



Productive Disciplinary Engagement in High- and Low-Outcome Student Groups: Observations From Three Collaborative Science Learning Contexts

Milo D. Koretsky¹ · Marja Vauras²  · Cheryl Jones³ · Tuike Iiskala² · Simone Volet³

Published online: 04 April 2019
© The Author(s) 2019

Abstract

This study explored how productive disciplinary engagement (PDE) is associated with the level of cognitive activity and collective group outcome in collaborative learning across multiple contexts. Traditionally, PDE has been studied in a single collaborative learning environment, without analysis of how these environments fulfill the supporting conditions for PDE. In addition, research on the quality of a collective learning outcome and product in relation to the extent of the group's PDE during actual collaborative learning processes is scarce. In this study, the learning processes of low- and high-outcome small groups were compared within three collaborative learning contexts: high school general science, second year university veterinary science, and fourth year university engineering. Two meaningful and self-contained phases from each context were selected for analysis. The same theory-based analytical methods were used across contexts. The findings revealed similar patterns in the high school science and second year university veterinary science data sets, where high-outcome groups displayed a greater proportion of high-level cognitive activity while working on the task. Thus, they could be distinctively perceived as high- and low-performing groups. These high-performing groups' interactions also reflected more of the supporting conditions associated with PDE than the low-performing groups. An opposite pattern was found in the fourth year university engineering data set, calling for interpretation grounded in the literature on the nature and development of expertise. This study reveals the criticality of using

✉ Marja Vauras
marja.vauras@utu.fi

Milo D. Koretsky
milo.koretsky@oregonstate.edu

Cheryl Jones
C.A.Jones@murdoch.edu.au

Tuike Iiskala
tuike.iiskala@utu.fi

Simone Volet
S.Volet@murdoch.edu.au

Extended author information available on the last page of the article

comparable analytical methods across different contexts to enable discrepancies to emerge, thus refining our contextualized understanding of PDE in collaborative science learning.

Keywords Productive disciplinary engagement · Cognitive activity · Learning outcome · Science learning · Senior high school students · University students

Introduction

Collaborative environments have been widely recommended to support students' science learning. There is evidence that working collaboratively in small groups is positively associated with students' achievement in science in high school (e.g., Bowen 2000; Kirschner et al. 2011) and at the university level (e.g., Springer et al. 1999; Stump et al. 2011). However, simply offering students the opportunity to work in a group setting is not sufficient to support high-level science learning (e.g., Khosa and Volet 2014; Sampson and Clark 2009; Volet et al. 2013).

Several necessary conditions have been identified for collaborative environments to increase learning in science. They range from the context and quality of the learning task to the groups' engagement in the interpersonal coordination of their cognitive processes. In regard to learning tasks and contexts, many researchers have described the advantages of authentic, complex, and challenging science learning activities that engage student groups in disciplinary practices (Brown et al. 1989; Ford and Forman 2006). When embedded in school science projects the likelihood of students transferring the skills learned at school to professional practice increases (Bransford et al. 2000; Johri and Olds 2011; Stevens et al. 2008). Such projects are most effective when they are "group-worthy," where the task is complex enough that each student needs to contribute his or her ideas but also to draw from the resources and perspectives of other group members (Horn 2005; Lotan 2003).

The present study capitalizes on recent conceptualizations of productive disciplinary engagement (PDE; Engle 2012; Engle and Conant 2002). It extends to what is currently known about productive engagement in disciplinary thinking and practices in collaborative science learning beyond isolated settings to establish how PDE manifests across disciplinary learning contexts at different levels of education. We use the same methodological approach to investigate the nature of productive collaborative learning in three group-worthy projects completed in three different disciplinary learning contexts.

Productive Disciplinary Engagement in Science Learning

During the past few decades, many instructional improvement efforts in science, technology, engineering, and mathematics (STEM) have attempted to implement classroom practices that resemble disciplinary practices, such as modeling or argumentation (National Council of Teachers of Mathematics 2014; National Research Council 2012). This movement has occurred in North America, South America, Europe, and elsewhere in the world (Engle 2012; Forman et al. 2014). Productive engagement in science learning has its own research tradition. A general definition of engagement is "active, goal-directed, flexible, constructive, persistent, focused interactions with social and physical environments" (Furrer and Skinner 2003, p. 149).

In the present study, we applied Engle and Conant's (2002; Engle 2012) framework of productive disciplinary engagement (PDE) because it aims to capture the kinds of interaction with people and objects that are likely to result in deep learning of science concepts and

practices. PDE was developed to serve as a common framework to help learning scientists make comparisons across case studies of innovative instructional projects (Forman et al. 2014). For Engle and Conant (2002; see also Engle 2012), engagement is productive to the extent that conceptual or practical progress on a problem is made over time. Disciplinary engagement involves “real work” coordinated between individuals and groups, where “action [is] informed by meaning drawn from a particular group context” (Cook and Brown 1999, p. 387) and to the degree that participants use the concepts and practices of the discipline to achieve their aims (Engle 2012; Ford and Forman 2006). In short, PDE is assumed to occur when learners use the language, concepts, and practices of the discipline in authentic tasks to “get somewhere” (e.g., develop a product or improve a process) over time.

PDE in Designing Science Learning Environments and Understanding the Quality of Learning

The PDE framework was developed around the turn of the century in response to a general challenge posed to the community of learning scientists immersed in development and research on an array of innovative learning environments (see Engle 2012). It intends to provide general guiding principles for both designing and understanding the quality of learning environments. Fundamentally, PDE represents a state of learner engagement (Engle and Conant 2002). PDE is characterized by student interactions around issues core to a discipline (e.g., chemistry, biology, veterinary science, and chemical engineering) and manifests by richer use of the practices and discourse of the discipline than in traditional learning environments. In addition, student interactions that result in meaningful disciplinary progress can be seen as expressions of PDE. When learners’ interactions are characterized by PDE, learners are more likely to demonstrate a deep understanding of concepts and the incorporation of disciplinary practices (e.g., Comelius and Herrenkohl 2004; Scott et al. 2006; Venturini and Amade-Escot 2014).

In order to support instructional design, Engle (2012) described four supportive conditions for PDE as a system of tensions along two poles: authority-accountability and problematizing-resources, defined as follows:

- *Authority*: Students have authority to address problems. Specifically, (i) students have agency in identifying, formulating, and solving problems; and (ii) instructors publicly position students as stakeholders and producers of knowledge.
- *Accountability*: Students’ work is made accountable to others and to disciplinary norms. Ford (2008) elaborated this idea further by claiming that, in addition to accountability to the norms of the discipline, students (and professionals) in science are also accountable to nature.
- *Problematizing*: Students are encouraged to identify and take on disciplinary problems. Engle (2012) identified three aspects of problematizing: (i) it engenders genuine uncertainty; (ii) it is in some way responsive to the learner’s commitments; and (iii) it embodies “big ideas” or some other central aspect of the discipline.
- *Resources*: Students are provided with sufficient resources to do this work, including both elements such as sufficient background information and time, and resources for promoting authority and accountability—for example, through class presentations or meetings with disciplinary experts.

Engle (2012) examined 23 innovative learning environments and showed that instructional environments that substantially embody the four elements (two poles) foster PDE, while those

that have missing elements or even disproportionate emphasis on one of the poles (e.g., too much authority and not enough accountability) do not. Engle and colleagues' conceptualization of PDE has been connected to several innovative science instructional environments, and findings based on PDE have provided a foundation for design-based research projects throughout the learning sciences and science education communities (Forman et al. 2014). While empirical investigations of PDE have been carried out primarily in activities constructed to develop students' conceptual knowledge directly through argumentation, there is an emerging research on the use of design tasks to promote PDE (Apedoe and Ford 2010; Koretsky et al. 2014a; Koretsky et al. 2015). PDE clearly combines "doing" and "thinking" in practice.

To date, however, there is still a paucity of empirical studies grounded in Engle and Conant's conceptualization of PDE (Engle 2012; Engle and Conant 2002) in STEM collaborative learning contexts, and these studies present limitations. Most studies have investigated single collaborative learning contexts (see, e.g., Koretsky et al. 2014b; Sinatra et al. 2015) and provided only brief illustrations of the features of PDE (Forman et al. 2014). It is only recently that some researchers have analyzed collaborative groups' interaction data regarding a subset of the supporting conditions for PDE poles: for example, authority and accountability in mathematics learning (Boaler and Sengupta-Irving 2016) or problematizing in physics (Phillips et al. 2017). Furthermore, it is plausible to claim that prior studies have not made explicit how learning environments display the supporting conditions for PDE and, in turn, how this is reflected in students' activities. In other words, the crucial link between the supporting conditions that support design of learning environments and subsequent student groups' PDE in those environments has remained unexplained. To refine our understanding of the PDE framework and, in particular, how PDE manifests during collaborative learning, empirical research needs to be conducted across contexts with the same method of data analysis, and it should explicitly take into account the supporting features of learning environments. The study reported in this article aimed to address this issue through the use of a theoretically driven research design, data from three distinct learning contexts that varied in discipline and educational level in which the studies were implemented, and the adoption of a common analytical approach for all datasets.

