Peer discussions in lecture-based tutorials in introductory physics

Risto Leinonen,^{*} Mervi A. Asikainen, and Pekka E. Hirvonen

*Department of Physics and Mathematics, University of Eastern Finland,
Yliopistokatu 7, 80100 Joensuu, Finland*

(Received 20 December 2016; published 31 March 2017)

This study analyzes the types of peer discussion that occur during lecture-based tutorial sessions. It focuses in particular on whether discussions of this kind have certain characteristics that might indicate success in the post-testing phase. The data were collected during an introductory physics course. The main data set was gathered with the aid of audio recordings. Data-driven content analysis was applied in the analysis to facilitate the placement of students' discussions in particular categories related to different types of discussions. Four major discussion types were found: discussions related to the content knowledge, metalevel discussions including metaconceptual and metacognitive elements, discussions related to practical issues, and creating a base for discussion, seen here in the order of their prevalence. These categories were found to possess individual substructures that involved, for example, asking and answering questions, participating in a dialogue, or disagreeing with a peer. Analyzing the substructures of the categories revealed that there were evident differences between the groups, some of them related to the group size. With regard to the characteristics of discussions considered to be connected to a better learning outcome, it was observed that a great number of lines uttered related to the physics content or metalevel discussions seemed to have a direct bearing on success in the post test at the group level. For individual students, answering content-related questions posed by their peers might also indicate success in the post test. We would encourage researchers to continue this type of research in order to discover the essential characteristics of students' discussions that facilitate learning.

DOI: [10.1103/PhysRevPhysEducRes.13.010114](https://doi.org/10.1103/PhysRevPhysEducRes.13.010114)

I. INTRODUCTION

Because of their evident benefits for learning, peer discussions are often used in teaching. They are also essential in *tutorials*, a research-based curriculum for teaching physics at the introductory level [1]. In the present study, tutorial discussions are analyzed to discover the kinds of discussion that are used and also to find whether some of them are connected with students' answers in the post-testing phase.

A. Peer discussions in learning

The underlying learning theory for peer discussions relies on social constructivism [2]. This means that social interactions with critical thinking processes comprise some of the key factors for better learning. The idea of a zone of proximal development (ZPD) is strongly related to this; ZPD refers to one's personal imaginary zone where learning will not occur individually but can take place when properly aided. Typically, this aid is provided by other

people. Such people can include peers or instructors, with improved learning occurring in such a context [3,4]. The advantages of peer discussions are acknowledged by learners themselves, which coincide well with the social constructivist view of learning. For example, learners realize that explaining their thinking and listening to their peers' explanations help them to learn [3]. We would also claim that the benefits of peer discussions can be explained with the aid of social constructivism, even if there can be other variables affecting cognitive processes, such as assignments and experiments [5].

An essential benefit of discussions is the improvement in the learning outcome, as shown in numerous different types of studies [6–9]. For example, discussions in peer instruction consist of numerous short (2–4 min) sections where, after pondering on them individually, students discuss their answers to so-called ConcepTests during lectures. Peer instruction has improved students' conceptual reasoning and problem-solving abilities [7]. In tutorials, students in small groups hold discussions guided by the tutorial worksheets. Tutorial teaching has helped students in constructing and applying important concepts and principles in the field of physics [1,6]. In the case of cooperative group problem solving, students are assigned certain roles and a procedure to follow while solving context-rich problems. This type of instruction has been shown to be beneficial in developing students' problem-solving abilities [10,11].

^{*}risto.leinonen@uef.fi

Published by the American Physical Society under the terms of the *Creative Commons Attribution 4.0 International license*. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Peer discussions can also improve critical thinking skills, such as helping students to view topics from multiple perspectives [9,12,13]. Moreover, they can stimulate students' interest in the content itself [9,10]. For those interested in the details, the literature reviews published by Pollock *et al.* [9] and Prince [12] are recommended.

