Frontline Learning Research
An official journal of EARLI

Frontline **L**earning **R**esearch Vol. 10 No. 2 (2022) 1 - 21
ISSN 2295-3159

A Multidimensional Framework of Collaborative Groups' Disciplinary Engagement

Toni Kempler Rogat[1], Cindy E. Hmelo-Silver[2], Britte Haugan Cheng[3], Anne Traynor[1], Temitope F. Adeoye[1], Andrea Gomoll[4] & Brenda K. Downing[1]

[1] Purdue University, USA
[2] Indiana University, USA
[3] Menlo Education Research, USA
[4] Gomoll Research & Design, USA

Article received 17 May 2021 / Article revised 12 March 2022 / Accepted 27 July / Available online 9 September 2022

Abstract

This research is aimed at developing novel theory to advance innovative methods for examining how collaborative groups progress toward productively engaging during classroom activity that integrates disciplinary practices. This work draws on a situative perspective, along with prior framings of individual engagement, to conceptualize engagement as a shared and multidimensional phenomenon. A multidimensional conceptualization affords the study of distinct engagement dimensions, as well as the interrelationships of engagement dimensions that together are productive. Development and exploration of an observational rubric evaluating collaborative group disciplinary engagement (GDE) is presented, leveraging the benefits of observational methods with a rubric specifying quality ratings, enabling the potential for analyses of larger samples more efficiently than prior approaches, but with similar ability to richly characterize the shared and multidimensional nature of group engagement. Mixed-methods analyses, including case illustrations and profile analysis, showcase the synergistic interrelations among engagement dimensions constituting GDE. The rubric effectively captured engagement features that could be identified via intensive video analysis, while affording the evaluation of broader claims about group engagement patterns. Application of the rubric across curricular contexts, and within and between lessons across a curricular unit, will enable comparative studies that can inform theory about collaborative engagement, as well as instructional design and practice.

Keywords: Engagement; collaborative learning; STEM education; observational rubric

Corresponding author: Toni Kempler Rogat, Purdue University, 100 N. University Street, West Lafayette, IN 47906, USA, trogat@purdue.edu *DOI:https://doi.org/10.14786/flr.v10i2.863*

1. Introduction

In response to science standards, which call for students' integrated understanding of STEM content and practices, as part of collaborative activity (e.g., NGSS, 2013; Forsthuber et al., 2011), this research is aimed at advancing theory and methods toward understanding how collaborative groups come to productively engage in STEM activities. In this research, we build from work on individual engagement that advances a multidimensional conceptualization from individual engagement (Fredricks, et al., 2004) to the collaborative group context. To accomplish this, we need to consider interpersonal engagement and to account for collective group engagement practices. Further, individual perspectives on engagement do not represent recent theoretical advances regarding the social and situated nature of engagement (Gresalfi, et al., 2009; Ryu & Lombardi, 2015).

Other engagement-related frameworks, including Engle and Conant's *productive disciplinary engagement* (PDE) framework, have integrated developments from situative perspectives (Danish & Gresalfi, 2018; Engle & Conant, 2002; Greeno, 2006; Gresalfi, et al., 2009; Hand & Gresalfi, 2015; Hickey, 2003), including assumptions that engagement is co-negotiated in collective interaction, evolves in moment-by-moment interactions, and is contextualized in activity systems. These activity systems are comprised of instructional opportunities that support and constrain engagement, given curriculum materials, teacher scaffolds, tasks, disciplinary content and practices, and interactions among learners (Greeno, 2006; Shechtman, et al., 2012). Productive disciplinary engagement reflects deep-level engagement yielding intellectual progress during authentic disciplinary tasks (Engle & Conant, 2002) in which students grapple with central domain concepts while participating in the authentic disciplinary practices (Duschl, 2008; Forman & Ford, 2014). Engagement, and its interrelated constituent dimensions, are not merely *influences* on learning, but instead are *central to and inseparable* from learning (Gresalfi, et al., 2009). In this view, invested effort or persistence in the face of challenge, including interpersonal interactions, coordinated activity, and being strategic while making meaningful connections are central to what and how learners come to understand. This earlier work has characterized how shared practices are established in the collective, encompassing teacher-student and whole class negotiation, but with limited focus on the negotiation of norms within the collaborative group.

Drawing on this multidimensional conceptualization and situated framework of engagement, our goal is to understand collaborative *groups' disciplinary engagement* (GDE) as being comprised of interrelated, but distinguishable aspects of interaction in group activity. We omitted productive from this descriptor because we wanted to capture the range of variation in quality of disciplinary engagement from none or superficial (i.e., low) to high quality disciplinary engagement that is likely to be productive. We investigate this primary goal within three STEM curricula integrating collaboration and disciplinary practices as central design features. We developed and applied an observational rubric to assign quality ratings to explore the engagement profiles of more and less productive groups, and to characterize the synergies among engagement dimensions.

Building on earlier work (Sinha et al, 2015), we delineate five dimensions of group engagement (see Table 1). *Behavioral engagement (BE)* characterizes the degree to which a group jointly participates and persists on assigned tasks or chooses to go off-task (Fredricks, et al., 2004). Sustained group participation amongst group members yields potential for building from others' perspectives, while temporary off-task exchanges can reinvigorate positive interpersonal interactions when returning to task (Barron, 2000; Langer-Osuna, et al., 2020). Both socioemotional engagement and collaborative engagement are extensions to individual engagement dimensions, accounting for the interpersonal nature of group engagement (Linnenbrink-Garcia, et al., 2011). *Socioemotional engagement (SE)* characterizes the group's interpersonal interaction quality and climate, where positive climate involves the negotiation and maintenance of respectful and inclusive interactions, team cohesion, and psychological safety (Rogat & Linnenbrink-Garcia, 2011; Rogat & Adams-Wiggins, 2015). Negative socioemotional engagement involves disrespect and competence put-downs, which may result from challenge, conflict, or status differences can derail collaborative engagement (Adams-Wiggins, 2020;

Näykki, et al., 2014). Research on learning in collaborative groups indicates that positive socioemotional interactions elevate the quality of joint task work (e.g., friendly; supportive; fostering risk-taking) (Barron, 2000; Kreijns et al., 2002). *Collaborative engagement (CE)* considers groups' task and conceptual coordination in constructing knowledge, as well as the balance of participation amongst group members in making contributions. High-quality CE undergirds joint knowledge construction accounting for multiple perspectives and promotes the development of a shared problem space (Roschelle & Teasley, 1995), whereas low-quality CE is characterized by independent task contributions (i.e., low coordination) or a group member's efforts to control and direct the task (i.e., imbalance). *Metacognitive engagement (ME)* describes groups' use of regulatory strategies, including planning, monitoring, and evaluation (Rogat & Linnenbrink-Garcia, 2011; Järvelä, et al., 2016; Schoor, et al., 2015). Recent findings show that high-quality ME is differentiated by effective shared regulation which is goal-focused toward understanding and progress on the task, content, and/or disciplinary practices (Rogat & Linnenbrink-Garcia, 2013; Molenaar & Chiu, 2014; Khosa & Volet, 2014), which is supported by, rather than being the sole focus of, regulation of behavior, time, group process, and task completion. Finally, *disciplinary engagement (DE)* refers to the nature of the group's content and disciplinary activity, with high-quality DE reflecting connections toward integration of conceptual and disciplinary competencies to solve lesson problems. Low-quality DE reflects fragmented discussion of content with limited elaboration, and a focus on recall, which may reflect initial understand early in the task or unit or may reflect task or instructional constraints. Prior research suggests that high-quality DE leads to growth in disciplinary achievement (Hmelo-Silver et al, 2015).