Main Aims

Through the lens of PDE (Engle 2012), this study characterizes the ways that students working in small groups in different collaborative science learning activities co-construct their conceptual understanding as they make (produce) something together. For this purpose, this study encompassed three different learning contexts (varying in discipline and educational level, as described in the "Method" section) that were designed intentionally to foster PDE, thus substantially embodying Engle's (2012) conceptualization of PDE in collaborative learning but differing in the nature of the activities. Using three distinct and varied contexts instead of only one was expected to strengthen the findings of this study. In each context, we selected a high- and a low-outcome group and assessed their level of cognitive activity in terms of students' engagement with the science content as they interacted. Consistent with prior research (Khosa and Volet 2014), cognitive activity was operationalized as any talk related to the task and science content processing. It could range from gathering and collating information (low level) to knowledge co-construction and scientific meaning making (high level). Examining systematically the extent to which groups engaged in high-level cognitive

activity was important, since the conditions for PDE, such as exercising authority, displaying accountability to peers and the discipline, problematizing, and using sufficient resources as a group, all imply deep-level knowledge processing. From the discourse, we identified evidence and features of PDE as it unfolded in real time during their collaborative work.

The overall objective of the study was to explore PDE in a systematic way across distinct and varied educational contexts and levels. First, we sought to characterize the extent to which the quality of the small groups' collective learning outcome, in collaborative science learning environments designed to support PDE, is related to the quality of their cognitive activity during their work on the task. We then compared the manifestations of PDE in high- and low-outcome groups, focusing in particular on how these groups took up affordances for PDE, in relation to the four supporting conditions of Engle and Conant (2002; Engle 2012).

Two research questions guided this study, as follows:

- ResQ1. How do the high- and low-outcome student groups in the three learning contexts differ in the level of their cognitive activity during a collaborative learning task? Based on prior research (Khosha and Volet 2014; Volet et al. 2013), it was hypothesized that, within each learning context, the group that produced a high-level outcome would display a greater proportion of high-level cognitive activity than the group that produced a low-level outcome (H1).
- ResQ2. To what extent do the interactions of high- and low-outcome student groups in the three learning contexts reflect the supporting conditions for PDE (authority, accountability, problematizing, and resources) afforded to them by their learning context? In light of the scarce empirical evidence in the PDE literature to date, only the following hypothesis could be generated. It was hypothesized that the student interactions in the higher outcome groups would more frequently reflect Engle's (2012) supporting conditions for PDE than their lower outcome counterpart (H2).

Method

Participants

The study participants were high school students from Finland and university students from Australia and the USA. Table 1 provides information on the background of each sample and each of the collaborative tasks. The students' age range varied from 16 to 18 years old (senior high school general science) to predominantly 19–25-year-old students (second year university veterinary science and final year university engineering). In high school, groups were partly self-selected but the teachers tried to ensure that they were heterogeneous in terms of content knowledge, and that at least one student was skilled enough in English (the virtual environment was in English) in order to make it possible for the group to complete the demanding task. At the university level, students self-selected into groups, either on the basis of personal preference (veterinary students) or individual choice of virtual laboratory (engineering students). Two groups from the larger pool who completed the project were selected for analysis at each research site—one group with a low-quality outcome and one with a high-quality outcome.

All participation in the research was voluntary, and written permission for video or audio recording the groups' interactions was sought from all students (and custodians

Table 1 Data of the study

Country	Level of education	Discipline	Task phases by same groups	Selection of extreme groups based on independent assessment of outcome quality of the product	Task and context particulars
Type of data	Group composition	Material/tools			
Finland	High school (16–18 y.o.)	Phase 1: generate hypothesis	Quality of PPT presentation	Phase 1: groups had to generate a hypothesis. Phase 2: groups had to analyze results based on an experiment that they conducted, and prepare a presentation based on their study. Context: groups had to carry out a scientific study within a collaborative virtual science learning environment—Virtual Baltic Sea Explorer (ViBSE). They needed to generate hypotheses, create a study plan, conduct the study, analyze results, and give a presentation.	
Video and transcript	Chemistry and biology Two groups of 3 female students	Phase 2: analyze results/prepare a presentation Virtual laboratory	2 science teachers Scoring: inductive, content-based scale 0–5 Gp A: score of 3/Gp B: score of 5 (extreme scores out of 29 groups)		
Australia	University 2nd year undergraduates (mostly 19–25 y.o.)	Phase 1: generate learning objectives	Quality of concept map Expert teachers	Phase 1: groups had to generate learning objectives linked to their unique (teacher-allocated) clinical case study, Phase 2: groups had to create a case map with associative and causal linkages using supplied “concept” (symptoms, diagnostics, treatment) cards related to a clinical case. Context: cases treated at university veterinary hospital allocated to groups to reconstruct. Each group’s learning objectives underpinned a presentation of their case study.	
Video and transcript	Veterinary science One group of 5 female and 1 male and one of 4 females and 1 male	Phase 2: construct a concept map of a clinical case Glass board: set of concept/case cards	Scoring: % age agreement between group maps and expert maps (selected concepts and links between them, associative or causal) Gp A: score 56%/Gp B: score 92% (extreme scores out of 11 groups)		
USA	University 4th year undergraduates (mostly 21–25 y.o.)	Phase 1: exploring resources	Quality of memorandum and meeting interactions with instructor	Phase 1: groups had to develop initial experimental plans. Data were selected once students completed initial orientation and project management activities (setting meeting times; examining the assignment). Phase 2: groups had to finalize their experimental plan and articulate it in a design strategy memorandum. Context: groups placed in the role of real-world engineers tasked with optimizing an industrial process.	
Audio and transcript	Chemical, biological, or environmental engineering Two groups of 2 males and 1 female	Phase 2: constructing the memorandum Virtual laboratory	The project supervisor Scoring: content-based, scale 0–10 Gp A score: 8/Gp B score: 10 (largest difference in grade within cohort of available data out of 22 groups)		

if students were under 18 years of age). The participants were fully informed as to why they were participating in the research, what data would be used for research purposes, and were permitted to withdraw their participation during any phase of the research. Data were anonymized and original materials were stored according to the regulations of each country. In all phases of the research, ethical principles were strictly followed in accordance with the national guidelines of each country (Finland, Australia, and the USA).

Data on Interaction and Collective Outcome

In all learning contexts, small groups of students were recorded as they worked on a collaborative science learning task over an extended period. In all tasks, the groups had to deliver a tangible collective product. Data for this study included transcripts of the groups' interactions during two distinct phases of the overall task and the outcome of each group's collective product (see Table 1). For the high school general science groups, the outcome was the group's joint, final presentation to the class. For the university veterinary science students, the outcome was the jointly constructed map of their clinical case, and for the university engineering students, it was the memorandum jointly written, which formed the basis of a meeting with the instructor to discuss their experimental design for the project. The quality of outcome was assessed separately at each research site by experts in the discipline and independently from the process data (see Table 1).

Table 2 Summary of the PDE design principles and supporting task features in the three contexts

	High school general science	Second year university veterinary science	Fourth year university engineering
Authority	Authority to conduct field-based scientific inquiry.	Authority to formulate and argue on cause and effect relations in regard to a clinical case.	Authority to undertake an iterative design project with many possible solution paths.
Accountability	Held accountable to other students in their group, and to the entire class and the teacher through public presentation.	Held accountable to other students in their group, and to the entire class, instructors and clinicians through public presentation and discussion.	Held accountable to others in their group, and to the entire class and the instructors/supervisors through public presentation. Accountable to nature through the simulated reactor data.
Problematizing	Embedded and needed to progress in the experimental research task.	Generating and researching learning objectives that pertained to key aspects of the clinical case; reasoning and solving a genuine clinical case.	Embedded and needed to address a design task with competing constraints and limited financial resources.
Resources	Authentic data of marine biologists; various knowledge sources embedded in the software; virtual environment to perform scientific inquiry.	Authentic case material from the clinicians who attended that case in the hospital; scientific publications relating to the clinical condition.	Prior coursework, introductory lectures, internet resources, computation resources, and virtual environment to perform engineering design.