A number of problems related to peer discussions have also been reported. It has been observed, for example, that peer discussions may not lead students to the desired learning outcome, if their preknowledge is insufficient. If they lack a sufficient knowledge of the relevant physics concepts, students tend to search for the relevant formulas in the textbook, an activity that does not result in productive discussions [10,14]. Another issue concerns the problems posed by certain classroom environments: large classrooms tend to make functional discussions challenging to implement [3].

Student characteristics, such as their learning style, their attitude to learning activities, and the number of messages delivered, have been claimed to have more impact on learning than group characteristics [15]. The passivity of some group members can lead to less productive discussions; the same effect may be observed when group members do not trust each other sufficiently [10]. Moreover, it has been shown that the most successful group in terms of learning outcome seems to be the one with the most regular and active participation in the discussions [16].

One important issue is that assignments should be challenging enough for students to become enthusiastic about understanding what is going on. If the assignments are too straightforward or simple, students may tend to rely on oversimplified explanations and ignore some of the essential factors [10].

Regarding the optimal group size for beneficial discussions, Alexopoulou and Driver [17] compared the relative effectiveness of discussions in pairs and in fours. Their finding was that discussions in groups of four were more beneficial. This appeared to result from the fact that pair discussions constrained interactive modes more than discussions involving four students. For their part, Heller and Hollabaugh [11] suggest that the optimal size would be three students. This finding was based on the observation that in larger groups some students tended to some extent to be excluded, while in pair discussions there is no mechanism for choosing between two opposing viewpoints.

B. Tutorials in physics education research

A major emphasis in physics education research has been placed on studying students' conceptions [18–20]. This ongoing work has been going on for several decades, and the view of students' conceptions concerning the most essential physics topics is quite wide ranging. Based on this body of research, the PER community has developed various types of materials and curricula to improve the learning of physics. One of the most famous outcomes is

that of tutorials that have been designed to supplement conventional university teaching to help students to overcome their most common learning difficulties and develop their scientific reasoning skills [1]. The tutorials include a particular method of teaching physics in a classroom setting, typically at the introductory university level, with their focus placed especially on developing the meanings of the essential concepts of physics.

The tutorial sessions are based on the textbook *Tutorials in Introductory Physics* [1], which contains approximately 50 tutorials related to different physics topics. A separate homework book includes related homework assignments [21]. The Instructor's Guide includes pretests, exam questions, and instructions and hints concerning individual tutorials [8].

Conventionally, tutorial sessions are held after the content has been addressed in lectures. Tutorial sessions start with a 10 min pretest, typically implemented as an online test. Pretests inform both instructors and students about the level of students' conceptual understanding.

Generally speaking, the tutorials are designed to be held in a classroom with approximately 20 students and two teaching assistants. During the tutorial sessions the students work through the tutorial worksheets in groups of four or five. The worksheets concentrate on constructing the meanings of the concepts and laws of physics. The emphasis is on helping students to find their own answers to the tasks with the aid of discussions with their peers and tutorial instructors. The tutorials are designed to be completed in 50–60 min. It is quite common for some students to be unable to complete all of the tutorial tasks during the time available, but the tutorials have been designed so that the most essential elements are introduced at the beginning of each session. Some of the tutorials also include experimental work such as conducting rolling experiments with balls or constructing dc circuits. Following each tutorial session the students are given homework exercises that focus on the most important issues dealt with in the tutorial.

The key factor of the tutorials from the perspective of learning is that they promote students' *active mental engagement* in the learning process [1,22]. Active engagement is related to *active learning*, which refers to the instructional methods engaging students in the learning process. In addition to the meaningful learning activities undertaken by the students, the sessions also include an element for which the students can think about what they are doing [12]. Studies of the process of active learning have shown that actively engaging students in the learning process evidently improves their learning, a finding that also holds true in other contexts than tutorials [12,23,24]. In addition, the learning outcome seems to be longer lasting than can be achieved using more conventional methods [23].

