Asia Pacific Journal of Education

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/cape20

Refining the opportunity to learn construct through the Productive Disciplinary Engagement framework

Yann Shiou Ong, Lee Yew-Jin & Miechie Leowardy

ASIA PACIFIC JOURNAL OF EDUCATION
2024, VOL. 44, NO. 3, 582–596
https://doi.org/10.1080/02188791.2024.2390665

Refining the opportunity to learn construct through the Productive Disciplinary Engagement framework

Yann Shiou Ong, Lee Yew-Jin and Miechie Leowardy

Natural Science and Science Education, National Institute of Education, Nanyang Technological University, Singapore

ABSTRACT

Opportunity to learn (OTL) is a ubiquitous measure of the likelihood of learning in educational research, which typically has been characterized by three dimensions: time, coverage of content, and quality of instruction. The last dimension has been defined in highly divergent ways, which gives it a double-edged nature. While it may be operationalized to better serve a specific problem or context, it also makes it harder to achieve consensus or derive broader implications across studies. Using the four design principles from the Productive Disciplinary Engagement (PDE) framework, we show how PDE can characterize the quality of instruction dimension in OTL that reflects contemporary understandings of the end-goals of science education i.e. learning science as practice. This revision is valuable for science educators and evaluators who rely on OTL measures for it helps them: i) evaluate the quality of instruction consistent with reformed science teaching in valid and reliable ways, and ii) address questions about the adequacy of content coverage in the subject. We exemplify analysis using this refined OTL model with a case study of middle-school learners in Singapore engaged in the challenging scientific practice of argumentation.

ARTICLE HISTORY
Received 6 October 2023
Accepted 26 May 2024

KEYWORDS
Productive disciplinary engagement; opportunity to learn; quality of instruction; scientific practices; science education

Introduction

If there exists a simple indicator of the rising interest in evaluation research in science education, it may be found by checking the content pages across the three editions of the *Handbook of Research on Science Education*. The latter have been bellwether reference guides for science educators since 1993: The first edition published a general chapter on assessment in science education (Doran et al., 1993), while the second saw a dedicated chapter on programme evaluation in science education for the first time (Lawrenz & Thao, 2014). The third and current edition is privileged to contain two chapters devoted to evaluation research; one that reviewed project and programme evaluation focusing on the North American situation (Woodruff & Nian, 2023) and the other an updated review of evaluation studies in science education (Lawrenz & Goodyear, 2023). Although it can be argued that assessment is a broader term that encompasses and overlaps with evaluation studies (see Moed, 2017), we believe that the presence of standalone chapters in the last two editions is an acknowledgement that evaluation research is an important foundational activity that supports and advances science education.

Another impetus for evaluation research in science education has been the *Next Generation Science Standards* (NGSS) reforms in the United States of America (Woodruff & Nian, 2023).

CONTACT Lee Yew-Jin yewjin.lee@nie.edu.sg Natural Science and Science Education, National Institute of Education, Nanyang Technological University, Singapore 637616

A decade ago, the NGSS painted a vision of what and how to teach science that aligned with contemporary understandings of achieving scientific literacy in K-12 education (National Research Council [NRC], 2012, NGSS Lead States, 2013). Compared to previous eras that either emphasized the teaching of science through inquiry or the acquisition of content knowledge, the NGSS instead proposed a three-dimensional structure of instruction where disciplinary core ideas, cross-cutting concepts, and science and engineering practices (i.e., a suite of ways of thinking and doing science/engineering) were intertwined. This three-dimensional science learning reflected the "practice turn" in science education that aimed to engage students in learning science through actual participation in practices resembling those of professional scientists (Forman, 2018).

Elements of these ambitious reforms have spread to other regions including Singapore, but more significantly they have spurred renewed interest in evaluating teacher effectiveness. This is because implementing such demanding forms of instruction requires science teachers who are knowledgeable in the subject, able to facilitate students navigating between practices such as asking questions, planning and carrying out investigations, constructing explanations, coordinating evidence, and being comfortable when working with uncertainty similar to what scientists and engineers encounter daily in their work (Osborne, 2014). Indeed, teaching in the spirit of NGSS especially with the scientific practices are acknowledged to present numerous challenges. Hence, knowing the factors and conditions that can support such reformed teaching has assumed great importance. Others have also reported that the NGSS prioritizes measurable and reproducible performances, which understandably encouraged the rise of evaluation in/for science education too (Hoeg & Bencze, 2017). In any case, programme logic and policy wisdom normatively underscore how high-quality teaching is associated with, if not a precursor for student achievement in school (Woodruff & Nian, 2023). Identifying and being able to confidently determine what counts as instructional quality, therefore, is essential to the success of any intervention that seeks to improve (science) education through teaching.

In this paper, we propose a new means of evaluating science teaching quality in the opportunity to learn (OTL) construct that is consistent with international/local reform mandates that foreground teaching with scientific practices. By respecifying the quality of instruction dimension according to four design principles from the Productive Disciplinary Engagement (PDE) framework (Engle & Conant, 2002) that we explain later, it will satisfy two important concerns among science educators and evaluators who use OTL. First, this revision can characterize with validity and reliability what is meant by quality of instruction (& also content coverage) in OTL that hitherto has been highly dependent on specific research priorities and study contexts without denying the value of the latter when deemed necessary. Second, it affords better alignment of science instruction with current practice-based reforms for achieving scientific literacy in schools. The PDE design principles have recently been applied as a means of planning and measuring effective forms of integrated STEM activities (Ong et al., 2023), which we believe can similarly offer a practical contribution among science educators and evaluators who use OTL. Following the literature review in the next section, we describe the implementation of a middle-school science activity that we designed to show the development and analytic utility of our revised OTL construct.