Collaborative Science Learning Environments

All three learning environments placed students in groups that positioned them in disciplinary roles. The learning environments were specifically designed to offer opportunities for students to engage in the four supporting conditions of PDE postulated by Engle (2012), namely *authority*, *accountability*, *problematizing*, and *resources*. The specific features of each context that were designed to support PDE are summarized in Table 2 and presented next.

High School General Science The Virtual Baltic Sea Explorer (ViBSE) web-based software was designed to provide high school students with adequate and inspiring tools and resources to build new integrated knowledge in two disciplines, biology and chemistry. The virtual learning environment provided a bridge between the school and science worlds by positioning students as researchers and fostering their adoption of the practices, objectives, and methods that guide the authentic research of professional scientists. In being challenged with a complex research task, the student groups had authority to formulate research questions, hypotheses, and research design; carry out their experiments by using authentic data from marine biologists; and draw conclusions based on outcomes. Problematizing was a natural and necessary function throughout the whole research process. Students had the responsibility to consult original knowledge sources and to share their scientific understandings with their groupmates, which fostered reciprocal accountability to each other, to the discipline, and to nature. After experimentation, the groups prepared a joint presentation to the whole class, followed by a discussion, thus further incorporating accountability to others and the discipline.

Second Year University Veterinary Science The clinical case-based task was designed to provide students in veterinary science with an opportunity to apply their primary preclinical knowledge, studied so far at a university, to the underlying principles of treatment and management of a real-life clinical case. Through generating and researching learning objectives that addressed group-selected key aspects of their clinical case, students had opportunities to problematize their case within the group naturally and made decisions on the areas of specific interest to the group. After researching their designated aspect, each group member could then position her/himself as the author of the content ideas s/he explored in greater depth. Students' responsibility to consult original sources and share their scientific understandings with peers was expected to foster reciprocal accountability to each other and the discipline. Accountability to the discipline was further incorporated by having each group present their clinical case as a whole to the class, followed by an extended question and answer session with teachers and clinicians.

Fourth Year University Engineering The Virtual Chemical Vapor Deposition (VCVD) process development project was designed to provide a bridge between university and industry by appropriating the practices, values, and goals of an industrial work group in a professional context. The groups were challenged with an open-ended design task, to develop a "recipe" of input parameters for an industrial reactor while considering competing constraints. The groups had the authority to develop one of many possible solutions using their own solution path. Concurrently, group members were accountable to their groupmates, to the discipline (through design meetings with the instructor who acted as their industrial supervisor), and to nature (through the output values of virtual experiments). Problematizing was critical as the groups addressed a design task with competing constraints and limited financial resources (students were

charged virtual money to perform experiments). However, they also had access to resources that included their prior coursework, introductory lectures, internet resources, and feedback in the design meetings. The first part of this project is studied here, in which the groups collaboratively generated an initial design strategy that they described in a memorandum.

In all learning environments, the students were assumed to adopt a role as a practicing professional: i.e., environmental research trainee, veterinary clinician, or process development engineer. The engineering learning context was, however, distinct from the other two learning contexts in that it involved a complex design process. In this context, there were no fixed algorithms or series of steps and the groups navigated through the design process using non-linear and iterative cycles of design ideation and analysis. While the project is structured around weekly milestones and deliverables, the work itself follows a unique path that unfolds differently for each group. Within the resources that are available to the students (e.g., feedback from the instructor, output data from the experiments, information in the technical literature, and knowledge and skills from prior engineering coursework), groups develop strategies and adapt or abandon them in a way that is iterative, open-ended, and temporally emergent (Pickering 1995). For any fixed point in the project, groups may be at very different stages in such a design approach. Therefore, taking a small sample of data at a fixed time may catch groups at very different places in task completion, and thus impede the comparability of the two groups relative to the other two contexts studied.

Interaction Data Analysis

Coding Cognitive Activity Two distinct meaningful and self-contained collaborative interaction phases were identified and included in the analyses (see Table 1). Identifying the distinct phases was relatively easy for the high school science data and for the university veterinary science data, both of which had instructor-defined project phases. It was more challenging for the university engineering data, where the phases emerged through how the student groups chose to proceed. The two distinct task phases were as follows:

- For the high school science task: (i) generating a hypothesis and (ii) analyzing results and preparing a presentation;
- For the university veterinary science task: (i) generating learning objectives and (ii) constructing a clinical case map; and
- For the university engineering task: (i) initial information gathering and problem scoping and (ii) writing the design memorandum.

Next, a segment (approximately 10–16 min) of each phase was selected for in-depth analyses. Criteria for selection were that both segments were crucial for task performance and completion, and also required students' collaboration (verbal interaction). The decision to use segments of around 10 to 16 min for in-depth analysis was necessary to keep the systematic and in-depth data coding manageable and to ensure the length of the selected segments were similar across phases and groups. The finding that no statistical differences in cognitive activity were observed between phases for each group (see the "Results" section for more detail) supports the choice of 10 to 16 min for in-depth analysis.

Across all sites, the small groups' verbal interactions formed the basis of the analyses. One group member's talk, whether it represented one word, or a nod (yes) or shaking head (no), or several

Table 3 Cognitive activity coding categories and examples across learning contexts

Level	Low	High
Definition (see Khosa and Volet 2014)	Talking about the content without conceptual justification or explanation, e.g., identifying, collating, or sharing factual information or discussing practical considerations	Talking about the content with justification or explanation, that is displaying engagement in the deeper understanding. Using concepts and practices of the discipline, e.g., meaning making with conceptual linkage
High-school general science	Anna: <i>“How many samples do we need?” (the task asks the number of samples)</i> Emma: <i>“Do not put any number, let’s write only.”</i> Anna: <i>“Let’s put that we need many samples...”</i>	Emma: <i>“What does happen if pH is higher but temperature is the same...”</i> Anna: <i>“And when it is higher temperature...”</i> (continues thinking that Emma started and tries to explain)
Second year university veterinary science	Renee: <i>“Okay so blood, gas, and electrolytes.”</i> Maddox: <i>“That’s normal.”</i> Renee: <i>“Yeah it does not seem very interesting to me.”</i> Thea: <i>“Did he have toxic neutrophils?”</i> Renee: <i>“Toxic neutrophils? Where are you reading that?”</i>	Thea: <i>“Theoretically weight loss can cause Azotaemia...”</i> Renee: <i>“It can? Okay...”</i> (Goes back to board and continues to draw arrows) Thea: <i>“By breaking down more protein, therefore you have got more urea in your system...”</i>
Fourth year university engineering	Bodhi: <i>“Yeah other testing...”</i> Alex: <i>“...for our next set of testing, I agree with it for our first testing...”</i> Alex: <i>“For our next set of testing, we test the middle, we test the radial we test like some middle point somewhere else.”</i> Marsden: <i>“Sure.”</i>	Alex: <i>“Yeah. SCCM a little bit of that, we are working here and we are here. So, we are really close to um some of their ratios so here where you see the 4.3, this is a 40 Å per minute. Um they were working at 350 mTorr which is what we are going to be working at um and a 5:1 ratio and um...”</i> Bodhi: <i>“... but we are doing 10:1.”</i> Alex: <i>“... yeah we are doing 10:1 but so I mean uh and then the other zero of flow rate of 66 um so if we go to somewhere where we kind of so I guess we should get ...”</i>

sentences of talk, was counted as one interaction turn until another group member spoke. Cognitive activity was coded into three categories: off-task, low-level, and high-level cognitive activity. These categories are based on Volet et al. (2009), Volet et al. (2013), Khosa and Volet (2014), and Koretsky et al. (2014a). Low-level cognitive activity included organizational tasks or confirming information. High-level cognitive activity included interactions that represented knowledge co-constructing, scientific meaning making, and conceptual linkage. Table 3 presents examples of low- and high-level cognitive activity from each learning context.

Due to the nature of the tasks and the number of group members participating (ranging from three to six), the number of interaction turns that were analyzed differed in the three learning contexts. For example, the high school science groups spent more time exploring information on the computer screen together and had fewer verbal interactions, whereas veterinary science students were engaged in joint discussion most of the time. Also, since all interactions were included, even if three students simultaneously said “yes,” the number of group members could also influence the total number of turns (interactions).

Inter-Rater Reliability In all three data sets, the principal coder was a native speaker. The principal coders also acted as inter-coders across data sets to ensure a reliable coding process.

After the principal coder had completed the coding, the inter-coder assessed approximately 20% (19.9–25.6%) of the excerpts from both groups (low- and high-outcome groups) and both phases (phases 1 and 2) of the other sites. Inter-rater reliability for cognitive activity was at least substantial (see Landis and Koch 1977, p. 165) in all datasets: high school general science, 87.6% (Cohen's kappa = 0.74); university veterinary science, 93.2% (Cohen's kappa = 0.85); and university engineering, 90.6% (Cohen's kappa = 0.79). There were only minor disagreements and these were resolved by discussion.