C. Research questions

In the present study the tutorials were implemented in a lecture setting because of the constraints posed by the

limited resources that our institution, like many others, could provide. Nevertheless, the most important features, that is, the peer discussions and worksheets, were present when the tutorials were implemented in this manner [25]. The greatest difference was caused by the number of tutorial instructors: for the purposes of our study we had two instructors dealing with approximately 60 students, which inevitably reduced the availability of student-instructor discussions and placed a larger emphasis on the peer discussions.

Regardless of the wealth of evidence concerning the benefits of discussions and tutorials, we would claim that tutorial discussions *per se* would deserve more attention. Their benefits for learning are known, but what students actually discuss in the tutorials remains unclear. Hence, one of our aims has been to find whether it is possible to define some of the indicators that could predict success in post testing. In consequence, our research questions were formulated as follows:

- (1) What types of peer discussions take place during tutorial sessions in a lecture setting?
- (2) Which types of discussions, if any, appear to predict success in the post test?

We were less interested in the accuracy of students' discussions in terms of physics than in how much of the discussions addressed actual physics content, practical issues, and whether they were devoted, for example, to debating the required topics. With respect to the second research question, we aimed to discover whether certain types of discussion were connected with students' success in the post test.

II. IMPLEMENTING THE STUDY

The following section is concerned primarily with the description of the context of the study, and also with descriptions of the methods of data gathering and analysis.

A. Context

The data-gathering occurred in the Basic Physics IV introductory course at the Department of Physics and Mathematics at the University of Eastern Finland during spring semester 2015. The course consists of lectures (35×45 min), homework sessions (16×45 min), and three tutorial sessions (2×45 min each). The course is generally based on a textbook by Knight [26]. The content of the course consists of superposition, wave optics, ray optics, optical instruments, the foundations of modern physics, and quantization.

In the course of the lectures the lecturer (Asikainen) went through the topics with the aid of PowerPoint slides. The lectures were frequently supplemented with assignments that required peer discussions. In the homework sessions the students were expected to present their solutions to the homework assignments. The role of the teaching assistant

(Leinonen) was to comment, supplement, and evaluate these solutions, and to show sample solutions to the most challenging tasks. In addition, the students received a number of extra assignments for the homework sessions, which they were encouraged to solve in groups rather than individually. The students were rewarded with extra points for the homework done in the course evaluation. Participation in the lectures and homework sessions was voluntary for the students, with the percentage of students participating in lectures and homework sessions standing at approximately 50%.

Prior to this course the students had taken the Basic Physics I–III courses covering topics related to kinematics, mechanics, thermal physics, oscillations, and electricity and magnetism [26]. The students had also taken—or were simultaneously taking—a laboratory course dealing with the topics covered in the lecture courses.

The number of students enrolling for the course stood at approximately 90. Their major subjects consisted of mathematics, physics, chemistry, pedagogy, and computer science. Since the teachers and scientists had also been educated in the department, the course cohort was relatively heterogeneous.

B. The tutorial

The *Two-Source Interference* [1] tutorial was the first of those held in the Basic Physics IV course. The tutorial was held after superposition, standing waves, and interference had been introduced in the lectures. Participating in the tutorial was voluntary for the students, but the participants were rewarded with a few extra points in the course evaluation. The total number of students participating in the tutorial was 61. The tutorial was staged in a lecture hall, with the students permitted to sit only in the odd-numbered rows in order to leave space for the two instructors to participate in the discussions. The students were instructed to work in small groups of any size. Typically, they chose to work in groups of two or three, but some bigger groups and also individual students were observed, as well.

The tutorial consisted of three cycles. Each cycle began with one or two prequestions presented with the aid of PowerPoint slides. The students answered these questions individually. Next, they were asked to proceed on to the tutorial worksheets in groups. These worksheets concentrated on the content related to the physics themes addressed in the prequestions. After a suitable amount of time with the worksheets, the students were then asked to answer the same question(s) again. This cycle was repeated three times, each cycle taking 25 to 30 min.