Table 1

Productive and Low-quality Indicators Collaborative Group Disciplinary Engagement

	Indicators of Productive Engagement	**Indicators of Low-quality Engagement**
Behavioral	Sustained on-task behavior, participation, persistence, and effort, even in the face of challenge	Primarily off-task behavior, disengagement, and limited focus on shared task work
Socioemotional	Respectful, inclusive, cohesive, with a climate characterized by psychological safety	Disrespectful interactions, exclusion, lacking cohesion and underlying tension reflecting strain to the climate
Collaborative	Coordinated and responsive interactions, with balanced participation and diverse perspectives solicited when building and engaging in knowledge co-construction	Lack of coordination given separate and unrelated contributions, without attempts or willingness to link (i.e., imbalance due to power differential)
Metacognitive	Planning, monitoring and evaluation focused on content and/or discipline, and meeting task expectations, aimed toward understanding, improvement, progress, integration, consensus, revisions, and task quality	Regulation focuses on basic task completion or is ineffective (e.g., unable to cohere around a plan; regulation is not pursued, accepted or ignored; heavy focus on regulating behavior), obstructing task progress
Disciplinary	Conceptual and disciplinary connections from prior lessons, across domains, or everyday experiences, with extended elaboration and rationale	Fragmented and surface-processing of content and practice, with no elaboration or attempts to connect, such as when memorizing or recalling facts or eliciting prior relevant terms when brainstorming

1.1 Beyond Current Methods for Studying Engagement

We operationalize this multidimensional conceptualization in a rubric, enabling the study of distinct engagement dimensions as well as the interrelationships of engagement dimensions that together describe groups' productive (or unproductive) progress. Although extant theory has conceptualized an

individual's engagement as multidimensional (Fredricks, et al., 2004), much existing *observational* research has assessed group engagement narrowly as a single dimension, such as on-task behaviors (Hmelo, et al., 1998; Lipponen, et al., 2013) or disciplinary engagement (Gresalfi & Barnes, 2016; Koretsky, et al., 2021; Mortimer & Oliveira de Araujo, 2014; Sengupta-Irving & Agarwal, 2017). There has been some qualitative research toward investigating interrelations among two dimensions, including socioemotional and metacognitive engagement, socioemotional and cognitive engagement, and metacognitive and cognitive engagement (Isohätälä, et al., 2020; Khosa & Volet, 2014; Rogat & Adams-Wiggins, 2015; Rogat & Linnenbrink-Garcia, 2011). Thus, we have limited understanding of the interplay among multiple dimensions. In our own prior research, we posited a multidimensional conceptualization of engagement with some initial exploration of the role of a threshold of engagement practices for on-task behavior and respectful climate to further collaborative engagement toward understanding for differentiating high and low case illustrations (Sinha et al, 2015).

The observational research that draws on a multidimensional approach tends to focus on single cases in specific curricular and disciplinary contexts (Engle & Conant, 2002; Järvenoja, et al., 2018; Sinha et al., 2015). Moreover, these approaches rely on intensive and often line-by-line analysis of few cases and thus are not sufficient, alone, for evaluating broader claims concerning how group engagement yields productivity or how engagement fluctuates over time. Addressing broader questions would require the use of moderate-to-large samples of many groups and multiple observations per group. We seek to enable analyses of these larger samples by proposing a method that can be applied directly to video or during real-time observations, more efficiently than prior approaches, but with similar ability to richly document the nature and quality of group engagement.

Empirical study of individual student engagement has assessed engagement as stable, an artifact of self-report surveys capturing of particular moments or retrospective accounts of extended time periods (Perry & Winne, 2006). Further, self-report surveys limit access to information about the nature and quality of interactional processes within the context of particular activities or practices (Ryu & Lombardi, 2015; Vriesema & McCaslin 2020). In alignment with situative views, we prioritize theorizing engagement as involving dynamic change, as groups come to negotiate engagement practices in particular unit phases, on specific tasks and instructional circumstances. Capturing engagement as dynamic enables specification of fluctuations and change in group engagement over time, such as during a task within a lesson. Moreover, we seek to understand how groups reach productive levels of DE, as groups make intellectual progress in their understanding of disciplinary content and/or practices.

Further, our conceptualization of disciplinary engagement concerns groups' collective reasoning with domain content knowledge and/or practices authentic to the discipline within specific problem solving, design, or modeling tasks. Opportunities for students to engage disciplinarily is important within collaborative exchanges, which offer opportunities for peer-to-peer knowledge co-development of concepts, explanation, and critical examination of arguments presented by other group members (Roschelle & Teasley, 1995). This conceptualization is an extension of prior notions of cognitive engagement as context and discipline-general (Fredricks, et al., 2004; Pintrich & DeGroot, 1990) to *disciplinary engagement* common to STEM fields, but instantiated differently within mathematics, science and engineering. Toward these ends, we focus on groups' conceptual work as integrated with discipline-specific reasoning (e.g., modeling, argumentation, design) to generate knowledge needed to solve problems.

Taken together, we draw on and extend prior theory and methods in several critical ways including 1) developing a multidimensional conceptualization of group engagement, 2) developing a rubric to apply that conceptualization in less time and with less labor intensive analyses, 3) focusing on dynamics of group engagement toward groups' productive disciplinary engagement, and 4) by comparing patterns of group engagement across disciplinary contexts to explore relationships between patterns of engagement and the contexts in which they emerge.

Extending these theoretical and methodological precursors, our research team developed and piloted a rubric for describing group GDE in three STEM curricular contexts. We examine the interrelations among the five engagement dimensions and how, together, these constitute GDE. We also illustrate the synergy and mutual influence among dimensions using a case example and a profile

analysis with a larger sample drawn from our video corpus. Finally, we illustrate how we access discipline-specific patterns of engagement via our multidimensional rubric, drawing on a comparison case from a second curricular context. Here, we aim to interrogate, through converging sources of evidence, whether the theoretical framework that we developed and rubric ratings can approximate access to GDE gained from intensive qualitative analysis. Toward these aims we pose the following research questions:

1. How can a multidimensional conceptualization of engagement account for collaborative group disciplinary activity in the context of particular science, mathematics, and engineering practices?

2. What levels of engagement quality together promote groups' productive disciplinary engagement?

2. Method

Group disciplinary engagement (GDE) is contextualized in collaborative tasks involving modeling, design, and argumentation in middle school math, science, and engineering. We draw on a rich corpus of video data collected from three projects with common features including STEM content, disciplinary modeling and argumentation practices, and group work as central to unit goals and what groups came to understand. That is, groups worked together for the majority of lessons in each unit, if not daily, and teachers expected that group work should be the primary mode of activity. Within this corpus, the variation in STEM domain (science, math and engineering) and the disciplinary practices and curricular features (e.g., technology tools, scaffolds) as contextualized in each curriculum has enriched our theoretical development efforts (Koretsky, et al., 2019). The three projects from which collected video was available for rubric development are summarized below.

The Promoting Reasoning and Conceptual Change in Science (PRACCIS) project developed three inquiry units aimed to involve students in scientific practices of evaluating evidence and model fit based on evidence (Chinn, et al., 2018). Collaborative groups develop, evaluate, and revise explanatory models. For this study, video was selected from the third unit, which focused on evolution and natural selection. Video data stemmed from 9 groups in 4 teachers' classrooms. The school district's demographics include a student body that is 49% White, 5% Black, 34% Asian, and 9% Hispanic/Latino. 5% of students are English language learners. Moreover, 14% of students are eligible for free and reduced priced lunch.