Literature review

Evaluation research in science education

In the latest review of evaluation research in science education by Lawrenz and Goodyear (2023), it was reported that the creation of evaluation instruments/frameworks across grade levels, subdisciplines of science (e.g., physics, chemistry), general teaching approaches (e.g., inquiry, Nature of Science, reasoning, teacher self-efficacy), and psychological constructs (e.g., attitudes, perceptions) occupied the bulk of such studies and reflected the strongly functional nature of this work. The remainder of scholarly interest dwelt on insights gleaned from evaluations of programmes and curricula, improvement of evaluation/assessment methods and programmes, or comparisons between different types of evaluations, which have likewise served to further scientific literacy writ large. Such patterns had already been observed in an earlier review where Lawrenz and Thao (2014) highlighted the need for greater consideration of equity and diversity in evaluation research as well as capacity building for conducting rigorous, high-quality evaluations.

The use of OTL concepts although not widespread among science educators and evaluators has been diverse. Mo et al. (2013) were more aligned with OTL used in general education when they drew on topic coverage, teachers' quality, and school enrichment/remedial programmes as their criteria whereas Wang (1998) focused on content (i.e., aspects of coverage, exposure, emphasis) and instructional quality albeit without considering time allocation. Applying measures of OTL that specifically catered to their study, Kang et al. (2016) considered student learning experiences (e.g., what students do and talk about, production of artefacts, range of participation) as their indicators of learning opportunities for scientific thinking. Gerber et al. (2001) adopted a different approach by creating a survey involving subscales of social activities, school activities, nature of participation, and work/domestic chores among others to assess informal learning activities in science.

Opportunity to learn construct

OTL is a ubiquitous yet evolving, multi-faceted construct for determining the likelihood of learning as an outcome of instruction be it at the system, school, classroom or individual level (Kurz, 2011). Starting out as a research tool, it has been widely incorporated into policymaking because "[p]roviding students learning opportunities is what schooling is about" (Schmidt & Maier, 2009, p. 556). OTL is frequently associated with student learning although through a probabilistic rather than a deterministic manner that accepts the presence of known and unknown mediating factors. One early definition of OTL was "whether or not …students have had the opportunity to study a particular topic or learn how to solve a particular type of problem presented by the test" (Husen, 1967, pp. 162–163). Since then, this generic indicator of progress has been widely employed in international large-scale achievement tests such as the Programme for International Student Assessment (PISA) and Trends in International Mathematics and Science Study (TIMSS) albeit operationalized in different ways (see Luyten, 2016; Nilsen & Gustafsson, 2016). Others have adapted OTL to sample aspects of content coverage/exposure in the curriculum/teaching, instructional time, emphases of specific topics, teacher knowledge/qualifications, pedagogical strategies, depth/sequencing of intellectual tasks, learning resources, social supports, classroom climate, classroom management, venues of learning (e.g., out-of-school learning) and numerous other permutations of proximal as well as distal factors that may influence learning.

Some researchers have argued that classroom learning opportunities should rightfully involve interactions among three factors; namely, tasks, teaching, and students although without being able to pinpoint the precise nature of their measurement (see Cai et al., 2020). Indeed, since the 1960s, OTL measures were conceptualized in terms of three broad dimensions at the classroom level, namely the: i) amount of time engaged in learning, ii) coverage of relevant content, and iii) quality of instruction that has been flexibly defined thus far (Elliott & Bartlett, 2016; Kurz, 2011). While the first two OTL dimensions provoke fewer disagreements with respect to their inclusion or measurement, the quality of teaching dimension has been susceptible to differing interpretations by researchers, teachers as well as by students (Hansen & Strietholt, 2018). For example, Tate (2001) encouraged urban educators to consider science education as a civil right and thus extended the quality of teaching dimension in OTL to incorporate organizational and fiscal issues. While paying attention to instructional time, but ignoring content coverage, the quality of teaching was separated into its generic (i.e., how well opportunities were provided) and local aspects (i.e., cultural relevance of learning opportunities) by Jensen et al. (2016). Among mathematics educators where curriculum research has arguably been given more sustained attention than in science, there still appears to be no consensus as to what should comprise the quality of instruction (see Walkowiak et al., 2017).

Such variability in OTL especially in the quality of instruction dimension is double-edged. On the one hand, measures here can be customized for the contexts of the study. On the other hand, it makes it hard to compare findings across cases leading to problems of generalization as individual studies prioritize a mix of proximal and distal factors (Luyten, 2016; Nilsen & Gustafsson, 2016). Nonetheless, teachers' instructional practices, which constitutes the enacted curriculum in the classroom has garnered widespread appeal together with content coverage (Kurz & Elliott, 2011). As researchers continue to refine their characterization of OTL and understand what explanations it can offer (or not), "[i]ts greatest contribution has been to help unpack the proverbial 'black box' representing the schooling processes that translate resource inputs into student outcomes" (McDonnell, 1995, p. 317). The latter underscores the primacy of teacher-level effects in OTL (Stevens, 1996) that are recognized as the main though certainly not sole determinant of student achievement.

As science educators, we recently designed novel science activities to support middle-school teachers in Singapore, which had recently reformed its science curriculum. The rationales for adopting OTL measures during the evaluation of these activities were due to the short (both in duration & frequency) nature of the activities as well as the research design that eschewed an experimental format (i.e., controlling for important dependent variables such as teacher experience, stakeholder familiarity with ambitious pedagogies, classroom climate, student ability, type of school etc.). OTL was thus adopted to determine the affordances of these activities for achieving desired student outcomes in science. However, we also wanted to adopt instruments that could evaluate instruction that was aligned with the desired end-goals of science education as much as we wanted tools that could analyse the quality of instruction with clear, easy to understand criteria. For this task, we turned to ideas originating from another framework.