The first cycle concentrated on graphical presentations of two-dimensional water waves, to discover the initial phase and then the displacement of water when crests meet crests or crests meet troughs. The second cycle started by addressing the idea of the change in displacement over time at various spots when two waves met, and then

TABLE I. Discussion group composition. Female discussants are presented in italics.

Group	Members
G1	S1, S2
G2	S3, S4
G3	S5, S6
G4	S7, S8, S9
G5	S10, S11, S12
G6	S13, S14, S15
G7	S16, S17, S18, S19

defining the path length differences from two sources to these positions. This led to the idea of nodal and antinodal lines, and also to the role played by phase difference in interference. The third cycle concentrated on the distance of two sources and how distance affected the locations and numbers of nodal and antinodal lines.

As a result of the focus of this present article, a more detailed description of the tutorials will not be provided here, but a few examples of the tutorial questions appear in the Appendix.

C. Data gathering and analysis

Two separate data sets were gathered during the tutorial. The first set consisted of students' tutorial worksheets, including their answers to the test items. The second set consisted of audio-recordings of the students' discussions during the tutorial.

The tutorial had three pairs (pre + post) of test items.¹ The students' answers to these items were investigated by means of content analysis [27]. The underlying idea of this analysis was to find similarities between the students' answers so that they could be categorized. The main emphasis in the analysis was on the students' explanations, while their selections in the multiple-choice questions were simply used to support the conclusions that had been drawn. Finally, the students' responses were placed in three categories (acceptable answers, inadequate answers, and empty or uncategorized answers), which were later used to mirror the students' discussions.

The audio recordings were gathered from seven groups (G1–G7) consisting of a total of 19 students. The groups consisted of two, three, or four members, as seen in Table I, which also includes information about the genders of the discussants.

The responses provided by the recorded students helped us to reduce the data to manageable proportions. In light of our research aim, we hoped to obtain a wealth of audio data

in terms of their diversity. Hence, we decided to concentrate on the audio data related to test item 2(a) addressing the concept of a path length difference (Fig. 1). This was a discretionary choice based on the fact that the responses of the recorded students in this item appeared to provide more insightful views of the students' ideas than did the others, possibly due to the fact that this item was open in nature in comparison to the test items 1 and 3 that were multiple-choice questions. Moreover, in the course of the analysis it became apparent that not all of the recorded students managed to work through the whole tutorial worksheets in their discussions, but all of them covered the themes related to path length difference related to test item 2(a); discussions related to test item 2(b), addressing phase difference, were not finished among all recorded groups.

The discussions were analyzed by means of data-driven content analysis [27], although we had some preliminary understanding of the categorization framework based on the literature [28]. In fact, our categorization system was inspired by the framework rather than being entirely based on it, with some of the category characters instead following those of Hogan, Nastasi, and Pressley [28].

The procedure began by transcribing the discussions. This phase was done almost word for word, with only the unnecessary interjections (e.g., “well” and “erm”) omitted. Whenever students happened to speak simultaneously, their utterances were transcribed as if they had occurred sequentially. In a very few instances the students' speech was incomprehensible, and after due consideration such instances were generally omitted from the analysis.

In the next stage, all of the discussions were read closely so that an overview of the types of discussions could be obtained. In addition, note was made of the various preliminary discussion types observed. It was also observed that, due to their ambiguity, certain individual words and sounds (e.g., “so” and “Mm-m”) would contribute little to the analysis and hence they were removed. Given our research aims, the discussions with the instructors and irrelevant discussion parts were also removed. The percentages of line reductions varied between 27% and 47%, the average being 36%. This may suggest that large amounts of data were lost, but this was not in fact the case. When, for example, interjections aimed at intervening in another student's utterance were deleted, some of the utterances produced by a single individual could be combined into one whenever they were related to one specific issue. In terms of the word count, the reductions varied between 7% and 30%, the average being 19%. There were plenty of variations in the numbers of lines eventually subjected to analysis² between the groups, varying between 69 and 121 lines, the average being 100. With respect to the word count, the numbers varied between 652 and 1897 words, the average being 1149. The discussion segments

¹All these items addressed different elements related to interference, such as the distance of sources, path length difference, phase difference, and wavelength. The test items 1 and 3 were multiple-choice questions while the test items 2(a) and 2(b) were more open in nature.