Statistical Analyses We used statistical analysis to determine if there were differences in the frequency of low- and high-outcome cognitive activity based on group outcome. Statistical significance was determined using non-parametric statistical tests with an alpha level of 0.05. The differences in coding distributions were first evaluated using logistic regression with cognitive activity (high or low) as the dependent variable, group outcome as the independent variable, and phase and site as covariates. Based on the results of that first regression, the cognitive activity at each site was then analyzed individually with group outcome as the independent variable.

Identification of the Nature of Students' PDE Exploration of PDE (ResQ2) involved qualitative analyses of the same 10–16 min of data used for the cognitive activity analyses from the three contexts and included both the high- and low-outcome groups.

First, transcripts were analyzed in the three contexts independently of one another for evidence of students' engagement that reflected one or more of Engle's (2012) four supporting conditions for PDE (authority, accountability, problematizing, and resources). In particular, the high-level cognitive activity episodes were explored for each group from both phases for instances in which PDE was present. Researchers in each context also focused on exploring the segments for interactions in which more than one of Engle's (2012) PDE conditions were simultaneously present. Once high-level cognitive-coded interactions had been explored, episodes of low-level cognitive activity were also searched.

Next, the research team agreed on two examples from each of the low- and high-outcome groups at each site, in order to provide best examples of data in which the presence of Engle's (2012) supportive conditions for PDE was visible in group members' interactions. The aim was that examples represent different aspects of the conditions for PDE (authority, accountability, problematizing, and resources) relative to the group's particular learning context. If a group had minimal high-level cognitive activity and PDE was not evident, and instead mostly low-level cognitive activity (e.g., this was case in the low-outcome high school and the veterinary group's collaboration in the second phase), then examples identified from an episode of low-level cognitive activity were selected to highlight missed opportunities for PDE (for example, as shown in Tables 5 and 6).

The length of the high-level cognitive activity episodes, from within which evidence of PDE was explicitly searched for, varied. The minimum requirement for an episode was two turns from different students within the group, but there was no upper limit of the length of the episodes. Cognitive activity episodes in which PDE was evident varied from this minimum amount of turns to an episode with over 100 turns in the veterinary science low-outcome group; however, that particularly long episode was an exception. For the high school groups, the length of high- or low-cognitive episodes was typically from two turns to over ten turns. For the engineering groups, it varied from two turns to over 30 turns. The 12 resulting

Table 4 Level of cognitive activity in different outcome groups across learning contexts

Level of education and discipline	High school general science		Second year university veterinary science		Fourth year university engineering	
	Low	High	Low	High	Low	High
Total cognitive activity, f	303	239	605	580	278	270
At high level, f	30	64	50	196	136	69
% at high level	9.9%	26.8%	8.3%	33.8%	48.9%	25.6%
Across groups	16.5%		20.7%		37.4%	

examples, which are shown in Tables 4, 5, and 6 below, illustrate each group's thinking processes and how that thinking directed the way the group proceeded with the task.

Results

Findings from the systematic analyses of the cognitive activity are presented first, since they serve as a basis for the subsequent PDE analyses. The findings and illustrations of the student groups' interactions in light of PDE are presented next, in an attempt to explain the possible differences between the groups from the perspective of disciplinary discourse.

Cognitive Activity in Three Distinct Learning Contexts

We first determined if there were statistical differences in the frequency of low- and high-level cognitive activity based on group outcome. Logistic regression of the entire data set revealed that high-level outcome groups showed significantly more high-level cognitive activity than low-level outcome groups (chi-square = 46.95; $df = 1$; $p < 0.001$). The covariate project phase did not significantly correlate with cognitive activity (chi-square = 3.49; $df = 1$; $p = 0.066$); however, the covariate site was significant, with high-level cognitive activity significantly increasing from the high school level to the 2nd year university level and then to the 4th year university level (chi-square = 87.86; $df = 2$; $p < 0.001$). In fact, site had a larger effect size than group outcome.

Table 4 shows the breakdown by group outcome level within site. Since phase was not significantly different, we removed that variable from analysis and performed separate regressions for each site. The high-outcome groups at the two lower education levels (high school and 2nd year university) spent a significantly higher percentage of their cognitive activity at a high level (26.8 and 33.8%, respectively) than did their low-outcome group counterparts (9.9 and 8.3%, respectively) (HS: chi-square = 29.64; $df = 1$; $p < 0.001$; V: chi-square = 99.27; $df = 1$; $p < 0.001$). However, the opposite was observed for the 4th year university site, where the lower outcome group showed more high-level cognitive activity (E: chi-square = 18.56; $df = 1$; $p < 0.001$).

In summary, it was clear that the low-outcome groups in high school general science and second year university veterinary science were also the *low-performing groups* in terms of process data. They worked at the lower (i.e., more superficial) cognitive level more frequently than the high-outcome groups, which supports the first hypotheses (H1). The same did not apply to 4th year engineering students, in which case task outcome and cognitive activity did

Table 5 Manifestations of PDE among high school students' interactions in general science

Analytical comments		
Low-outcome group		
Extract 1		
Emma: <i>Yes, it would have been good to know that if higher temperature and the cool is the same, so what is then ...</i>	Two different extracts were selected for this group displaying dominantly low-level cognitive activity. Extract 1 illustrates one of the rare episodes where the students' discussion partly features high-level cognitive activity. The group is thinking about the knowledge they lack while trying to figure out what kind of hypothesis to make. This episode displays some connection to the supporting conditions for PDE, namely, how the students are basing their thinking on the <i>resources</i> provided by the ViBSE. Further, this discussion shows <i>evidence of problematizing</i> and the end a <i>hint of authority</i> , but this remains kind of hanging in the air, unelaborated, as the group is prone to seek answers from resources rather than through shared reasoning. Engagement in deeper problem-solving halts, and attempts to use the available resources is not sufficient to achieve deep learning of science concepts and practices.	
Olivia: <i>Let's read once again, put it back.</i> (takes mouse, and the all start to read)		
Olivia: What is it? ... Here it is said about pH. ... What is this?		
Emma: To change, I guess.		
Olivia: Changes between levels depending on whether it is daytime or night		
Emma: <i>So do they come up at night?</i>		
Olivia: No, it is sad just to change here, here is...		
Anna: Or so		
Olivia: So (like agrees)		
Anna: There is variations and ...		
Olivia: Hmm...		
Anna: Should we put, do we start so that "We hypothesize that"?		
Emma: <i>So what do we hypothesize?</i>		
Extract 2		
Anna: First is planning (refers how to proceed in their research process)		Extract 2 shows that the group remains at a low cognitive level without a focus on the content. Although Anna and Emma display an awareness of their planning difficulties, their engagement is not productive. As in Extract 1, there is <i>no evidence of problematizing</i> the task <i>or displaying authority</i> in the research process. Rather, they take the task "as given," expecting the problem to be ready-made ("It's not clear what is a question in this research"). Also, they <i>do not show accountability to each other or the discipline</i> . They seem to proceed mechanically, just performing the task, which is a low-level cognitive activity.
Emma: Research plan (elaborates Anna's turn)		
Anna: It's not clear what is a question in this research (starts to create a research question as a part of planning)		
Emma: Confusing (shows uncertainty of the research question)		
Anna: There are so many things here (explains reason for uncertainty)		
Emma: (starts to list things)		
Anna: How many samples do we need? (continues planning)		
Emma: Don't put any number, let's write only		
Anna: Let's put that we need many samples so that we get needed results of copepods, so many as possible		
High outcome group		
Extract 1		
Ellen: <i>What do we know about these issues?</i>	Both extracts from this group's interactions feature high-level cognitive activity. In Extract 1, the group displays evidence of taking up the opportunities for PDE offered by the ViBSE learning environment in almost every individual contribution. Extract 1 was the starting point of the group collaborative learning process. Unlike their low-outcome counterpart, this group <i>problematized</i> the task at hand by pointing to their lack of prior content knowledge (e.g., "What do we know about these issues?") and did not just check what the material stated. This group also displayed <i>authority and accountability to each other</i> by sharing what they were thinking of the phenomenon and concluding they did not know much. Importantly, they also showed <i>accountability to the</i>	
Sofia: <i>So what do we know about ...?</i>		
Ellen: And about other these things?...		
Ellen: <i>I do not know so much of this, I do not know about...</i>		
Sofia: <i>I don't know anything, really, so it's bad ...</i>		
Ellen: <i>Why pH is important? So pH is, I don't know what's the situation but it's probably in a bad condition</i>		
Sofia: In plankton, ecosystem (links the meaning to the larger entity), <i>but what things are included in it</i>		
Paula: <i>I don't know anything else that...</i>		
Sofia: <i>Why is pH important?</i>		

Table 5 (continued)

	Analytical comments
Ellen: <i>Its effect on who can live there, is not it so?</i> (in Baltic Sea)	<i>discipline</i> , especially, in attempting to account for why pH is important.
Sofia: <i>Probably so that some planktons cannot live there...</i>	
Paula: <i>We knew a lot, indeed</i> (laughs because means opposite)	
Extract 2	
Sofia: <i>So what did happen to the eggs... our hypothesis was wrong because...?</i>	This extract also features high-level cognitive activity, aimed at assessing the group thinking while generating and testing their hypotheses empirically. They are <i>problematizing</i> the phenomenon while creating a hypothesis, conducting an experiment, and drawing conclusions. The group displayed <i>authorship and accountability</i> of their ideas by comparing them with the discipline knowledge provided by the software.
Paula: <i>...they developed</i>	
Ellen: Yes, it was	
Paula: <i>...it affected positively</i>	
Ellen: <i>... why does it go this way</i>	
Paula: <i>We were quite wrong</i>	
Ellen: <i>We were completely wrong</i>	
Paula: It doesn't matter because it has to be learning	
Ellen: Yes, this is learning	

not match as it did in the other two groups. Next, PDE analyses were aimed to elucidate these qualitative performance differences.