Productive Disciplinary Engagement (PDE)

The PDE framework from Engle and Conant (2002) and later elaborated by Engle (2012) is a framework for designing/evaluating learning environments aimed at fostering students' engagement in classroom activities as a community of learners in disciplinary ways that are productive. Originating from the learning sciences community, the framework comprises two components: three-dimensional student outcomes of PDE and a set of four design principles. The outcomes of PDE revolve around student engagement that Engle and Conant (2002) described as the extent to which students contribute to a discussion about a topic in coordination with other students and continue to be involved attentively over a period of time and/or re-engage in the discussion promptly. Such engagement is disciplinary in nature when "'there is some contact between what students are doing and the issues and practices of a discipline's discourse'" (ibid, p. 402). These authors were aware that there may be a diversity of conceptions on what counts as disciplinary engagement, and thus they offered flexibility in how "disciplinarity" could look like in one's own context. PDE also entails being productive as an essential characteristic of students' disciplinary engagement in a task. The productive nature of the disciplinary engagement considers the intellectual improvements/progress made in a disciplinary issue that students are engaged from the start to the end of their engagement with the issue. Progress could include recognizing a confusion or problem with an idea, making connections across ideas, or constructing a scientifically sounder argument. Thus, PDE outcomes situate learning as socially constructed and anchor the development of understanding on students' collaborative and interactive discourse as well as participatory actions.

To achieve these outcomes, a set of four design principles – problematizing, resources, accountability, and authority – should be met by the instructional designer. *Problematising* entails having students work on a problem that is meaningful to the disciplinary community, i.e., scientific communities in our case. Students could also further define the problem, pose further problems or generate new questions. To help students solve the problem, students need to be provided with *resources* that could include sufficient time (corresponding to the amount of time dimension of OTL), materials (including relevant information), and adequate scaffolding. Additionally, students should be given the *authority* to define, address, and solve the problem. Having authority also means the teacher and other learners position the students in more active roles such as stakeholders, contributors, and experts, which encourages them to be joint producers of knowledge. This may lead to the development of a sense of ownership in their work that motivates their engagement with the tasks. While students are encouraged to have the authority to produce knowledge, their ideas need to be held accountable to relevant standards too. *Accountability* refers to the expectation that students would be responsive towards other perspectives, especially those holding alternative yet valid perspectives. This design principle asks that students reconcile why their ideas differ from others' valid ideas, including authoritative disciplinary ideas such as consensual scientific knowledge. It also reflects the need to hold students' ideas accountable to epistemic criteria or ideals valued by scientific communities such as the notions of reliability and validity (Chinn et al., 2014). Both the teacher as well as the students in the learning community should have shared authority to hold ideas accountable. In summary, the PDE design principles highlight the need for balancing problematizing with resources and balancing authority with accountability (Engle, 2012) in instructional design in order to foster students' PDE outcomes. These four principles have also been utilized to evaluate how learning environments/instructions were designed to foster students' engagement in scientific practices, including scientific argumentation around an orca classification problem (Engle & Conant, 2002) and scientific modelling around modelling a backyard ecosystem (Manz, 2012).

Evaluating the quality of instruction in OTL

By interpreting disciplinary engagement as students' engagement in scientific practices – which includes scientific conceptual knowledge as well as epistemic and social aspects of science (Duschl, 2008) – the four PDE design principles were deemed appropriate for characterizing the quality of science instruction in OTL that also confirms with reformed understandings of what really matters in science teaching. In addition, unpacking what is problematized (through the problematizing design principle) could inform content coverage in OTL, which typically focuses on conceptual knowledge while neglecting the epistemic and social aspects of science education (Duschl, 2008). Through the unpacking of content coverage, the required time in learning in OTL (resource design principle), which includes time for students to engage with the relevant scientific practices as well as learn the needful conceptual, epistemic, and social aspects of science to solve the problem, becomes salient.

Table 1 represents the relationships between four PDE design principles and three OTL dimensions in the context of science education. Based on this framework, we now show how to analyse the quality of instruction dimension in OTL with a case study of a science activity among middle-school learners in Singapore.

Case study context

The case study data came from a larger project aimed at developing new science activities foregrounding scientific practices for middle-school learners in Singapore. All necessary consent/assent were obtained from teachers, students and parents to participate in the study following university IRB guidelines. Multiple data sources were collected; those of relevance for the case study reported here include video- and audio-recordings of lessons, student focus group discussions (FGD), and students' completed handouts. Semi-structured FGDs of 30–45 min were conducted within one week following the activity where students were asked "what have you learned from this activity? Is there any part(s) of the activity that is/are challenging?" to gain insights on how the students engaged in the activity and their views about their learning experiences. All the FGDs were transcribed and analysed.

The activity reported in the case study was implemented at two different secondary schools in Singapore. Jasper Secondary School (JSS) was an academically average secondary school for girls. At

Table 1. PDE-informed OTL constructs for scientific practices.

OTL dimensions	PDE design principles			
	Problematising	*Resources*	*Authority*	*Accountability*
Quality of instruction	Problem students engage with is also meaningful to scientific communities.	Resources are provided to help students solve the problem through engagement in the targeted scientific practice(s).	Students are given the authority to define, address, and solve the problem thus producing knowledge.	Students' ideas are held accountable by others to the epistemic criteria valued by scientific communities.
Content coverage	Conceptual, epistemic, and social aspects of science to solve the problem are included in instruction.			
Time in learning		Sufficient time as a resource for students to engage with relevant practices and learn needful conceptual, epistemic, and social aspects of science to solve the problem.		

JSS, the activity ran for three hours and was enacted by Mr Teo, a chemistry teacher with five years of teaching experience. The participating class J1 comprised 38 consented seventh graders in the Express stream i.e., students with stronger academic aptitudes. Twelve students participated in the FGD. Damian Secondary School (DSS) was a secondary school with strong emphasis in STEM education. At DSS, the activity ran for one hour and was enacted by Ms Saara, a chemistry and biology teacher with 13 years of teaching experience. The participating class D1 comprised 11 consented eighth graders also in the Express stream. Three students participated in the FGD.