²Following the data reduction described earlier.

2a. Estimate the path length difference at points A, B, and C. Express your answer in terms of wavelengths. Explain how you reached this conclusion.

	Point A	Point B	Point C
Δr

FIG. 1. The test item 2(a), designed to test students' ideas about path length difference.

that were analyzed lasted between 15 and 18 min, depending on whether the students had proceeded to the next assignments.

Following this data reduction, the students' speeches were precategorized based on the discussion types observed during the preliminary reading. The basic unit for the categorization was an *utterance* or *line*. One line refers to something said by an individual student without any noticeable breaks or interruptions. A line could consist of one word (e.g., "Really?", "Why?"), one sentence (e.g., "Weren't these lines already marked here?"), or numerous sentences (e.g., "Yeah, but still. Because shouldn't it get weaker as it goes further? Because if you throw a rock into the water, first there are big waves, and then they get smaller because they do not proceed to the other side of the lake."). Whatever their length, all of the lines were treated equally in the analysis. One line could include more than one type of discussion, and our categorization system was constructed to permit one line to be placed in more than one category. The reason for our concentration on the frequencies of the lines uttered, rather than using interval-based coding, is that in this kind of study the time used does not necessarily correlate with the value of something said. For example, stating "it does not go like that" takes only a few seconds but it was observed that these types of comments frequently stimulated valuable discussions.

Leinonen had the primary responsibility for the analysis and development of the coding criteria. During the analysis it was observed that the precategorization did not describe the data adequately, and so new discussion types had to be found. This made analyzing the data a cyclic process during which the categories were combined, split, elaborated, etc. In this phase, the categorization and the criteria of categories and their substructures were discussed a number of times between the researchers (authors) in order to enhance the trustworthiness of the study. Besides the categories and their criteria, ambiguous and unclear lines uttered were discussed between the researchers. In practical terms, coding was conducted with the aid of Microsoft Excel software, where each line was coded according to the main categories and their substructures. At the end of this process, a categorization system describing the data adequately had been constructed, and all the discussions

concerning the lines had been assigned to four main categories. These categories were identified as *creating a base for discussions*, *discussions related to the physics content*, *discussions related to practical issues*, and *metalevel discussions*. Table II elaborates on the substructure of the categories, using authentic quotes. At the end, the categorization system proved to be relatively unambiguous (see subcategories and example quotes in Table II). Some individual ambiguities were located in the category of metalevel discussions, but these were discussed inside the research group until consensus for the subcategory descriptions and criteria was reached.

The first three categories are quite self-explanatory in terms of their substructural descriptions, but the meaning of "metalevel discussions" would deserve further explanation. This category includes elements from two separate, yet connected, metalevels, namely, the metacognitive and the metaconceptual. The metacognitive level refers to the knowledge that the individual has about their own cognition [29]. This description can be elaborated with three types of knowledge related to metacognition: one's knowledge concerning cognition, a functional awareness of one's knowledge, and controlling one's cognitive processes [30]. Because of the context of the present study, we cannot concentrate solely on the individual student's metacognition, but the matter does also include a social element. This means that metacognition cannot necessarily be linked to individual learners but is social in nature. This is referred to as socially shared metacognition [31].