Productive Disciplinary Group Engagement in the Three Learning Contexts

The key question of this study was to examine how the low- and high-outcome student groups across learning contexts take up affordances for PDE in learning environments designed to support Engle's (2012) conditions for PDE. Across all contexts, PDE was reflected in discussions and the differences were characteristically not only quantitative but also qualitative. The PDE of the high school students typically focused on *problematizing* and was quite often paired with *authority* or, to a lesser degree, *accountability*. Similarly, the 4th year engineering student groups' discussions focused on *problematizing*, often accompanied by *authority* or *accountability*. In contrast, second year veterinary science student groups engaged more evenly in all four supporting conditions: *authority*, *accountability*, *problematizing*, and *resources*. Interestingly, from time to time, the university students' talk reflected simultaneously *all* four supporting conditions for PDE within a single episode (though much less so for the low-outcome veterinary science group, as described below), which never occurred with the high school students. Across all contexts, indications of PDE were more discernible when the students were discussing at the high cognitive level than at the low level.

Further, there were distinct differences between the low- and high-outcome groups in high school science and veterinary science, supporting the second hypothesis (H2); that is, compared to low-outcome groups, high-outcome groups' discussions typically took place at the high cognitive level and more frequently contained the four supporting conditions of PDE. In both the high school and second year university contexts, the differences between the low- and high-outcome groups' PDE were striking (noting, as earlier mentioned, the way in which the level of cognitive activity also differed considerably between these groups). Differences in PDE are well illuminated in the extracts that follow. As in the case of cognitive activity, the second hypothesis was not supported by the 4th year engineering students, where PDE was more evident in the low-outcome group.

Table 6 Manifestations of PDE among second year university students' interactions in veterinary science

	Analytical comments
<p>Low-outcome group</p> <p>Extract 1</p> <p>Ghita: Yeah, I think dog fight can go off to the side (referring to the card on the board)</p> <p>Mark: Yeah, I was thinking that</p> <p>Sheila: <i>But dog fight could be a cause</i></p> <p>Jade: It could be a cause</p> <p>Ghita: <i>But it's not in our case we do not know what the cause is</i></p> <p>Sheila: <i>It's possible</i></p> <p>Ghita: It is</p> <p>Extract 2</p> <p>Mark: And then, each one of these leading to all of those? (Referring to cards on board)</p> <p>Sheila and Drita: Yep</p> <p>Ghita: Yep</p> <p>Jade: Mmmhmm</p> <p>Mark: Ah what have I got? Like that really (... draws the lines and arrows)</p> <p>Jade: yep</p> <p>Ghita: yeah</p>	<p>Both extracts for this group feature low-level cognitive talk. Here, the group is trying to determine which concept cards belong in the case and which might be distractors. Their productive engagement with the task is in the realm of <i>problematizing</i>, as the group attempts to define the parameters of the case while not altogether ruling out possibilities that may warrant consideration. The talk is opinion-based and not explicitly supported by disciplinary knowledge.</p> <p>Extract 2 is representative of a long (100+ turns) episode of low-level cognitive talk, with minimal linkage of construction of the case map with the group's clinical case file or background research. Members tend to readily agree, with little evidence of scientific engagement, clinical exploration, or justification of how and why concept map linkages were made. Tentative, <i>low-level accountability or authority</i>, often unsourced or unreferenced, was occasionally observed. When one member did <i>attempt to halt the map-making to problematize</i>, uptake was typically brief or cursory and met with generic low-level reference to resources, often posed as a question such as "did not it say somewhere...?", i.e., without accountability or authoritative use of resources. Task completion appeared mechanical and divorced from disciplinary engagement.</p>
<p>High-outcome group</p> <p>Extract 1</p> <p>Thea: ... <i>because retinal detachment relates to why the blood pressure was taken in the first place. That was our clinical sign as well as-</i></p> <p>Renee: Okay so it would have a link here and here?</p> <p>Blanca: Yeah</p> <p>Thea: <i>Yep. Probably between the retinal detachment and the amlodipine because the amlodipine would have stopped...</i></p> <p>Renee: Do you want me to start drawing?</p> <p>Thea: Yes</p> <p>Renee: Okay. So one-way arrow?</p> <p>Thea: Two-way</p> <p>Renee: Well it's a cause and effect is not it?</p> <p>Thea: <i>I'd, I'm not sure because you treat hypertension with amlodipine and then amlodipine modifies your hypertension so I'm not sure which way... we would put that arrow</i></p> <p>Extract 2</p> <p>Thea: ...<i>what I'm saying is we've looked into proteinuria quite a bit and proteinuria is exacerbated by hypertension and yet causes hypertension so, it's a key thing</i></p> <p>Blanca: Yeah</p> <p>Thea: <i>I haven't looked into azotaemia so much because I thought proteinuria was key in this case</i></p>	<p>Extract 1 displays evidence of high cognitive activity, during which members offer or request clarification of understanding, and continually challenge one another's interactions, which are steeped in the discourse of the discipline. Elements of <i>accountability and authority</i> come into play as resources, and clinical knowledge, background research, and wider discipline knowledge are deployed through problematizing the issues. In contrast to the low-performing group which, even when well into the task maintained comfortable low-level interactions about card placement for many turns, this high-outcome group offered and <i>demanding accountability, engaged authoritatively and utilized clinical and discipline knowledge</i> in their map construction from early in the task and throughout the entire process.</p> <p>Extract 2 typifies the way in which this group went about the task. It shows how the group continued to undertake the construction of their map jointly, and also how they engaged in interactions that reflected all four of Engle's (2012) supporting conditions of PDE. For example, the <i>authoritative</i> use of resources closely linked to the group clinical case notes and</p>

Table 6 (continued)

	Analytical comments
<p><i>but you would expect that azotaemia means that you're retaining more urea in your blood which would lead to lower concentration of your urine... Urea absorption...</i></p> <p>Renee: <i>Hang on if you're having more urea in your blood...</i></p> <p>Thea: Yes</p> <p>Renee: <i>...because your kidneys have packed up wouldn't your blood pressure rise too? 'Cos you've got more things in your blood?</i></p>	<p>research produced accountability. Like the high-performing high school group, this group built from <i>problematizing</i> early on, evolving over the task to increasingly engage in ways that reflected all four of Engle's conditions for PDE at increasingly advanced levels and from all group members. Engaging in a disciplinary talk, during which they also debated one another's thinking, thus enabled the learning affordances of the task to be fully realized, leading to a high-level outcome.</p>

Examples of Interactions of Low- and High-Outcome Groups in the Three Learning Contexts

Extracts of interactions that best illustrate how low- and high-outcome groups seized opportunities provided by their respective learning environment to engage in PDE were selected from each of the three learning contexts. Two extracts from each outcome level (low, high) are presented in Tables 5, 6, and 7, accompanied by analytical comments on the nature of the group's engagement in PDE. The most visible indicators of interactions that mirror Engle's (2012) supporting conditions for PDE (problematizing, accountability, authority, and resources) are shown in italics. A cross-context analysis is presented at the end of the section.

High School Students' Interactions in General Science Table 5 shows extracts from this learning context. For both groups, extract 1 occurs during the phase in which students had to generate a hypothesis for their project and extract 2 during the phase in which groups had to analyze their results and prepare a class presentation. The low-outcome group's interactions are consistent with their low performance, and the qualitative search for interactions that mirrored the supportive conditions of PDE offered to them revealed the nature of "missed opportunities" for PDE in this group. Overall, the analysis of these two groups' interactions highlight that although the learning environment of ViBSE provided opportunities for all groups to engage in problematizing, using disciplinary resources such as real research data, assuming intellectual authority, and testing how their ideas made sense in relation to those of disciplinary experts, these opportunities were taken up only in the high-outcome group. Importantly, group differences in the quality of their engagement were consistent with differences in their outcome level.