The lactose intolerance activity

The lactose intolerance (LI) activity coupled relevant science content knowledge learning outcomes in the Singapore middle-school science theme of Systems (MOE, 2020) i.e., the human digestive system with an open-ended scientific issue that could engage students in the practice of scientific argumentation. The LI activity materials were adapted from curriculum resources from *Biointeractive*, a website which houses teaching and learning resources, and other existing resources (e.g., Gerbault et al., 2011; Hertzler & Savaiano, 1996; Ingram et al., 2009). The overarching problem or puzzle presented in the LI activity was: *"Why are there more people who are lactose tolerant in some places, but not in others?"* Solving the LI problem would require students to draw upon concepts of evolution and genetics to explain the non-random geographical distribution of lactose intolerant populations in the world. Not only is LI a common medical condition that might resonate with Asian students, it is also likely that they might hold some intuitive ideas about its causes.

The current scientific consensus involves a genetic explanation: the genes that are activated for lactase (lactase is the enzyme that digests lactose) remain active after weaning among some people; those who are not lactose intolerant. However, it is unclear why this happens only among some ancestral populations in the world, but not in others (Gerbault et al., 2011). This presents an interesting puzzle because scientists have yet to achieve full understanding regarding this phenomenon that draws on genetics, biological evolution, human immigration, and cultural practices of food preparation as new scientific and archaeological evidence appears over time.

The LI activity comprises three sections: content knowledge (30 minutes), practical work (30 minutes), and argumentation (60 minutes). The first section provides students with the opportunity to learn the relevant content, including: (i) What is LI condition; (ii) what causes LI (i.e., how the presence and action of enzyme lactase affect LI); and (iii) the symptoms of LI. Such content

Table 2. A list of three possible explanations/claims to the main question: *"why are there more people who are lactose tolerant in some places, but not in others?"*.

Explanation/ claim	Statement
A	People in some areas are just born lactose tolerant!
B	Some people regularly drank milk and eventually became lactose tolerant!
C	Some lactose intolerant people found ways to reduce lactose in food, so it didn't matter if they were lactose tolerant or not!

Table 3. List of evidence cards that could support the explanation/claim.

Evidence card	Description
1	A written statement of the history of the lactose tolerance gene
2	A map of the genetic heritage of some regions
3	A summary of a scientific research study on the effect of daily lactose feeding to the symptoms of lactose intolerance
4	A map of the distribution of global milk consumption
5	A table of percentage of lactose content in some dairy products
6	A map of the distribution of global milk production

knowledge is needed to support students' engagement with the remaining parts of the activity. The practical work involves the conversion of lactose in milk into glucose by the enzyme lactase. Students conduct a short investigation to determine which of two unlabelled glasses of milk contain milk with lactose (i.e., regular milk) using lactase pills and urine strips. To do so, students measure the amount of glucose using the urine strip before and after lactase pills are added into the milk. As lactase converts lactose into glucose, the glucose content in regular milk increases whereas that of lactose-free milk remains the same. The LI activity culminates in the argumentation section. Having understood LI-related conceptual knowledge from the previous two sections, students are shown a world map with the global distribution of lactose tolerant populations (i.e., non-LI people) and posed the anchor question: *"Why are there more people who are lactose tolerant in some places, but not in others?"*. Given three different possible explanations by the teacher, students work in groups to determine the best explanation that answers the question. These explanations are listed in Table 2.

Six evidence cards (ECs), each containing information that supports/refutes only one of the explanations, are provided to the students. Each explanation may be supported by more than one evidence card. A brief description of the six cards is listed in Table 3.

Student groups have to reach consensus on which explanation best answers the anchor question, select the relevant evidence card(s) to support that explanation, and justify their argument with reasons. The group discussion is followed by whole class discussion where groups share their argument and have the opportunity to engage in peer critique. Peer critique is cued by the handout questions: *"Do you agree with the other groups' views? Why or why not?"*.

Findings

We present the analysis of the enacted argumentation section in both classes using the PDE-informed OTL construct. The analysis is summarized in Table 4 and then elaborated.

Quality of instruction

Regarding the PDE design principles, the two classes faced issues with problematizing the goal of the argumentation activity, lack of resources to support students' critique of arguments as well as inadequacy in holding students' arguments accountable to disciplinary criteria despite giving students the authority to construct and critique arguments. We explain these points in turn.

Table 4. Analysis of enacted argumentation section using PDE-informed OTL construct. ✓ = design principle for OTL was met; × = design principle for OTL was not met.

OTL dimensions	PDE design principles			
	Problematizing	*Resources*	*Authority*	*Accountability*
Quality of instruction	LI activity anchor question was to determine the best scientific explanation with justifications for the global distribution of lactose tolerant populations (i.e., function of the argument to persuade others of an answer using evidence). × However, teachers focused on the form of an argument rather than its function.	Provided resources include: ✓ Information about LI (videos and self-research). ✓ Three possible explanations and six evidence cards. × Limited structural support for constructing arguments; inadequate support for critiquing arguments.	✓ Students had authority to construct their own argument. ✓ Students were invited to critique their peers' arguments publicly during class presentation or privately as written critique on handout.	× Students' arguments were not held accountable by their teacher or peers during whole class presentation. × Students' arguments were not held accountable to epistemic criteria except the need to be supported by at least one evidence card.
Content coverage	Conceptual ✓ Students learnt about lactose intolerance (LI) causes and symptoms. Epistemic × Students were not taught the epistemic criteria for constructing/critiquing scientific arguments, except for the reminder to include evidence to support their argument. Social × Students were not instructed on how to interpret the given multiple representations (e.g., distribution on world maps and tables in evidence cards) for communicating scientific ideas. × Students were not instructed on how to reach group consensus			
Time in learning		✓ Students had sufficient time to learn the content knowledge. × Students did not carry out the practical work. × One class had sufficient time to engage in the argumentation task but not the other.		