The SimCalc Engagement Project (SEP) leveraged student's use of multiple dynamic representations of mathematical relationships (i.e., modeling) to support their understanding of rate and proportion in a technology-based instructional unit (Roschelle et al., 2010). Study videos were drawn from 3 focal groups in 2 teachers' classrooms. One teacher taught in a school where the student body is 56% Hispanic/Latino, 12 % Asian, 10% White, 9% Filipino, 4% Black, and 1% Native American. In this school, 53% of students are eligible for free and reduced priced lunch and 22% are classified as English language learners. The second teacher taught in a school where the student body is 63% White, 14% Asian, 13 % multi-racial, 9% Hispanic and 1% Filipino. In this school, 1% of students are eligible for free and reduced priced lunch and 1% are classified as English language learners.

The Human-centered Robotics project aimed to inspire youth interest in STEM topics to develop robotic technologies in response to people's needs (Gomoll et al., 2018). Video data were used for two groups during two curriculum implementations of a STEM elective course in one classroom. Study videos were drawn from 2 focal groups in one teacher's classrooms in a rural midwestern US school district that was largely White (~90%), with 39% eligible for free and reduced lunch.

The sample consisted of 36 five-minute segments drawn from the larger video corpus of 77 lessons. This included 15 groups, each observed once or twice. In these videos, students were primarily

assigned to work as groups of three or four, but occasionally were assigned to work in pairs and then self-organized into groups. Video segments were balanced across the three units. We aimed to observe some variation in the sample by including videos from across unit phase and disciplinary practices integrated in tasks. Moreover, given our interest in exploring the dynamics of group engagement, we intentionally included some additional segments to observe groups throughout a lesson (i.e., consecutive time segments or segments from later in a lesson).

2.1 Measuring Collaborative Disciplinary Engagement

We developed a rubric to employ when observing groups collaborating during joint activity with the aim of characterizing their co-negotiated engagement practices. The rubric encompasses five engagement dimensions using primarily 3-point rating scales, with DE specified using a 4-point scale. We recognized that the DE dimension benefited from an extended scale based on our observations during rubric development, both in the case of limited discourse and non-verbal interactions such as during independent activity in which productive disciplinary engagement could still be happening, and in the case of the high-end of the continuum where elaboration and justification as well as initial and brief conceptual connections are made. Beyond the engagement dimensions, the rubric also captures group structure (i.e., whether groups opt to or are assigned to work in pairs, as a full group, and even individually) to characterize fluctuations in how groups are organized over time. Initial rubric drafts were informed by a review of extant research on engagement, empirical studies, and self-report items. These initial drafts were also informed by conducting joint analyses of group interactions (N = 4 groups), drawing on the range of expertise of the various project team members when describing observable group engagement along multiple dimensions from across curriculum contexts (Jordan & Henderson, 1995). Iterative revision of this initial framework was informed by a pilot study, expert feedback, and questions during rater training. Raters observe group interactions including members' behaviors, discourse content, and nonverbal behaviors (e.g., gaze, gesture, leaning in toward joint task, spatial closeness) (Chi & Wylie, 2014). Using this rubric, we assign group engagement ratings based the predominant character of group interactions for the majority of the selected segment.

Quality indicators differentiated *behavioral engagement* as primarily on-task or off-task joint participation, *socioemotional engagement* as the respectful, inclusive and cohesive group climate and alleviation of tension and frustration stemming from different perspectives or interpersonal dynamics, *collaborative engagement* as responsive and coordinated knowledge building, *metacognitive engagement* as regulation focused on the progress and understanding of content and disciplinary practice, and *disciplinary engagement* as the integrated conceptual and disciplinary contributions with rationale. Across dimensions, we assume that high-level ratings facilitate the likelihood of attaining productive GDE, with some potential exceptions of low or moderate ratings also supporting productive GDE (e.g., temporary off-task joking (low BE) may benefit group cohesion (high SE)). Rubrics for each engagement dimension can be found in the Appendix.

2.2 Study procedures

In applying the rubric, raters assigned quality ratings for each of the five engagement dimensions, for five-minute time segments of videotaped group work. This time segment choice was based in our own prior research and prior observational studies of engagement (Lee & Brophy, 1996; Sinha, et al., 2015). The time segment also afforded more than a single turn in conversation, with sufficient time for group members to respond and negotiate a task direction and/or understanding. Raters independently viewed video segments twice, first to familiarize themselves with the activity and a second time to assign ratings. Ratings were applied by three teams, made up of 2 to 3 raters, with each team specializing in one curricular context. The study proceeded in four rating cycles, followed by calculating interrater agreement. In cycles 3 and 4, rater teams had Krippendorff's alpha interrater agreement coefficients of .53 for BE, .64 for CE, .59 for SE, .30 for ME, and .48 for DE. LeBreton and Senter (2008, p. 836) suggest that for chance-corrected interrater agreement coefficients, values of 0 to .30 should be interpreted as "lack of agreement," .31 to .50 as "weak agreement," and .51 to .70 as "moderate agreement." After recording individual ratings and calculating initial interrater agreement,

the pair worked together with the project team's master raters to resolve areas of disagreement and come to consensus; these conversations were structured to clarify discrepancies and produce coding clarifications and exemplars to inform subsequent coding. It is these consensus ratings that we employ in the presented analyses.

2.3 Analysis

To gain an understanding of relationships between DE and each of the other dimensions, we calculated correlations and conducted a profile analysis. Given that the five dimensions are theorized to jointly constitute productive group engagement, we would anticipate moderate positive correlations among the dimensions. Profile analysis is a variant of multivariate analysis of variance (MANOVA) that allows testing of hypotheses about patterns of means across groups. Specifically, if omnibus *F* tests indicated significant differences in the pattern of engagement dimension means across the high and low DE subsamples, we would proceed to univariate analyses of dimension mean differences. We conceptualized the four DE rating scale levels as possibly representing two categorically different types of DE, relatively low (scale categories 1 and 2) and relatively high (scale categories 3 and 4), rather than as linearly increasing. Thus, we identified two subsamples based on the observed level of DE and compared profiles of means on the other engagement dimensions across those two subsamples. Because fewer than 40% of the student groups were observed more than once over time, we planned to interpret the results only as indicating any cross-sectional relationships that exist among the engagement dimensions.

Five student group observation cases, a small fraction of the sample, were missing ratings on metacognitive engagement because none was observed during the sampled time period. Rather than excluding these cases from the analysis, which would be expected to cause estimation bias unless the missingness was completely at random (e.g., Graham, 2009), we implemented multiple imputation for 10 replications. Stata was used to compute estimated coefficients and standard errors for each replicated dataset, and to combine the results using Rubin's (1987) rules.

A case illustration was purposefully selected from those groups (n = 5) showcasing high-level DE of a 4 rating, with assigned dimensional ratings similar to remaining groups with high DE ratings. We intentionally drew from two curricular contexts that would require observers to specifically evaluate the disciplinarity of DE. We reviewed the video segments and described with rich narrative the central dimensions specified in the rubric. Analyses focused on how engagement dimensions worked in synergy to produce high-level disciplinary engagement.