Second Year University Students' Interactions in Veterinary Science In this learning context, all extracts are selected from the phase of the overall task during which students constructed their clinical case map after having undertaken background research (Table 6). In the low-outcome group, there was only minimal problematizing during this phase and instead, there was mostly general agreement among members about placing cards and arrows. Although the task afforded opportunities for problematizing and deep disciplinary engagement, most of the time this group "skimmed the surface" of the requirement to produce the map. This level of engagement resulted in the construction of a map largely disconnected from their clinical case and background research regarding direct, definitive linkages. In contrast, the high-outcome group displayed escalating robust engagement in disciplinary talk that reflected

Table 7 Manifestations of PDE among fourth year university students' interactions in engineering

	Analytical comments
<p>Low-outcome group</p> <p>Extract 1</p> <p>Marsden: <i>So, uh, you'd want an excess of the ammonia though</i></p> <p>Alex: <i>But, okay, but when we have, so, so that was my point. If we have an excess of the ammo-. Well wait. Just because the, the dichlorosilane is nasty stuff, doesn't, why, that doesn't mean that we need to have more ammonia. That's not a reason for it</i></p> <p>Bodhi: <i>So, the-</i></p> <p>Marsden: <i>Yeah, but the ammonia means it's going to speed up the reaction</i></p> <p>Alex: <i>Yeah</i></p> <p>Extract 2</p> <p>Alex: [talking while writing on whiteboard] <i>Well, I think we ... need to know ... these reaction rates because if we have extra ... of our ammonia, it's going to end up reacting with this ... That's why we need to know. And we need how much</i></p> <p>Bodhi: <i>Well it's going to do it anyway</i></p> <p>Alex: <i>It is, but ... we can try to minimize it</i></p> <p>Bodhi: <i>Or you just pump the ... out of ammonia and 'cause it's cheap and easy</i></p> <p>Alex: <i>We don't care about everything else that comes out?</i></p> <p>Marsden: <i>Nope</i></p> <p>Alex: <i>Well</i></p> <p>Bodhi: <i>I mean if the overall reaction has no NHCl in it, or not NH. Sorry. HCl. Then, no your overall has none</i></p> <p>High-outcome group</p> <p>Extract 1</p> <p>Morgan: <i>It does have ballpark ranges for ... flow rates. But I don't know how useful it will be.</i></p> <p>Charlie: <i>I mean it's a different-</i></p> <p>Morgan: <i>Different system</i></p> <p>Charlie: <i>Yeah, it's- I mean it's a different- totally different thing. Yeah, I would assume there would be a big difference.</i></p> <p>Extract 2</p> <p>Charlie: <i>Yeah 'cause I mean like—so like right here you are gonna have your highest concentration</i></p> <p>Morgan: <i>Yeah</i></p> <p>Charlie: <i>And you could [pause]. Like I guess you - you probably want - you probably want your lower temperature here, because you don't want the reaction rate to go that fast right here.</i></p> <p>Morgan: <i>Yeah</i></p> <p>Charlie: <i>Because if you react faster down here you're gonna have less material up here and you're gonna have way thicker wafers down here.</i></p> <p>Morgan: <i>Yeah, yeah- no, I swear that- we had to find like the...</i></p>	<p>Both extracts selected for this group feature high-level cognitive activity. In stark contrast to the low-outcome groups in the other two contexts, this low-outcome group displayed sustained engagement in ways that reflected Engle's supporting conditions for PDE for example, <i>problematizing</i>, e.g., "Just because the, the dichlorosilane is nasty stuff doesn't, why, that does not mean that ... " and <i>accountability to science</i>, e.g., "the ammonia means it's going to speed up the reaction."</p> <p>Deep-level engagement in the discourse of the discipline is also evident in Extract 2, which typifies interactions over an episode within which the issues at hand were <i>problematized</i> and jointly regulated. Members challenged one another in ways that reflected Engle's four supporting conditions for PDE. For example, Alex debates ideas, during which Alex exhibits disciplinary <i>accountability</i>, while also demanding <i>accountability</i> from peers. Marsden and Bodhi respond with disciplinary <i>authority</i>. The extract also demonstrates wider professional <i>accountability</i>, for example to community and environment: "We do not care about everything else that comes out?"</p> <p>Extract 1 features the group working on its first test parameter values for the memorandum, and evaluating resources. Morgan realizes that the resource may or may not be useful, and Charlie responded with a disciplinary assumption. This brief interaction focused on resource evaluation was promptly co-determined, suggesting that one another's evaluations were "taken as given." There was <i>no evidence of problematizing or resources brought in with authority and accountability</i>.</p> <p>Here, the group is discussing the test parameter values. They demonstrate a co-constructed understanding that forms the basis from which they <i>problematize</i> the issue at hand, progressing by building on contributions more so than questioning one another. Their contributions converge in a way that suggests implicit disciplinary <i>accountability</i> and <i>authority</i> among members, e.g., Charlie displays <i>accountability</i> and <i>authority</i> in explaining: "right here you are gonna have your highest concentration," and "... because you don't want the reaction rate..." Morgan listens, agrees: "Yeah, yeah," then seamlessly contributes understanding: "oh ... you need to know if</p>

Table 7 (continued)

	Analytical comments
<p>Charlie: <i>So I say T5 is probably gonna be ...</i> Morgan: <i>Oh, you need to know if it's gonna be hotter as you go up-</i> Charlie and Foster: <i>Yeah</i> Morgan: <i>- then you have to find like the concentration gradient, and then match it to your temperature gradient, but we need to know the kinetics to do that.</i> Charlie: <i>Yeah.</i> Morgan: <i>Otherwise, we don't know what temperatures to use</i></p>	<p>it's gonna be hotter ...” which Morgan’s peers confirm, and Morgan: “then you have to find like the concentration gradient ...,” and so on. The group efficiently problematize and progress with a tacit agreement that appears underpinned by a bedrock of <i>accountability</i> and <i>authority</i>, enabling a disciplinary flow that results in a quality outcome.</p>

all four of Engle’s PDE supporting conditions, which led to the construction of an accurate concept map of their clinical case.

Fourth Year University Students’ Interactions in Engineering Both extracts from this learning context are from the phase in which the groups undertook initial information gathering and problem scoping toward production of a design memorandum to present to their supervisor (Table 6). In combination, the two extracts of the low-outcome engineering group reveal how all three members were deeply engaged in ways that exhibited all four of Engle’s supporting conditions for PDE. There was evidence that their high-level cognitive activity was directed at developing a shared understanding of the issue at hand, rather than assuming everyone was “on the same page” about it. Their deep-level engagement appeared remarkably similar to that of the high-outcome veterinary science group and was, therefore, inconsistent with their low-outcome level. In contrast to the low-outcome group’s interactions that reflected all of Engle’s supporting conditions for PDE in their high-level cognitive talk, the high-outcome group seems to demonstrate implicit agreement of assumed and reciprocal disciplinary authority, and a degree of intellectual trust from which they proceeded.

Summarizing PDE across Three Learning Contexts

We summarize the PDE results as follows:

First, both the low-outcome high school science group and the low-outcome second year veterinary science group showed minimal accountability to science in their learning process. While there was evidence of uptake of some of Engle’s (2012) supporting conditions for PDE (mainly through problematizing within the realm of low-level cognitive activity) within these groups, this was not sufficient to facilitate the development of enriched disciplinary understanding in the absence of other aspects, as illustrated by their low outcome and the nature of their science talk.

Second, notable in all high-outcome groups across the three contexts, was that group members tended to define parameters and gaps in their disciplinary knowledge. In the high school science high-outcome group, for example, interactions such as “What do we know about these issues?” (Extract 1). In the high-outcome veterinary science group, gaps were expressed using language such as “I’d, I’m not sure because you treat hypertension with amlodipine and then ...” (Extract 1) and, in the engineering group, as “But I don’t know how

useful it will be” (Extract 1). This practice showed that the most successful groups continually and explicitly evaluated aspects of their disciplinary knowledge.

Third, at the most advanced level, the high-outcome group in engineering departed from the observed trend. Contributions to PDE appeared to be increasingly collective among group members at this site. While this group was engaged in deep-level processing of the task, their implicit shared understanding may have prevented manifestations of the supporting conditions for PDE to be observed in their verbalizations to the same extent as the low-outcome group.