Problematising

The argumentation goal of the LI activity is for students to construct a scientifically sound argument for the distribution (and not solely the existence) of lactose tolerant populations i.e., the distribution problem. This goal was framed by the question, *"Why are there more people* *who are lactose tolerant in some places, but not in others?"* Both teachers showed the lactose tolerant population distribution map and described the distribution (to various extents) to highlight the uneven geographical distribution of LI population i.e., the phenomenon in question. However, the teachers' presentation of the task emphasized constructing the form of an argument rather than our intended function or goal of the argument task as discussed below.

After reading aloud the anchor question, Mr Teo instructed the students to "*have a[n] [online] post that says which one, A B, or C, you agree with and which [evidence card] you choose to support and give me [the] explanation [for] why you choose that"*. Thus, Mr Teo emphasised the task was to select the explanation (i.e., claim) that the students agreed with and support it with evidence and reasoning, i.e., the form of an argument, instead of determining the explanation that best answered the distribution problem i.e., the function of an argument. This task interpretation was reflected through students' descriptions of their activity actions during the FGD. For example, Henny's group chose explanation A (People in some areas are just born lactose tolerant) because they thought they *"could explain it the most with the evidence card 1, 2, and 4"*. While Henny's group selected their explanation based on the scientific criterion of selecting the explanation supported by the most amount of evidence, the group did not consider whether their selected explanation would best answer the distribution problem.

Ms Saara's instruction for the argumentation task started with, *"[f]or each EC, you and your group will select one or two or three or more that you . . . think can support one or more of the three claims that you saw earlier"*. She then added, *"you need to be very clear which one of the three claims is supported by the EC you selected or the other way around; you are looking for the EC that supports the particular claim"*. Finally, she said, *"[c]hoose the relevant EC to support your explanation. So, it depends on which one your group feels is the most relevant claim and you choose the EC"*. Ms Saara's instructions might have confused the students about whether they were supposed to first select EC(s) and then identify the claim the EC(s) supported or to select one claim then identify EC(s) that support the selected claim. Either interpretation might likely have led students to think the task goal was to match the ECs with the claims/explanations rather than evaluate each claim's potential to answer the anchor question. Indeed, during the FGD with Ms Saara's students, Sachi, mentioned that all claims *"would be possible if backed up with the right type of evidence"*, to which Aditya, another student of Ms Saara, also agreed. The students did not consider whether a claim was sound based on how well it answered the distribution problem. Thus, both teachers seemingly foregrounded the form rather than function of an argument, which severely diminished the intended problematization of the argumentation activity.

Resources

During instruction, students were supported in acquiring the necessary content knowledge, but had limited resources/skills to help them engage productively with argumentation. Regarding content knowledge, students were provided with five questions about LI in the student handout. These questions guided both teachers to successfully cover the necessary LI-related content knowledge for students to engage in the subsequent practical work and argumentation activity. Both teachers showed their students a video with relevant LI information, allowed students to do additional online research to answer the questions, as well as reviewed and elaborated on students' responses.

For argumentation resources, at the start of the activity, students were given three possible explanations and six ECs as previously described to support their argument construction. While students were expected to interpret the ECs on their own, Mr Teo provided extended commentary of the ECs. Mr Teo's descriptions helped his students craft their own argument. As evident in students' written responses, a few students developed Mr Teo's descriptions into better interpretations in their argument. Ms Saara, on the other hand, had her students make sense of the given claims and ECs independently. Consequently, most of her students' arguments merely re-stated information in the ECs without further interpretation or reasoning.

Despite providing the same explanations and EC as resources, how the teachers supported their students in making sense of these resources had affected how the students utilized the resources.

Throughout the activity, Mr Teo repeatedly instructed the students to "support your claim with reasoning and evidence" which indicated his emphasis solely on the structure of an argument rather than on the epistemic criteria of what makes a good scientific argument (e.g., validity, coherency, agreement with scientific concepts, support by evidences produced through reliable processes) (Sampson & Clark, 2008). Most students could provide claims, evidence, and reasoning, but more than half of the constructed arguments were not of high quality for these arguments mainly consisted of the ECs with unclear reasoning. For example, group 1 in Mr Teo's class wrote, *"We choose student B's explanation. This is evident in evidence card 6 because when more milk is produced around the world, more people will drink milk often. Thus there will be more lactase enzymes in their bodies so they will get used to the enzymes and thus, they become more tolerant to lactose"*. This argument did not take into account that the geographical locations of high/low milk production and milk consumption did not map well onto one another. Similar observations were made in Ms Saara's class where she instructed the students to "*choose the relevant EC to support your explanations*". Consequently, several groups did not provide clear reasonings in their arguments.

In addition to constructing arguments, students were expected to critique peers' arguments at the end of the argumentation activity to reach consensus on the best scientific explanation for the distribution problem. However, students had inadequate support to critique arguments. Mr Teo provided some guidance on how to critique an alternative argument, specifically the quality of its evidence, with the following prompts: *"Do you think there is a lack of evidence or better evidence for that claim? Do you think their evidence is good enough? Is it convincing to you? Do you think yours is better?"*. However, such prompts might be insufficient as we noted that many students merely provided a general commentary without critiquing peer groups' arguments. For example, Carissa wrote: "I agree with group 5 (claim C supported with EC5) as people who are lactose intolerant change their food diet due to a great variety of food sources". Carissa's reason for choosing group 5's argument simply reworded the reasoning provided by group 5 without evaluating their argument. While Ms Saara encouraged her students to engage in peer critique verbally, apart from asking if any student "objects" or "think it is not possible" after each group presentation, she neither elaborated on how to evaluate if the arguments were valid or accurate nor demonstrated how to critique the arguments.