3. Results

Means, standard deviations and correlations among the engagement dimension ratings are presented in Table 2. SE had the highest mean rating, while CE had the lowest mean ratings. The moderate positive correlations, most of which are significantly greater than zero, suggest positive correspondence among all five dimensions. The high correlation between ME and DE occurs in all likelihood because the quality of ME is important for initiating and/or supporting DE. The low correlation between BE and SE likely indicates that positive socioemotional interactions (SE 3 rating) may be evident during off-task exchanges (BE 1 or 2 ratings).

Table 2

Correlations among Engagement Dimension Ratings (N = 36)

	Mean	*SD*	*BE*	*CE*	*SE*	*ME*	*DE*
Behavioral Engagement	2.44	0.73	1				
Collaborative Engagement	2.14	0.64	.33	1			
Socioemotional Engagement	2.69	0.62	.18	.27	1		

Metacognitive Engagement	2.29	0.74	.74***	.57**	.58*	1	
Disciplinary Engagement	2.61†	0.77	.60**	.58***	.29	.85***	1

Note. † Mean = 2.07, SD = .051 for Disciplinary Engagement ratings transformed to 3-point scale similar to that employed for the remaining dimensions. * $p < .05$; ** $p < .01$; *** $p < .001$.

3.1 Case illustration

We conducted a qualitative case analysis to richly characterize a group's efforts in making intellectual progress and to explore the potential for synergistic interrelations among engagement dimensions in explaining high-quality DE. This specific case stems from the SimCalc mathematics curricular context. The triad is made up of three girls (pseudonyms: Abby, Beth, Carly) (see Figure 1a). As shown in Figure 1b, the group is tasked with interpreting a line graph depicting motion of two fictional vehicles over time (i.e., their speed). The assigned ratings indicate some intermittent off-task activity (BE=2), with remaining ratings at a high-quality level (SE=3, CE=3, ME=3, DE=4). They were working on recording answers in their individual workbooks, with two open laptops. Abby was recently absent and was working to catch up in her workbook. The lesson took place in November and was mid-way through a 2-week unit during which this group had been working together on a daily basis. Students' comments indicated that they already knew each other before the unit.

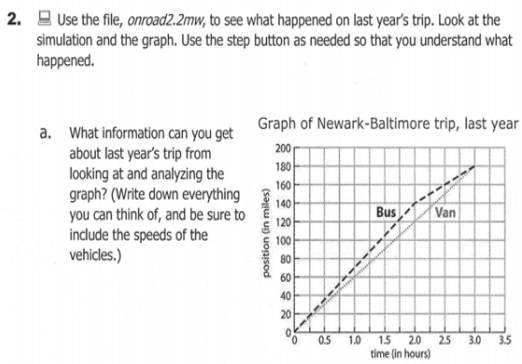
Just prior to the focal time segment, Abby raised a question, based on her metacognitive monitoring of the assigned question concerning how to read the graph in question 2a (ME)[1] (Figure 1b). Beth and Abby's beginning interactions in making sense of the graph illustrate both girls voicing a common misunderstanding that the line graph represents the physical characteristics of one of the vehicles' routes rather than its speed and relying on the simple recall of the slope formula. Abby's questions about the meaning of the slope formula (ME) as related to the graph, elicited group disciplinary engagement (DE) as they mutually monitored their understanding and grappled with interpreting the graph:

> Abby: I don't get how it [the graph] can describe the motion. Like, what does that mean? They're just lines. It doesn't say describe the graph, it says to describe the bus and the van.
> Beth: The motion. But it is still like in the graph
> Abby: Yeah, but the bus didn't drive tilted.
> Beth: The motion of the bus. The van had a straighter route, and the bus had a more curved route.
> Abby: I guess that made sense. Then we have to write the speed again.
> Beth: And it would be like distance over….distance over time.
> Abby: For one second, does that mean? Like what distance and what time?

The full group briefly disengaged from the task as Abby complimented Carly's hat (BE). This elicited Carly's sharing a peer's previous teasing about it. Beth and Abby showed their support by giggling along, actively listening, and agreeing that it was silly to dislike her unique hat (i.e., "Some people just don't know your style.") (SE). This temporary off-task exchange (BE) was positive in socioemotional interactions and fostered team cohesion (SE) amongst the group, which the group subsequently leveraged in the collaborative engagement (CE) which followed.

[1] Throughout the case we identify evidence of specific engagement dimensions by noting their abbreviations in parenthesis.

(a)



(b)

Figure 1. Case illustration collaborative group (a) and image of student worksheet (b).

Abby returned to the task (BE) and persisted in raising a question regarding her uncertainty in interpreting the line graph (ME). "Beth, so I don't get how they can just be…how can they go the same amount of miles and still go at the same speed? Like, it just doesn't make sense." Beth was responsive to Abby's question, turning toward her and leaning in to view the graph Abby was drawing (SE), precipitating this collaborative and disciplinary engagement. Beth leaned in and explained, "Because this one [the bus] is only a little more quicker." Abby further elaborated and built from this explanation (CE) stating, "Oh I get it because the bus starts going off faster…So it ...is like it has more slack, like kind of has more time and then they arrive at the same time." and "Because the bus stopped. So they are both going the same speed." Abby continued to negotiate their working understanding of the different speeds of the van and bus while gesturing to her graph (DE), "Actually, the bus still is going at a faster speed. Its speed would be faster than the van. Just because it stopped doesn't mean the speed changed, right?" (ME and DE).

Abby further questioned whether their previously employed effective math strategy for using the arrival or end time of the graph to calculate the slope would work for this new representation with vehicles at different speeds. "But, then so I am confused on how I would represent that as two things, fractions, because at the end they are the same" and, "I'm not going to do the end distance and the end time; I am going to do something a little earlier on" (ME and DE). Throughout, Beth and Carly showed their coordination by voicing agreement, responding "Yeah," ensuring they were looking at the same question in the packet and by prompting Abby's meaning making (i.e., Carly "So (inaudible) what do you think of the motion?") (CE). Carly further supported the group's knowledge construction as she moved her physical position to be proximal to Abby and facilitated the interpretation of the graph and the task, further suggesting that the former off-task but positive socioemotional interaction (SE) maintained her involvement in the conversation.

This case illustration showcases how the five-engagement dimensions interrelated across a five-minute time segment in ways that afforded intellectual progress at the integration of mathematical practice (generating a model, coordinating the use of a graphic representation) and content (conceptual understanding of slope) (DE). The group engagement practices reflect a willingness to persist in the face of uncertainty (BE) and metacognitively monitor for understanding concerning how to interpret the graphic representation of speed, the meaning behind the slope formula and different vehicle speeds, as well as whether their former mathematical strategy of using graphic endpoints was still relevant (ME), initiated by Abby but sustained by the joint group's high level of engagement. Carly and Beth's high-quality collaborative engagement was responsive to the metacognitive monitoring and yielded joint knowledge co-construction (CE). The group was mutually respectful and cohesive throughout the exchange, including when off-task behavior was similarly leveraged (SE). Ultimately, it is our multidimensional conceptualization of the group's engagement that enabled the examination of synergy among dimensions as facilitating the *joint accomplishment* of DE.