Discussion

This study aimed to explore PDE in collaborative science learning across different educational contexts and levels. The three collaborative learning environments in this study were all designed to encompass Engle’s four conditions for PDE (authority, accountability, problematizing, and resources). The first aim was to explore the student groups’ cognitive activity and, specifically, how the groups’ collective learning outcomes are associated with the quality of cognitive activity in different contexts. Based on outcomes of cognitive activity, the second key aim of the study was to establish how the learning environments’ supporting features of PDE manifest during group interactions and, particularly, how PDE is associated with low and high cognitive performance in collaborative science learning processes in these three different contexts.

The novel contributions of the study are twofold: (i) analyzing both cognitive activity and PDE from the same ongoing collaborative processes using the same theory-based analytical methods on three different sets of data, thus avoiding the danger of the same actions in different datasets being classified differently; and (ii) exploring PDE in relation to the quality of cognitive activity in terms of attempts at deep meaning making and high-level collective achievement. Bringing together three elements of collaborative learning, namely level of cognitive activity, product outcome, and PDE, allowed the identification of patterns in the groups’ interaction processes that can help us better analyze the intended influence of the features of learning environments on student learning, as well as differences in productive engagement in science learning.

High Performance and Engagement in Science: Relations of the Quality of Cognitive Activity, Product Outcome, and PDE in Real-Time Interactions

On the basis of theoretical conceptualizations supported by current empirical evidence, it was assumed in the first research question that high-outcome groups would display a greater proportion of high-level cognitive activity (H1) (Khosa and Volet 2014; Volet et al. 2013; Volet et al. 2017). It was expected that this same pattern would be found across all three studied contexts. Hypothesis H1 was confirmed in two contexts: senior high school general science and second year university veterinary science. However, an opposite pattern of results was found with the university engineering student groups, in which case the hypothesis was rejected. The high-outcome group demonstrated less high-level cognitive activity during the phases of the task that were studied. This unexpected outcome is discussed in more detail in the next section.

Engle’s (2012) four supporting conditions of PDE were mirrored in all groups’ collaborative interactions, and most typically paired with high-level cognitive activity. Due to the

novelty of the second research question (ResQ2), how small groups seize opportunities for PDE, only one broad hypothesis (H2) could be generated. In line with the hypothesis, the relation of PDE and a high-outcome product was found in high school and university second year veterinary science students, but not with the fourth year engineering students. As expected, the qualitative analyses revealed that the interactions of the low-outcome high school science and veterinary science groups reflected Engle's (2012) supporting conditions for PDE less frequently and less richly than their high-outcome counterparts. However, they did show missed opportunities for PDE, suggesting that there is opportunity for teachers to interact with these groups in ways that could help support their disciplinary talk. Interestingly, in the high-outcome veterinary science student group, three or even four supporting conditions were often present at the same time. In the engineering context, the low-outcome engineering group frequently demonstrated three supporting conditions for PDE, namely problematizing, authority, and accountability, and sometimes all three simultaneously. Although the high-outcome engineering group also demonstrated these supporting conditions, particularly problematizing most often paired with authority, their disciplinary thinking was not as explicit as that of their low-outcome counterpart.

In sum, the comparison between the low- and high-outcome groups' collaborative learning processes provided evidence of the important relationship between group discussions, high-level cognitive activity, and collective outcome that reflect the four supporting conditions of PDE. However, the results for the engineering context differed from the two other contexts for both research questions.

Disciplinary Expertise in Explaining PDE at Different Professional Levels

There are plausible explanations for the unexpected results in the engineering context. One is contextual. In contrast to high school science and veterinary science task contexts, the engineering learning context is a complex design process, in which there is no fixed algorithm or series of steps for completion. Rather, the groups navigate through the design process using non-linear and iterative cycles of design ideation and analysis. Therefore, data from a sample "slice" in time may capture students in very different stages of thinking and interaction, which might impede the comparability of the two groups. However, the fact that project phase was not significantly different makes this explanation less likely.

A more intriguing, conceptually driven explanation revolves around different social practices as the students progress toward disciplinary expertise. The 4th year engineering students were approaching graduation, on the cusp of entering the profession. The analyses of the present study revealed some evidence of "tacit agreement" among members of the engineering high-outcome group that was less evident in the low-outcome group. Specifically, members of the high-outcome group appeared to proceed in a way that built on recognized, mutual intellectual authority and shared understanding as evidenced, for example, by frequently completing one another's sentences. To the extent that each of these students *implicitly* acknowledged that their peers "knew their stuff" and shared a common conception of the project, one could argue that there was less need for extensive high-level cognitive talk. As the students' conceptions of the design task became the same (i.e., a joint enterprise), it is reasonable to assume that there could be a more implicit intellectual authority realized in interactions of tacit agreement as observed. This explanation is supported by the assumption that, by the final year of study, some engineering students have a robust understanding of disciplinary concepts and practices and are capable of "thinking like engineers," especially in

the context of an industrially situated task. In contrast, the low-outcome group interactions, which contained more questioning-type discussion as members negotiated their conceptual understanding with one another, are consistent with a group of students less confident in their knowledge, but determined to make progress on the task by learning with and from each other.

The distinct patterns of interactions displayed by high- and low-outcome groups are reminiscent of the distinctions found between science experts and novices in studies on the nature of expertise (e.g., Ericsson and Pool 2016; Feltovich et al. 2006). For example, research on novices and experts in medicine has revealed that, with higher levels of expertise, some of the reasoning processes of experts are automatic and unconscious and are activated only in cases of mismatches or conflicts in processing (Boshuizen and Schmidt 2008). In this respect, it can be argued that the high-outcome engineering group showed the compiled conceptual understanding of engineering experts (Redish and Smith 2008).

Additionally, while the data sampling strategy for the four year engineering groups provides equivalent slices of data in time (temporal slices), we suggest the groups might be in quite different “social states” regarding the task at hand. This project was the third project of the term in which the students in each group worked together and they had been in the same program for four years of study. Thus, they brought with them previous conceptions of one another’s competencies and a history of collaborative interactions. Once they were “all on the same page,” there was no need to renegotiate their approach continually and, in fact, they could use that shared understanding to make their interactions more efficient. In other words, the two engineering groups may have brought different sociotechnical histories, both from within the project and longer term, to the examined interactions. The data from the high-outcome group can be interpreted as stemming from interactions that were not noticeably focused on developing a shared understanding of the foundational scientific principles, but rather on “getting on” with the task of design (Vincenti 1990). However, they were clearly in PDE and meaningfully making progress. This explanation is supported by observed behaviors of an expert group in this same project who, when compared to student groups, spent much more time focusing on their specific strategies to create, test, and evaluate design options, rather than negotiating the underlying core conceptual knowledge (Nolen et al. 2018).

Limitations

While we used the same framework and methodology across three sites in this study, they each had different tasks, students, and intended outcomes. This variation enables exploration of disciplinary engagement and cognitive activity across educational contexts, but also makes the study design unsuitable for claims about progression based on education level. However, when site was considered as a covariate, the data revealed a significantly steady increase in cognitive activity from high school to second year university and then to fourth year university. This trend is consistent with the notion that PDE embodies professional disciplinary behaviors that are more likely to be found in advanced university professional programs than in senior high school science classes. Research designs that can isolate and investigate this developmental progression are needed.

Two of the researched tasks focus on engaging in roles that directly align with disciplinary practice (the high school experiment, the engineering design), whereas the veterinary task had a more academic conceptual focus and required the development of a concept map for a veterinary case. Groups engaged in disciplinary practice tasks tended to focus on *problematizing*, often accompanied by *authority* or *accountability*. In contrast, the groups

engaged in the more conceptual task more evenly engaged in all four of Engle's supporting conditions. Research is needed to identify the connection between the nature of the task and the ways collaborative groups take up PDE.

Concluding Remarks

To conclude, consistent and expected patterns of cognitive activity and PDE in relation to product outcomes emerged in the high school science and university veterinary science contexts, whereas a different pattern was found in engineering. The novel and unique feature of this study was the use of the same coding scheme and coders across datasets. This enables controlled identification of discrepancies and correspondences across different learning contexts. Although overgeneralizations must be avoided, such as claims about developmental patterns, the application of the same methods of analysis underlines the effectiveness of this study's strategy in revealing this result. This strategy produced outcomes that offer clear directions for future research, such as further investigation of the role of disciplinary expertise and group members' history, in interpreting a group's social interactions regarding the quality of cognitive activity and PDE. For example, observing groups' learning processes over a longer time period (e.g., follow-up of the same students groups across cultural and educational contexts), in varying learning tasks (e.g., routine/traditional vs. novel problems demanding innovative thinking, disciplinary practice vs. conceptual tasks), and/or comparing high-outcome expert groups (e.g., in different disciplines) with one another would be helpful in identifying whether similar patterns emerge across novice and expert groups.