Authority

In giving students authority, we focused on students taking on the dual role of constructors and critiquers of argument (Forman, 2018). As constructor, the students were given the opportunity to work in groups to construct an argument that best answered the main question of the argumentation activity by selecting a claim and evidence card(s) as well as their reasonings that link the evidence(s) to the claim. Students also had the opportunity to take on the role of critiquer during class presentation though this was enacted differently in the two classes. Mr Teo did not encourage his students to ask questions during the presentations. Instead, he asked them to answer the critique question in the handout, *"Do you agree with the other groups' views? Why or why not?"* He instructed his students to select a group's argument with a different claim than theirs and provide some critiques or comments. On the other hand, as previously described, Ms Saara prompted students to share if they had differing opinions towards the arguments presented by the three groups called on to present their arguments. Only one student shared that she agreed with the argument presented by one group because she had "*actually read an article about this exact thing*" i.e., claim B. This was followed by Ms Saara's affirmation. No other students demonstrated disagreement with the presenters, questioned the arguments presented or provided alternative arguments.

Accountability

As previously described, even when given the opportunity to critique, neither the students nor the teachers publicly highlighted problems with the groups' presented arguments. In other words, students' arguments were not held accountable by their teacher or peers. Furthermore, since both teachers focused on the form of an argument (i.e., claim, evidence, and reasoning) as previously discussed, they did not guide their students to think about how to critique an argument based on epistemic criteria for scientifically sound arguments valued by scientific experts. Only the requirement for groups' arguments to be supported by EC resembled the epistemic criterion that a claim should be supported by evidence.

Content coverage

Since the LI activity was designed by the authors and the teachers' instruction mostly followed the LI activity materials, our analysis of content coverage focused on what was included/excluded in the aforementioned. Based on the intended problematization of the argumentation activity, the instructional content of the LI activity should intertwine relevant conceptual, epistemic, and social aspects of science in order for students to construct the best scientific explanation for the distribution problem through engagement in scientific argumentation. However, our analysis suggests only the conceptual aspect of learning was covered.

Conceptual

Engagement with the argumentation problem requires the following conceptual knowledge about LI:

(1) LI is a digestive disorder caused by the inability to digest lactose due to a deficiency in the amount of the lactase enzyme, which is responsible for breaking down lactose into glucose and galactose, and
(2) Excessive consumption of lactose can cause stomach pain, bloating, diarrhoea, flatulence, and constipation to person with LI due to breaking down of lactose by naturally occurring bacteria (i.e., microflora) in colon into by-products such as methane, hydrogen gas, short-chain fatty acid and carbon dioxide.

Students learned the aforementioned conceptual knowledge in the first section of the LI activity, as intended in the design, through previously described instructional approaches. Additionally, the action of lactase (c.f. point a) was reinforced through the practical work. Thus, there was adequate conceptual knowledge coverage in the LI activity instruction.

Epistemic

While a main goal of the argumentation section was for students to construct the most scientifically sound argument for the distribution problem through argumentation practices, the epistemic aspect of scientific argumentation was not planned in the student handout. As previously mentioned, students were not provided with resources nor further instruction on how to critique arguments using epistemic criteria valued by scientific experts. Thus, while opportunities for construction and critique of arguments were provided in the argumentation section, students were not taught how to decide what counts as a scientifically sound argument and when an argument is sound/unsound. Although the provision of ECs as a resource prompted students and their teachers to require the use of evidence to support their claim, students were not taught how to determine whether the evidence is reliable, how to weigh arguments based on their supporting evidence and reasonings, etc. The lack of epistemic content from the LI activity design and subsequently its instruction might explain why the students did not meaningfully engage in argumentation to evaluate peer groups' arguments and reach consensus on the most scientifically sound explanation as intended by the authors.

Social

Two social aspects were important for students' productive engagement in scientific argumentation. Firstly, the distribution problem and the ECs included multiple representations, such as tables (EC5) and distributions of entities on world maps (EC2, EC4, EC6), which the students might have been unfamiliar with. Students' encounter with world maps might be limited to the humanities subjects (e.g., history and geography), and thus, find their use in science lessons unfamiliar. Hence, it was necessary for the teachers to demonstrate how to interpret these multiple representations to derive the scientific ideas they intended to communicate. Unfortunately, such directions for teachers were missing in the LI activity instructional package, although Mr Teo did provide descriptions for each EC (c.f. Resources). However, it is uncertain to what extent Mr Teo's students understood the ideas communicated by the representations.

Secondly, while students were tasked to work in groups to construct a group argument, no guidance was provided to teach students how to reach consensus. During the FGD, students reported they "compromised" when they held different opinions and went with the claim that "was easier to do" (Natalie, Mr Teo's class), or ensured "everyone's work is inside [the presentation]", thus avoiding any disagreement or argument (Sarah, Mr Teo's class). In other groups, uncertainties might be quickly resolved when one student dominated the discussion while others were less involved in the decision-making (e.g., one group in Ms Saara's class). We consider the abovementioned approaches for reaching group consensus to be undesirable as the groups did not deliberate the different opinions/arguments based on epistemic criteria for a scientifically sound argument. In summary, the exclusion of both social aspects necessary for engagement with the argumentation problem might support why the students did not engage in the argumentation practice as intended.

Time in learning

Both teachers spent different amounts of curriculum time on the three-section LI activity. While the authors intended the activity to be completed in two hours, Mr Teo took about three hours while Ms Saara took only one hour. With the additional time, Mr Teo demonstrated the practical work for his students and had them complete questions on the handout that were relevant to the demonstration (25 minutes). His students also had time to conduct their own research on lactose intolerance (10 minutes) and a longer duration to construct their group argument (45 minutes). On the other hand, due to time constraint, Ms Saara omitted the practical work and focused on the argumentation section. Even so, her students had only 10 minutes to work on the argumentation task, which they noted during the FGD was insufficient time to delve deeply into the claims and ECs. Both teachers allocated time for class presentation (all groups presented in Mr Teo's class; selected groups presented in Ms Saara's class). Hence, while the students had sufficient time to learn LI-related content knowledge, they did not carry out the practical work. Additionally, while Mr Teo's class had sufficient time to engage in the group argumentation task, Ms Saara's class did not.