3.2 Exploring the disciplinarity of engagement

Common curriculum features, including incorporation of the authentic disciplinary practices of modeling, argumentation, and design within the curriculum corpus, facilitated questions about how DE could be evaluated in ways that were common in three domains, while remaining sensitive to detecting discipline-specificity. During rubric development, collaborative analysis of videos from across curriculum contexts enabled the team to generate descriptions of DE that are applicable across disciplinary tasks, involving various disciplinary practices (see Appendix). Indicators that are specific to each context ground the DE rubric in examples that support raters; this is especially important for specifying disciplinary variation. This approach, a broader definition of DE with context-specific indicators, allows for comparative analyses across contexts, including patterns of and interrelations between various dimensions of the rubric. For example, observing in one curricular context where frequent student interaction was aligned with disciplinary norms, we noted long periods of independent work eroded team cohesion (SE) and ultimately constrained conceptual progress. However, in one segment from the Robotics dataset, students spent far more time in independent on-task work (BE=3), captured in group structure ratings of 'individual work,' with mid or low ratings on all but socioemotional engagement (SE=3, CE=1, ME=0, DE=2), as students each attended to building different components of the robot. In this context and similar disciplinary spaces, longer periods of independent work were common and group cohesion was maintained, marked by brief check-ins by group members (e.g., 'does this look right to you?') and intermittent gesture-based collaboration accomplished through physical indicators (e.g., leaning over to examine and/or gazing at each other's progress without comment). In later segments, the teams' disciplinary engagement was rated higher when their conversation turned to providing rationales for the ways their robot construction choices did and did not meet stakeholder stated preferences. We intended our final rubric to measure DE in a unified manner across contexts.

3.3 Profile analysis

To examine patterns in the relationships among disciplinary engagement and the other dimensions across the cross-sectional video sample, a profile analysis was conducted. Taking group disciplinary engagement ratings of 3 or 4 to indicate high (n=20 cases), and ratings of 1 or 2 to indicate low DE (n=16 cases) (see Section 2.2 and Appendix), we prepared a plot of the mean rating profiles (Figure 2). Visual inspection indicated the high DE group observations tended to have higher ratings across all four co-occurring engagement dimensions.

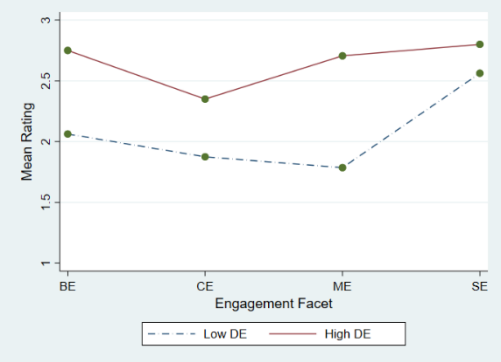


Figure 2. Profiles of (mean) group ratings on other engagement dimensions when Disciplinary Engagement (DE) is high and low

A preliminary multivariate analysis of variance (MANOVA) was used to test the hypothesis of no mean differences between the low and high disciplinary engagement observations on any of the other engagement ratings (in which case proceeding with profile analysis would be unwarranted). The null hypothesis was rejected for the Wilks' lambda omnibus test statistic [$F(4, 31) = 5.85$, $p = 0.001$], so we

continued with the profile analysis. A test of parallelism indicated that the profiles of mean ratings for the low and high DE observations had marginally significant differences in overall shape [$F(3, 32) = 2.82$, $p = 0.063$]. A test of "level" or group differences confirmed that the profiles were not coincident; that is, the group ratings on each dimension were not identical [$F(1, 34) = 14.81$, $p < 0.001$]. Given the significant group difference in overall rating outcomes found in the MANOVA, to identify specific engagement dimension(s) that were the source, we estimated univariate ANOVA models with each engagement dimension as a single outcome, and the high/low DE indicator again as a predictor. These "stepdown" analyses suggested significant differences between the high and low DE observations for behavioral [$t = 3.12$, $p < 0.01$], collaborative [$t = 2.35$, $p < 0.05$], and metacognitive [$t = 4.36$, $p < 0.001$] engagement, but not socioemotional engagement. The profiles suggest that the other engagement dimensions significantly vary with the quality of DE and that these dimensions may interrelate in fostering DE.

4. Discussion

In this research, we used multiple methods to investigate how group productive disciplinary engagement is constituted, using a multidimensional framework and our initial rubrics developed to embody this framework. Study findings suggest that the five dimensions of our disciplinary engagement framework and rubrics are positively interrelated, illustrating that interrelationships among dimensions mutually support the high-quality of disciplinary engagement observed among groups during joint activity. We used a case illustration to richly characterize the dynamic and synergistic nature of these dimensions for explaining DE, as well as the import of being sensitive to disciplinary specificity of DE. Ratings assigned to the time segments of group activity provide data complementing the nuance of in-depth cases, by allowing for the identification of common patterns across a wide range of groups in varying curricular contexts, as initially demonstrated by the profile analysis. The rubric corresponded to group engagement features that could have been identified via intensive video analysis, while affording the evaluation of broader claims of patterns of group engagement with larger datasets.

Although the profiles of high and low DE group observations suggest that the other engagement dimensions significantly vary with the quality of DE, this was not the case for socioemotional engagement (SE); results suggested that the high and low DE observations were both characterized by high-quality socioemotional interactions. One explanation is that the high SE (i.e., 3) rating included a range of quality indicators, primarily reflecting polite and collegial interpersonal interactions, climate norms aligned with working well together, expectations as part of classroom work but also interactions that added to or made efforts to maintain a positive climate (Rogat & Linnenbrink-Garcia, 2011; Summers, et al., 2005). The highest form of SE was observed during off-task group exchanges, characterized by positive and friendly interactions that could be carried over back to on-task interactions (Langer-Osuna, et al., 2020). It may be that the high-quality SE which differentiates high DE group engagement reflects the active negotiation and mutual accountability of a climate facilitative of risk taking and inclusion of diverse ideas, at the upper endpoints for this dimension. These interpersonal dynamics are likely a less regular occurrence, contextualized in particular situations (e.g., newly constituted group, following disagreement or provoked tension), requiring examination of longer time periods and the exploration of dynamics via analyses that were not yet conducted in this initial development work.

Our work aligns with a situative perspective on learning by investigating collaborative engagement as shared, with the group as the unit of analysis (e.g., Barron, 2000; Engle & Conant, 2002; Gresalfi & Barnes, 2016). This work contributes a method to the study of group engagement that leverages the benefits of observational methods (vs. self-report) and multi-dimensional conceptualizations of engagement in a framework and rubric. These tools enable descriptions and analyses of group engagement as situated and anchored in disciplinary content and practices, and as trajectories comprised of dynamically interrelated aspects of group activity. One important implication for research is that using this tool, researchers can address critical questions about collaborative learning

and engagement that have been inaccessible because of limitations of prior methods, including examination of the complex dynamics of separable but interdependent dimensions of engagement and, in particular, the potential variation in quality within and across dimensions. This work extends collaborative group research that has examined single engagement dimensions, which may lend to a conceptualization of these dimensions as separable and independent rather than as interrelated dimensions that present a more enriched characterization. Moreover, these phenomena can be examined across disciplinary learning contexts, across specific tasks, and as a function of time, while remaining sensitive to disciplinary specificity when investigating productive (and non-productive) engagement. Our rubric, by being broadly applicable but grounded by discipline-specific indicators, supports theory development about groups' GDE, including implications for instructional design, practice, and collaborative learning. In other research, Gomoll et al (2020) used these rubrics as a tool for professional development that the teacher could use as a lens for video analysis and subsequent facilitation of collaborative groups. Although this was a small-scale study, future research can build on these implications for teacher professional development.