The present study further raises new questions for future research and practice about the application of the PDE framework (Engle 2012; Forman et al. 2014). Large differences between the high- and low-outcome groups in relation to cognitive activity and PDE, particularly in senior high school, warrants the question raised by Kelly (2014; see also Vauras et al. 2018) of what it means to incorporate authentic disciplinary practices in classrooms and how classroom teachers can help their students make the difficult transition between school science and the kind of instruction embodied by PDE.

Funding Open access funding provided by University of Turku (UTU) including Turku University Central Hospital. This work was supported by the Academy of Finland Grant No. 274117, by the Australian Research Council under the Discovery Award (DP150101142) and Murdoch University, Research and Innovation Office, and by the US National Science Foundation Grant No. EEC 1261930 (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 National Science Foundation).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Apedoe, X., & Ford, M. (2010). The empirical attitude, material practice and design activities. *Science & Education, 19*, 165–186.

- Boaler, J., & Sengupta-Irving, T. (2016). The many colors of algebra: the impact of equity focused teaching upon student learning and engagement. *The Journal of Mathematical Behavior*, 41, 179–190.
- Boshuizen, H. P. A., & Schmidt, H. G. (2008). The development of clinical reasoning expertise: implications for teaching. In J. Higgs, M. Jones, S. Loftus, & N. Christensen (Eds.), *Clinical reasoning in the health professions* (3rd ed., pp. 113–121). Oxford: Butterworth-Heinemann/Elsevier.
- Bowen, C. W. (2000). A quantitative literature review of cooperative learning effects on high school and college chemistry achievement. *Journal of Chemical Education*, 77, 116–119.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. Washington, DC: National Academy Press.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 32–42.
- Cook, S. D., & Brown, J. S. (1999). Bridging epistemologies: the generative dance between organizational knowledge and organizational knowing. *Organization Science*, 10, 381–400.
- Cornelius, L. L., & Herrenkohl, L. R. (2004). Power in the classroom: how the classroom environment shapes students' relationships with each other and with concepts. *Cognition and Instruction*, 22(4), 467–498.
- Engle, R. A. (2012). The productive disciplinary engagement framework: origins, key concepts and developments. In Y. Dai (Ed.), *Design research on learning and thinking in educational settings: enhancing intellectual growth and functioning* (pp. 161–200). New York: Routledge.
- 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 and Instruction*, 20(4), 399–483.
- Ericsson, K. A., & Pool, R. (2016). *Peak: secrets from the new science of expertise*. New York: Eamon Dolan books/Houghton Mifflin & Harcourt.
- Feltovich, P., Prietula, M., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance*. New York: Cambridge University Press.
- Ford, M. (2008). Disciplinary authority and accountability in scientific practice and learning. *Science Education*, 92, 404–423.
- Ford, M. J., & Forman, E. A. (2006). Redefining disciplinary learning in classroom contexts. *Review of Research in Education*, 30, 1–32.
- Forman, E. A., Engle, R. A., Venturini, P., & Ford, M. J. (2014). Introduction to special issue: International examinations and extensions of the productive disciplinary engagement framework. *International Journal of Educational Research*, 64, 149–155.
- Furrer, C., & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95, 148–162.
- Horn, I. S. (2005). Learning on the job: a situated account of teacher learning in high school mathematics departments. *Cognition and Instruction*, 23, 207–236.
- Johri, A., & Olds, B. M. (2011). Situated engineering learning: bridging engineering education research and the learning sciences. *Journal of Engineering Education*, 100, 151–185.
- Kelly, G. J. (2014). The social bases of disciplinary knowledge and practice in productive disciplinary engagement. *International Journal of Educational Research*, 64, 211–214.
- Khosa, D. K., & Volet, S. (2014). Productive group engagement in cognitive activity and metacognitive regulation during collaborative learning: can it explain differences in students' conceptual understanding? *Metacognition and Learning*, 9, 287–307.
- Kirschner, F., Paas, F., Kirschner, P. A., & Janssen, J. (2011). Differential effects of problem-solving demands on individual and collaborative learning outcomes. *Learning and Instruction*, 21, 587–599.
- Koretsky M, Nolen SB, Gilbuena DM, Lehtinen E, Vauras M, Tierney G, Volet SE (2014a) Studying and supporting productive disciplinary engagement in STEM learning environments. *Frontiers in education*. Madrid, Spain
- Koretsky M, Nolen SB, Gilbuena DM, Lehtinen E, Vauras M, Tierney G, Volet SE (2014b) Studying and supporting productive disciplinary engagement in STEM learning environments. Paper presented at 2014 ASEE Annual Conference, Indianapolis, Indiana. <https://peer.asee.org/23070>
- Koretsky M, Nolen SB, Tierney G, Wetzstein L (2015) Productive disciplinary engagement in design tasks. Research in engineering education symposium. Dublin, Ireland.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Lotan, R. A. (2003). Group-worthy tasks. *Educational Leadership*, 60(6), 72–75.
- National Council of Teachers of Mathematics (2014) Principles to actions: ensuring mathematical success for all. Reston, VA: National Council of Teachers of Mathematics.
- National Research Council. (2012). *A framework for K-12 science education: practices, crosscutting concepts, and core ideas*. Washington, D.C.: National Academies Press.

- Nolen SB, Koretsky MD, Wetzstein L, Scalone G, Goodell A, Chindanon K, Tierney G (2018) Productive disciplinary engagement in STEM design projects. Manuscript submitted for publication.
- Phillips, A. M., Watkins, J., & Hammer, D. (2017). Problematising as a scientific endeavor. *Physical Review Physics Education Research*, 13, 020107.
- Pickering, A. (1995). *The mangle of practice: time, agency, and science*. Chicago: University of Chicago Press.
- Redish, E. F., & Smith, K. A. (2008). Looking beyond content: Skill development for engineers. *Journal of Engineering Education*, 97, 295–307.
- Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science Education*, 93, 448–484.
- Scott, P. H., Mortimer, E. F., & Aguiar, O. G. (2006). The tension between authoritative and dialogic discourse: a fundamental characteristic of meaning making interactions in high school science lessons. *Science Education*, 90, 605–631.
- Sinatra, G. M., Heddy, B. C., & Lombardi, D. (2015). The challenges of defining and measuring student engagement in science. *Educational Psychologist*, 50, 1–13.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: a meta-analysis. *Review of Educational Research*, 69, 21–51.
- Stevens, R., O'Connor, K., Garrison, L., Jocuns, A., & Amos, D. M. (2008). Becoming an engineer: toward a three dimensional view of engineering learning. *Journal of Engineering Education*, 97, 355–368.
- Stump, G. S., Hilpert, J. C., Husman, J., Chung, W.-T., & Kim, W. (2011). Collaborative learning in engineering students: gender and achievement. *Journal of Engineering Education*, 100, 475–497.
- Vauras, M., Volet, S., & Nolen, S. (2018). Supporting motivation in collaborative learning: challenges in the face of an uncertain future. In E. Gonida & M. Lemos (Eds.), *Motivation in education at a time of global change: theory, research, and implications for practice*. New York: Emerald.
- Venturini, P., & Amade-Escot, C. (2014). Analysis of conditions leading to a productive disciplinary engagement during a physics lesson in a disadvantaged area school. *International Journal of Educational Research*, 64, 170–183.
- Vincenti, W. G. (1990). *What engineers know and how they know it: analytical studies from aeronautical history*. Baltimore: Johns Hopkins University Press.
- Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: how does it emerge and how is it sustained? *Learning and Instruction*, 19, 128–143.
- Volet, S., Vauras, M., Khosa, D., & Iiskala, T. (2013). Metacognitive regulation in collaborative learning: conceptual developments and methodological contextualizations. In S. Volet & M. Vauras (Eds.), *Interpersonal regulation of learning and motivation: methodological advances* (pp. 67–101). New York: Routledge.
- Volet, S., Vauras, M., Salo, A.-E., & Khosa, D. (2017). Individual contributions in student-led collaborative learning: insights from two analytical approaches to explain the quality of group outcome. *Learning and Individual Differences*, 53, 79–92.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Milo D. Koretsky¹ · Marja Vauras² · Cheryl Jones³ · Tuiki Iiskala² · Simone Volet³

¹ School of Chemical, Biological, and Environmental Engineering, 116 Johnson Hall, Oregon State University, Corvallis, OR 97331-2702, USA

² Department of Teacher Education, Faculty of Education, University of Turku, Assistentinkatu, 5, 20014 Turku, Finland

³ School of Education, Murdoch University, 90 South Street, Murdoch, WA 6150, Australia