In summary, analysis of the argumentation section of the LI activity using the PDE-informed OTL construct revealed that the ***quality of instruction*** did not adequately support students' engagement in scientific argumentation. Only the design principle of authority was fulfilled; resources were partially provided whereas problematizing and accountability aspects were lacking. Unpacking the researchers' intended problematizing identified the necessary ***content coverage***. However, only conceptual content coverage was adequately designed for and implemented while the epistemic and social aspects were missing. Analysis of ***time in learning*** further reflected the emphasis on conceptual knowledge learning while practical work for students was eliminated and time for argumentation was inadequate in one class. While there seemed to be adequate time for argumentation in the other class, this might have been due to the teacher's simplified problematization of the task which prioritized the form of an argument over its function. The abovementioned inadequacies revealed through our analysis of the case study using the PDE-informed OTL construct suggest a number of future improvements in the LI activity design and implementation, which we argue demonstrates the utility of these changes in the construct.

Conclusion

In this paper, we have presented a new PDE-informed OTL construct for science educators and evaluators who use the former in their research (Table 1), which is our attempt at the persistent challenge of formulating a "theoretical and empirical rationale. . .[for the] operationalization of instructional quality" (Kurz, 2011, p. 113). Additionally, through the problematizing aspect of the quality of instruction, the relevant conceptual, epistemic, and social aspects of content coverage in science could be ascertained and thus evaluated. Through analysis of middle school students engaging in a LI activity for learning scientific argumentation, we illustrated how this OTL revision helped characterize issues with the problematizing, resources, and accountability aspects of the activity. Together, such information may lead to more valid and reliable judgements concerning the quality of science instruction in the OTL construct although these may still be subject to interpretation. Nonetheless, we believe that our framework offers a holistic set of ideas that most researchers can arrive at a broad agreement about what "counts". This OTL revision has also revealed inadequacies in the content coverage too, specifically among the epistemic and social aspects of scientific practices related to our LI activity in school. We do not deny that there are indeed many worthwhile activities and programmes to improve science education, but this refined OTL construct has significant merits for evaluating science instruction especially when it can be closely aligned with recent notions of reformed teaching in the discipline (Lawrenz & Goodyear, 2023). While we have not demonstrated its use under different instructional contexts such as instruction targeted at different scientific practices (e.g., designing investigations; modelling) or other grade levels, etc., we remain confident of its broader utility in science instruction and research. We therefore invite other interested science educators using OTL to utilize our proposed framework and empirically test its applicability in their own classrooms.

Abbreviations

DSS	Damian Secondary School
EC	Evidence cards
FGD	Focus group discussion
JSS	Jasper Secondary School
LI	Lactose intolerance
NGSS	Next Generation Science Standards
OTL	Opportunity to learn
PDE	Productive Disciplinary Engagement
STEM	Science, technology, engineering, and mathematics

Acknowledgements

This study was funded by Singapore Ministry of Education (MOE) under the Education Research Funding Programme (DEV 02/19 LYJ) and administered by National Institute of Education (NIE), Nanyang Technological University, Singapore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore MOE and NIE. We gratefully acknowledge the discussions on earlier versions of this paper with Dr Timothy Tan and Ms Shermayne Soh.

Disclosure statement

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

Funding

The work was supported by the Ministry of Education – Singapore [DEV 02/19 LYJ].

Notes on contributors

Ong Yann Shiou is an assistant professor at the National Institute of Education, Nanyang Technological University, Singapore.

Lee Yew-Jin is an associate professor at the National Institute of Education, Nanyang Technological University, Singapore.

Miechie Leowardy is a research assistant at the National Institute of Education, Nanyang Technological University, Singapore.

ORCID

Yann Shiou Ong http://orcid.org/0000-0002-6092-2803
Lee Yew-Jin http://orcid.org/0000-0002-3286-2778
Miechie Leowardy http://orcid.org/0009-0006-6809-0606

References

Cai, J., Morris, A., Hohensee, C., Hwang, S., Robison, V., Cirillo, M., Kramer, S.L., Hiebert, J., & Bakker, A. (2020). Maximizing the quality of learning opportunities for every student. *Journal for Research in Mathematics Education*, *51*(1), 12–25. https://doi.org/10.5951/jresematheduc.2019.0005

Chinn, C.A., Rinehart, R.W., & Buckland, L.A. (2014). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D.N. Rapp & J.L.G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 425–453). MIT Press.

Doran, R.L., Lawrenz, B., & Helgeson, S. (1993). Research on assessment in science. In D. Gabel (Eds.), *Handbook of research in science teaching and learning* (pp. 388–432). Macmillan.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education*, *32*(1), 268–291. https://doi.org/10.3102/0091732X07309371

Elliott, S.N., & Bartlett, B.J. (2016). Opportunity to learn In *Oxford handbook topics in psychology* (online) edn, Oxford Academic. 3 Mar. 2014. July 10, 2023. https://doi.org/10.1093/oxfordhb/9780199935291.013.70

Engle, R.A. (2012). The productive disciplinary engagement framework: Origins, key concepts, and developments. In D. Y. Dai (Eds.), *Design research on learning and thinking in educational settings* (pp. 161–200). Taylor & Francis.