Limitations of the current study include the examination of interrelations of dimensions of engagement in single time segments, precluding the examination of engagement as dynamic and fluctuations *across* multiple segments that make up a group task, which is the ultimate goal of this research program and is part of our current research activities. Moreover, the video sample included repeat observations from the same day's lesson, given our long-term interest in exploring lesson dynamics and short-term interest in testing alternative time intervals for recording ratings, but those were not modeled in the profile analysis given the modest sample size. Additionally, this study examines a case with high-quality engagement ratings for a majority of the rubric dimensions, but we do not assume that this pattern exclusively fosters GDE (e.g., sometime off-task behavior [lower BE] enables students to joke and bond socially [increasing SE], which may precede GDE); future research will aim to identify and illustrate other such patterns.

The rubric presented here is an initial exploration for operationalizing collaborative group disciplinary engagement as five dimensions using quality ratings. However, we faced challenges in obtaining inter-rater agreement prior to consensus meetings, with reliability indices indicating room for improvement particularly for metacognitive engagement (ME). We understand these challenges as attributable to (1) rater process, (2) the complexities of group data, and (3) unique challenges presented by the ME dimension. First, these data were from secondary sources collected from past projects and although raters were trained to become familiar with the curriculum materials, we anticipate that most users of this rubric would be studying curricular contexts with which they are highly familiar. Second, we asked raters to examine interactions among members of the group, which is more complex than observing individuals' engagement (e.g., Lee & Brophy, 1996), capturing the nature of collective interaction as students present a contribution and other groupmates' responses by accepting, ignoring or rejecting with or without rationale (Barron, 2000). Furthermore, raters observe engagement across 5-minutes, with fluctuation in quality typical within that timeframe, and being tasked with selecting the rating that best represents the majority of that time. In our current research with a revised rubric, we have modified the time segment to 2.5 minutes to address the noted disagreement provoked by raters' varying strategies for synthesizing across data. Finally, ME proved uniquely challenging as raters were asked to simultaneously evaluate several relevant elements in their rating, including the metacognitive target, the duration, and its quality. In other research employing qualitative analyses, these elements are considered in distinct analytic phases, with an initial step identifying the presence of ME in group interaction, and then classifying its target (e.g., Rogat & Linnenbrink-Garcia, 2011; Volet, Summers & Thurman, 2009). Building on this exploratory study, we have worked to address challenges specific to ME by removing the evaluation of duration and developing more specific indicators for each level of quality. Nonetheless, because of the initial low reliability, all ratings used here were consensus ratings of two or more raters.

Future research might also employ mixed methods for coupling the analysis of dynamic group patterns of engagement, such as through latent profile transition analysis on the five engagement dimensions across time (e.g., Nylund-Gibson, et al., 2014). This analysis would include an in-depth qualitative analysis of groups exemplifying these engagement trajectories. This convergence of methods

would offer an enriched understanding of how specific group processes unfold as trajectories. Another recommendation would be to investigate how the framework and rubric may prove valuable in classroom contexts by supporting teachers through professional development about the additional support and resources collaborative groups need, targeted toward specific engagement dimensions. Similarly, studies could examine the benefit of proximal feedback provided to teachers to identify groups at varying unit phases in need of monitoring and/or support relevant to specific group processes hindering their collective efforts at making intellectual progress. Although there is still work to done, this research demonstrates the importance of considering group disciplinary engagement as a complex and multidimensional phenomenon.

Keypoints

- We propose that engagement is a group-level construct that builds from and integrates theory on individual engagement and productive disciplinary engagement.
- We advance innovative methods for evaluating collaborative group engagement during disciplinary activity as shared, multidimensional and dynamic.
- We developed and piloted an observational rubric encompassing five engagement dimensions, each with quality ratings.
- Mixed-methods analyses included correlations, profile analyses, and case illustrations.
- Results illustrate the synergistic interrelations among engagement dimensions that together constitute group disciplinary engagement.

Acknowledgments

This research was funded by the United States' National Science Foundation (NSF) Grants # 1661266 and 1661234. 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. We thank Clark Chinn and Ravit Duncan for sharing the PRACCIS video, theSimCalc Research group at SRI, especially Nikki Shechtman and Patrik Lundh for SimCalc data and thought partnership. We thank the graduate and undergraduate researchers who were part of our project term who employed the described rubric and provided invaluable feedback, specifically Brianna Cermak, Alex Lindsay, Mehdi Ghahremani, and Skye Huffman. We also thank Megan Humberg for helpful comments on an earlier draft.

References

Adams-Wiggins, K. R. (2020). Whose meanings belong?: Marginality and the role of microexclusions in middle school inquiry science. *Learning, Culture and Social Interaction*, *24*, 100353. https://doi.org/10.1016/j.lcsi.2019.100353

Barron, B. (2000). Achieving coordination in collaborative problem-solving groups. *Journal of the Learning Sciences, 9*, 403-436. https://doi.org/10.1207/S15327809JLS0904_2

Chi, M.T.H. & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49,* 219-243. https://doi.org/10.1080/00461520.2014.965823

Chinn, C. A., Duncan, R. G., & Rinehart, R. (2018). Epistemic Design: Design to Promote Transferable Epistemic Growth in the PRACCIS Project. In E. Manalo, Y. Uesaka, & C. A. Chinn (Eds.), *Promoting Spontaneous Use of Learning and Reasoning Strategies*. pp. 242-259. New York: Routledge.

Danish, J. A., & Gresalfi, M. (2018). Cognitive and Sociocultural Perspective on Learning: Tensions and Synergy in the Cognitive and Sociocultural Perspective on Learning: Tensions and Synergy in the Learning Sciences. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (pp. 33-43). New York: Routledge.

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

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

Engle, R. A., Langer-Osuna, J. M., & McKinney de Royston, M. (2014). Toward a model of influence in persuasive discussions: Negotiating quality, authority, privilege, and access within a student-led argument. *Journal of the Learning Sciences*, *23*(2), 245-268. https://doi.org/10.1080/10508406.2014.883979

Forsthuber, B., Motiejunaite, A., & de Almeida Coutinho, A. S. (2011). *Science Education in Europe: National Policies, Practices and Research*. Education, Audiovisual and Culture Executive Agency, European Commission.

Fredricks, J., Blumenfeld, P., & Paris, P. (2004). A school engagement potential of the concept and state of the evidence. *Review of Educational Research*, *74*, 59-109. https://doi.org/10.3102/00346543074001059

Forman, E. A., & Ford, M. J. (2014). Authority and accountability in light of disciplinary practices in science. *International Journal of Educational Research*, *64*, 199-210. https://doi.org/10.1016/j.ijer.2013.07.009

Fulmer, S.M., & Frijters, J.C. (2009). A Review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review, 21,* 219-246. https://doi.org/10.1007/s10648-009-9107-x

Gomoll, A. S., Hillenburg, R., & Hmelo-Silver, C. E. (2020). "I have never had a pbl like this before": On viewing, reviewing, and co-design. *Interdisciplinary Journal of Problem-based Learning. 14*(1). https://doi.org/10.14434/ijpbl.v14i1.28802.

