517.2019.1692335

van Heerden M (2020) (How) do written comments feed-forward? A translation device for developing tutors' feedback-giving literacy. Innov Educ Teach Int 58:555–564. https://doi.org/10.1080/14703297.2020.1788411

Waddington D, Nentwig P, Schanze S (eds) (2007) Making it comparable: standards in science education. Waxmann

Webb NL (1997) Criteria for alignment of expectations and assessments in mathematics and science education. Council of Chief State School Officers and National Institute for Science Education Research Monograph, National Institution for Science Education

Webb's Depth of Knowledge guide (2009) Career and technical education definitions. https://www.aps.edu/sapr/documents/resources/Webbs_DOK_Guide.pdf