Engle, R.A., & Conant, F.R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition & Instruction*, *20*(4), 399–483. https://doi.org/10.1207/S1532690XCI2004_1

Forman, E.A. (2018). The practice turn in learning theory and science education. In D.W. Kritt (Ed.), *Constructivist education in an age of accountability* (pp. 97–111). Springer International Publishing. https://doi.org/10.1007/978-3-319-66050-9_5

Gerbault, P., Liebert, A., Itan, Y., Powell, A., Currat, M., Burger, J., Swallow, D.M., & Thomas, M.G. (2011). Evolution of lactase persistence: An example of human niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1566), 863–877. https://doi.org/10.1098/rstb.2010.0268

Gerber, B.L., Marek, E.A., & Cavallo, A.M. (2001). Development of an informal learning opportunities assay. *International Journal of Science Education*, *23*(6), 569–583. https://doi.org/10.1080/09500690116959

Hansen, K.Y., & Strietholt, R. (2018). Does schooling actually perpetuate educational inequality in mathematics performance? A validity question on the measures of opportunity to learn in PISA. *ZDM Mathematics Educations*, *50*(4), 643–658. https://doi.org/10.1007/s11858-018-0935-3

Hertzler, S.R., & Savaiano, D.A. (1996). Colonic adaptation to daily lactose feeding in lactose maldigesters reduces lactose intolerance. *The American Journal of Clinical Nutrition*, *64*(2), 232–236. https://doi.org/10.1093/ajcn/64.2.232

Hoeg, D.G., & Bencze, J.L. (2017). Values underpinning STEM education in the USA: An analysis of the next generation science standards. *Science Education*, *101*(2), 278–301. https://doi.org/10.1002/sce.21260

Husen, T. (Ed.). (1967). *International study of achievement in mathematics: A comparison of twelve countries* (Vol. I). John Wiley & Sons.

Ingram, C.J., Mulcare, C.A., Itan, Y., Thomas, M.G., & Swallow, D.M. (2009). Lactose digestion and the evolutionary genetics of lactase persistence. *Human Genetics*, *124*(6), 579–591. https://doi.org/10.1007/s00439-008-0593-6

Jensen, B., Pérez Martínez, M.G., & Escobar, A. (2016). Framing and assessing classroom opportunity to learn: The case of Mexico. *Assessment in Education Principles, Policy & Practice*, *23*(1), 149–172. https://doi.org/10.1080/0969594X.2015.1111192

Kang, H., Windschitl, M., Stroupe, D., & Thompson, J. (2016). Designing, launching, and implementing high quality learning opportunities for students that advance scientific thinking. *Journal of Research in Science Teaching*, *53*(9), 1316–1340. https://doi.org/10.1002/tea.21329

Kurz, A. (2011). Access to what should be taught and will be tested: Students' opportunity to learn the intended curriculum. In S.N. Elliott, R.J. Kettler, P.A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students: Bridging the gaps between research, practice, and policy* (pp. 99–130). Springer.

Kurz, A., & Elliott, S.N. (2011). Overcoming barriers to access for students with disabilities: Testing accommodations and beyond. In M. Russell (Eds.), *Assessing students in the margins: Challenges, strategies, and techniques* (pp. 31–58). Information Age Publishing.

Lawrenz, F., & Goodyear, L. (2023). Review of research about science education program evaluation. In N.G. Lederman, D. L. Zeidler, & J.S. Lederman (Eds.), *Handbook of research on science education: Volume III* (pp. 992–1009). Routledge.

Lawrenz, F., & Thao, M. (2014). Review of science education program evaluation. In N.G. Lederman & S.K. Abell (Eds.), *Handbook of research on science education: Volume II* (pp. 773–788). Routledge.

Luyten, H. (2016). Predictive power of OTL measures in TIMSS and PISA. In J. Scheerens (Eds.), *Opportunity to learn, curriculum alignment and test preparation* (pp. 103–119). Springer.

Manz, E. (2012). Understanding the codevelopment of modeling practice and ecological knowledge. *Science Education*, *96*(6), 1071–1105. https://doi.org/10.1002/sce.21030

McDonnell, L.M. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, *17*(3), 305–322. https://doi.org/10.3102/01623737017003305

Ministry of Education [MOE]. (2020). *Science syllabuses: Lower secondary express course, normal (academic) course*. Ministry of Education.

Mo, Y., Singh, K., & Chang, M. (2013). Opportunity to learn and student engagement: A HLM study on eighth grade science achievement. *Educational Research for Policy and Practice*, *12*(1), 3–19. https://doi.org/10.1007/s10671-011-9126-5

Moed, H.F. (2017). *Applied evaluative informetrics*. Springer.

National Research Council [NRC]. (2012). *A framework for K-12 science education: Practices, crosscutting concepts and core ideas*. National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Nilsen, T., & Gustafsson, J.-E. (Eds.). (2016). *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time*. IEA Research Institute of Education.

Ong, Y.S., Koh, J., Tan, A.L., & Ng, Y.S. (2023). Developing an integrated STEM classroom observation protocol using the productive disciplinary engagement framework. *Research in Science Education*, *54*(1), 101–118. https://doi.org/10.1007/s11165-023-10110-z

Osborne, J. (2014). Teaching scientific practices: Meeting the challenge of change. *Journal of Science Teacher Education*, *25*(2), 177–196. https://doi.org/10.1007/s10972-014-9384-1

Sampson, V., & Clark, D.B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*(3), 447–472. https://doi.org/10.1002/sce.20276

Schmidt, W.H., & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. Schneider, & D.N. Plank (Eds.), *Handbook of education policy research* (pp. 541–559). Routledge.

Stevens, F.I. (1996). *The need to expand the opportunity to learn conceptual framework: Should students, parents, and school resources be included?* https://files.eric.ed.gov/fulltext/ED397523.pdf

Tate, W. (2001). Science education as a civil right: Urban schools and opportunity-to-learn considerations. *Journal of Research in Science Teaching*, *38*(9), 1015–1028. https://doi.org/10.1002/tea.1045

Walkowiak, T.A., Pinter, H.H., & Better, R.Q. (2017). A reconceptualized framework for 'opportunity to learn' in school mathematics. *Journal of Mathematics Education*, *8*(1), 7–18. https://doi.org/10.7916/jmetc.v8i1.800

Wang, J. (1998). Opportunity to learn: The impacts and policy implications. *Educational Evaluation and Policy Analysis*, *20*(3), 137–156. https://doi.org/10.3102/01623737020003137

Woodruff, S.B., & Nian, Q. (2023). Project evaluation: Its history, importance, and current practice. In N.G. Lederman, D. L. Zeidler, & J.S. Lederman (Eds.), *Handbook of research on science education: Volume III* (pp. 930–959). Routledge.