Gomoll, A. S., Šabanović, S., Tolar, E., Hmelo-Silver, C. E., Francisco, M., Lawlor, O. (2018). Between the Social and the Technical: Negotiation of Human-Centered Robotics Design in a Middle School Classroom. *International Journal of Social Robotics, 10*, 309-324. https://doi.org/10.1007/s12369-017-0454-3

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. 10.1146/annurev.psych.58.110405.085530

Greeno, J. G. (2006). Learning in activity. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 79–96). Cambridge: Cambridge University Press

Gresalfi, M. S., & Barnes, J. (2016). Designing feedback in an immersive videogame: supporting student mathematical engagement. *Educational Technology Research and Development*, *64*(1), 65-86. https://doi.org/10.1007/s11423-015-9411-8

Gresalfi, M., Martin,T., Hand, V. & Greeno, J. (2009). Constructing competence: an analysis of student participation in the activity systems of mathematics classrooms. *Educational Studies in Mathematics*, *70,* 49-70. https://doi.org/10.1007/s10649-008-9141-5

Hand, V., & Gresalfi, M. (2015). The joint accomplishment of identity. *Educational Psychologist, 50,* 190-203. http://dx.doi.org/10.1080/00461520.2015.1075401

Hickey, D. T. (2003). Engaged participation versus marginal nonparticipation: A stridently sociocultural approach to achievement motivation. *Elementary School Journal*, *103*(4), 401-429. https://doi.org/10.1086/499733

Hmelo, C. E., Guzdial, M., & Turns, J. (1998). Computer-support for collaborative learning: Learning to support student engagement. *Journal of Interactive Learning Research, 9,* 107-130.

Hmelo-Silver, C. E., Eberbach, C., Jordan, J., Sinha, S., Rogat, T. K. (2015, August). *Trajectories for Engaged Learning about Complex Systems.* Presented at European Association for Research on Learning and Instruction Biennial Conference. Limassol, Cyprus.

Isohätälä, J., Naykki & Jarvela (2020). Convergences of Joint, Positive Interactions and Regulation in Collaborative Learning. *Small Group Research, 51,* 229-264. https://doi.org/10.1177/1046496419867760

Järvelä, S., Järvenoja, H., Malmberg, J., Isohätälä, J., & Sobocinski, M. (2016). How do types of interaction and phases of self-regulated learning set a stage for collaborative engagement?. *Learning and Instruction*, *43*, 39-51. http://dx.doi.org/10.1016/j.learninstruc.2016.01.005

Järvenoja, H., Järvelä, S., Törmänen, T., Näykki, P., Malmberg, J., Kurki, K., Mykkänen, A. & Isohätälä, J. (2018). Capturing motivation and emotion regulation during a learning process. *Frontline Learning Research*, *6*, 85-104. https://doi.org/10.14786/flr.v6i3.369

Khosa, D. K., & Volet, S. E. (2014). Productive group engagement in cognitive activity and metacognitive regulation during collaborative learning: can it explain differences in students' conceptual understanding?. *Metacognition and Learning*, 1-21. https://doi.org/10.1007/s11409-014-9117-z

Koretsky, M. D., Vauras, M., Jones, C., Iiskala, T., & Volet, S. (2021). Productive disciplinary engagement in high-and low-outcome student groups: Observations from three collaborative science learning contexts. *Research in Science Education, 51*, 159-182. https://doi.org/10.1007/s11165-019-9838-8

Kreijins, K., Kirschner, P.A. & Jochems, W. (2002). The sociability of computer-supported collaborative learning environments. *Journal of Education Technology and Society 5,* 8–22. http://www.jstor.org/stable/jeductechsoci.5.1.8

Langer-Osuna, J. M., Gargroetzi, E., Munson, J., & Chavez, R. (2020). Exploring the role of off-task activity on students' collaborative dynamics. *Journal of Educational Psychology*, *112*(3), 514-532. https://doi.org/10.1037/edu0000464

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Lee, O., & Brophy, J. (1996). Motivational patterns observed in sixth-grade science classrooms. *Journal of Research in Science Teaching*, *33*, 303-318. **https://doi.org/10.1002/**

Linnenbrink-Garcia, L., Rogat, T.K. & Koskey, K.L. (2011). Affect and Engagement during Small Group Instruction. *Contemporary Educational Psychology, 36,* 13-24. https://doi.org/10.1016/j.cedpsych.2010.09.001

Lipponen, L., Rahikainen, M., Lallimo, J., & Hakkarainen, K. (2003). Patterns of participation and discourse in elementary students' computer-supported collaborative learning. *Learning and Instruction*, *13*, 487-509. https://doi.org/10.1016/S0959-4752(02)00042-7

Molenaar, I., & Chiu, M. M. (2014). Dissecting sequences of regulation and cognition: statistical discourse analysis of primary school children's collaborative learning. *Metacognition and learning*, 1-24. https://doi.org/10.1007/s11409-013-9105-8

Mortimer, E. & de Araújo, A. (2014). Using productive disciplinary engagement and epistemic practices to evaluate a traditional Brazilian high school chemistry classroom. *International Journal of Educational Research, 64*, 156-169. https://doi.org/10.1016/j.ijer.2013.07.004

Näykki, P. Järvelä, S. Kirschner, P.A. & Järvenoja, H. (2014). Socio-emotional conflict in collaborative learning – A process-oriented case study in a higher education context. *International Journal of Educational Research, 68*, 1-14. https://doi.org/10.1016/j.ijer.2014.07.001

NGSS Lead States. (2013). *Next Generation Science Standards: For States, By States.* Washington, DC: The National Academies Press.

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling*, *21*, 439-454. https://doi.org/10.1080/10705511.2014.915375

Perry, N. E., & Winne, P. H. (2006). Learning from learning kits: gStudy traces of students' self-regulated engagements with computerized content. *Educational Psychology Review*, *18*(3), 211-228. https://doi.org/10.1007/s10648-006-9014-3

Pintrich, P.R. & DeGroot, E.V. (1990).Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*, 33-40. https://doi.org/10.1037/0022-0663.82.1.33

Rogat, T.K. & Adams-Wiggins, K.R. (2015). Facilitative versus Directive Other-regulation in Collaborative Groups: Implications for Socioemotional Interactions. *Computers in Human Behavior, 52*, 589-600. https://doi.org/10.1016/j.chb.2015.01.026

Rogat, T.K. & Linnenbrink-Garcia, L. (2011). Socially Shared Regulation in Collaborative Groups: An Analysis of the Interplay between Quality of Social Regulation and Group Processes. *Cognition and Instruction, 29,* 375-415. https://doi.org/10.1080/07370008.2011.607930

Rogat, T.K. & Linnenbrink-Garcia, L. (2013). Understanding the quality variation of socially shared regulation: A focus on methodology. In M. Vauras & S. Volet (Eds.), *Interpersonal regulation of learning and motivation: Methodological advances* (pp. 102-125). London: Routledge.

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S.,... & Gallagher, L. P. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. *American Educational Research Journal, 47*, 833-878. https://doi.org/10.3102/0002831210367426

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69-197). Berlin, Germany: Springer.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Ryu, S., & Lombardi, D. (2015). Coding Classroom Interactions for Collective and Individual Engagement. *Educational Psychologist, 50,* 70-83. https://doi.org/ 10.1080/00461520.2014.1001891

Sandoval, W. A. (2014). Conjecture mapping: an approach to systematic educational design research. Journal of the Learning Sciences, 23, 18-36. https://doi.org/10.1080/10508406.2013.778204

Schoor, C., Narciss, S., & Körndle, H. (2015). Regulation During Cooperative and Collaborative Learning: A Theory-Based Review of Terms and Concepts. *Educational Psychologist*, *50*, 97-119. https://doi.org/10.1080/00461520.2015.1038540

Sengupta-Irving, T., & Agarwal, P. (2017). Conceptualizing perseverance in problem solving as collective enterprise. *Mathematical Thinking and Learning*, *19*(2), 115-138. https://doi.org/10.1080/10986065.2017.1295417

Shechtman, N., Cheng, B. H., Lundh, P., & Trinidad, G. (2012). Unpacking the black box of engagement: Cognitive, behavioral, and affective engagement in learning mathematics. In van Aalst, J., Thompson, K., Jacobson, M. J., & Reimann, P. (Eds.) *The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) – Volume 2, Short Papers, Symposia, and Abstracts.* International Society of the Learning Sciences: Sydney, NSW, Australia, pp. 53-56.

Sinha, S., Rogat, T., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning, 10*, 273-307. https://doi.org/10.1007/s11412-015-9218-y

Summers, J. J., Gorin, J. S., Beretvas, S. N., & Svinicki, M. D. (2005). Evaluating collaborative learning and community. *The Journal of Experimental Education*, *73*(3), 165-188. https://doi.org/10.3200/JEXE.73.3.165-188

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. https://doi.org/10.1016/j.learninstruc.2008.03.001

Vriesema, C.C, & McCaslin, M. (2020). Experience and meaning in small-group contexts: Fusing observational and self-report data to capture self and other dynamics. *Frontline Learning Research*, *8,*136-139. https://doi.org/10.14786/flr.v8i3.493

Appendix

Collaborative Disciplinary Engagement Observational Rubric

Behavioral Engagement: Group norm can be characterized by on-task engagement, persistence, and effort investment, even in the face of challenge		
1: Low	***2: Moderate***	***3: High***
Group characterized by off-task behavior, with limited on-task activity Brief and intermittent on-task activity *Joking in off-task interactions* *Groupmates may attempt to distract on-task activity*	Group characterized by predominantly on-task activity for a majority of the time, but intermittent off-task activity.	Group characterized by sustained on-task activity, with brief intermittent off-task activity
Socioemotional Engagement: Socioemotional climate is respectful, cohesive, and characterized by psychological safety		
1: Low	***2: Moderate***	***3: High***
Group interactions characterized by negative climate reflective of the following qualities: • Disrespectful (put downs, harsh criticism of ideas; grabbing, shoving, pushing you out of the way, shouting) • Interactions showcase low cohesion/sense of team When tension and frustration are expressed, it is responded to with disrespect, resistance to difference in perspectives; tension may be sustained. *When the group makes mistakes, seek blame of groupmates; criticism.*	Group interactions characterized by mixed climate (indicators of both negative and positive climate are present). Tension brings strain to the group climate (although not overtly disrespectful or safe) *Laughter reflects mild tension*	Group interactions characterized by a positive climate reflective of working well together or promoting high-quality positive climate: • Respectful, polite, collegial • Encouraging of groupmates/team • Climate is comfortable in terms of allowing for risk-taking, mistakes as well. • Cohesive/team; warmth and caring about one another • Good-natured and friendly during off-task interactions (e.g., friendly joking) When tension and frustration are expressed, it is alleviated, responded to with safe climate and respect

		When the group makes mistakes – respect and sense of team is fostered.
Collaborative Engagement: Group norm characterized as coordinated and responsive		
1: Low	***2: Moderate***	***3: High***
Group interactions characterized by lack of coordination with the following qualities: • Separate contributions without attempts or an unwillingness to link (i.e., parallel play); contributions may be unrelated. • Imbalance in perspectives due to dominant/power differential • No attempts to revisit a groupmate's previous contribution. *Reject without (conceptual) rationale* *Ignoring (and not returning to idea)* *Unresponsive to questions* *Repetition of one idea, without modifications to incorporate other's ideas* ***Physicality*** Limited eye contact, turning away to another task, spatial distancing Low ratings are assigned when there is no content, practices or assigned task to coordinate around, such as during off-task activity.	Group interactions characterized by intermittent or mixed interactions with the following qualities: A subset of high-quality indicators are present and/or are inconsistent OR Limited coordination because first response is taken-up as group response with limited or no discussion, elaboration, modification or checking for agreement	Group characterized by coordinated interactions, in consistent ways with the following qualities: • Students build from and are responsive to ideas • Diversity in perspectives are solicited and integrated in ways that are balanced among the group • Reject groupmates' ideas with rationale *Elaborating, integrating and /or adding on to one another's contributions* *Responsive to questions, feedback* *When multiple ideas are voiced or solicited, each is considered* *Efforts to build a group response, consensus, and reconcile across contributions, perspectives, or negotiate; taking up one perspective with rationales* ***Physicality*** Coordinated, seamless activity with flow, including nonverbal activity; eye contact; spatial closeness; leaning in, turning toward

Metacognitive Engagement Dimension: Group norm characterized by socially shared regulation and co-regulation focused on content and/or practice, and supported by regulation aimed at maintaining on-task behavior, monitoring of group process, time use, productive emotions, and following task directions.			
No rating	***1: Low***	***2: Moderate***	***3: High***
No observed regulation	Group norm characterized by ineffective regulation (low-quality regulation or regulation is not pursued not taken up/ accepted, or it is ignored), obstructing task progress *Unable to cohere around a common task goal or plan.* *Planning occurring late in the task; Repeated return to plan with limited task progress or enactment of task.* *Sustained emphasis on behavioral regulation, distracting other regulation and task engagement.* *Monitoring reveals problems with planning, rather than task.* *No time remains at the end of the task for evaluation.*	Group norm characterized by effective regulation taken up/accepted, but merely focuses on task completion, task directions, processes. *Planning and monitoring toward task completion (e.g., checking spelling and formatting; meeting task requirements), but not more. This yields lower quality regulation of the task goal.* *Evaluation is brief, and recognizes completion or meeting minimal task requirements.*	Group norm characterized by effective regulation toward high-quality understanding reflected in the task, group-set goals for understanding. Regulation is taken up/accepted. *Planning and monitoring toward task or group-set goals (as supported by the curriculum) focus on understanding, improvement, progress, integration, consensus, revisions, task quality as exemplified in task expectations* *Evaluation at the end of a task of whether making progress or meeting their goals.*

Disciplinary Engagement Dimension: Content of collaborative talk or physical activity characterized by new contributions aimed at making intellectual progress, involving integrated conceptual and disciplinary activity			
1: ***Low***	**2:** ***Moderate-Low***	**3:** ***Moderate***	***4: High***
Group norm characterized by limited to no collaborative content or disciplinary talk and physical activity *Group disengagement; limited task work* *Independent activity with no content/disciplinary talk or gesture*	Group norm characterized by fragmented talk, with no elaboration or attempts to connect (e.g., restating terms; recall of discrete facts) *OR* Focus on content and practices as facts, memorization, recall, or reproduction of practices *Brainstorming, eliciting prior relevant knowledge,* *Building or independent activity with gesture and physicality, without on-task discourse*	Group norm characterized by content of collaborative talk or physical activity involves some brief elaboration or connections of facts, terms, content and/or practices; elaborative telling *Brief elaboration of a term* *Brief or initial work toward a connection*	Group norm characterized by content of collaborative talk or physical activity integrates content with practice OR content or practice, with rationale/explanation, toward solving lesson/unit problem; intellectual progress Group explicitly identifies how their content and/or practice activity generates needed knowledge to solve task/problem *Synthesis, conceptual connections, connections between content and practice, extended elaboration that informs conceptual development, Justifications/rationale*

Note. Italics provide example indicators of ratings. An updated version of the rubric is available from the authors.