Metaconceptual processes³ refer to processes related to one's conceptual system [30,32]. It has been claimed that these processes include four components: metaconceptual knowledge, awareness, monitoring, and evaluation [30]. For our purposes it is enough to state that it refers to processes where a learner reflects, evaluates, or relies on his or her current conceptions in speaking. As with metacognition, these processes can be social in nature, even if the actual body of conceptions is a property of the individual student.

³Sometimes metaconceptual processes are seen as metacognitive processes, but this depends on the authors.

TABLE II. A substructure of the four main categories observed in students' discussions. Sample quotes for the subcategories are included. Each line placed in a given category has been underlined; the surrounding text extracts are included discretionally to illustrate the context of the particular discussion. The symbols in parentheses after the quotation refer to the individual students (S1–S19).

The main category	Subcategories	Example quote
Creating a base for discussions	A student reads worksheet assignments aloud	<u>How does the surface of the water change at this stage?</u> (S4)
Discussions related to the physics content	Asking questions	<u>Wait a moment. Isn't one wave like two times pi?</u> (S8)
	Answers to the questions from a peer	Yeah. (S9) <u>And what is that path length difference? Is it just a remainder of those two?"</u> (S6) <u>Well, I think it is that one as a positive figure. So it's the absolute value.</u> (S5)
	Answers to the questions in the tutorial worksheet	<u>How does the surface of the water change at this point?"</u> (tutorial) <u>Well, it varies between 2A and -2A.</u> (S19) <u>You have one</u> (S16)
	Dialogue where more than one student participates actively, and the discussion is not simply a process of asking and answering questions	<u>And from r2 it is..</u> (S19) <u>... two and a half</u> (S18) <u>There was one there. So two and a half, yup.</u> (S19) <u>Doesn't it go like up and down again, or does it?</u> (S8)
	One student is teaching the others in a way that goes beyond answering the questions presented by his/her peers	<u>Well. Let's start with one point, so of course we have...</u> <u>Let's take just one moving part. Or one, say a wave source. Of course, during the time we're watching it, there is a crest. Let's take a look at that same part and proceed in time, and it will take on this up and down motion. Like a buoy in the waves. But now we should think about what that buoy would do in a cross sea.</u> (S7)
	Other	...
	Asking questions	<u>Is my point in a different location, or is it in the same one?</u> (S2)
	Answers to questions posed by a peer	<u>Hey, do you have a ruler?</u> (S5) <u>Yes I do.</u> (S6)
	Other	...
	Agrees with something said by his/her peers	<u>If you place it somewhere where there is a crest or a trough, it will go crest, trough, crest, trough.</u> (S5) <u>Yeah, that's where it moves most, true.</u> (S6)
Metalevel discussions	Disagrees with something said by his/her peers	<u>But here it is zero for both.</u> (S10) <u>No, it is not.</u> (S11)
	Reflects on something	<u>Or is it the whole axis?</u> (S11) <u>I think that it is in practice.</u> (S10) <u>Ah, yes it is!</u> (S11)
	Reflects on problems in understanding or expressing uncertainty (two examples)	Example 1: <u>Well, of course it is moving least at those zero points because there is no motion at all.</u> (S7) <u>I don't get it.</u> (S8) Example 2: <u>But how? Is this wrong now? Or should it always be half? Or why?</u> (S8)
	Asks for confirmation of his/her own answers	<u>If you make this kind of triangle here. Isn't it possible then, in theory?</u> (S15)
	Asks for more explanations or elaborations from peers	<u>I also think it looks like it's moving when seen in nature, but why?</u> (S8)
	Other	...

It might appear that these two metalevels are easy to distinguish, but significant challenges have to be faced while analyzing students' discussions in an authentic context. This difficulty is also seen in the literature, where researchers use metacognition as a general term for situations where students' thinking about their conceptions is analyzed rather than considering their actual cognition [28,33]. Students' comments often reflect some kind of metalevel thinking, but this remains ambiguous if they are evaluating their own thinking or conceptions. For example, if a student states that "This is illogical," it is impossible to judge whether that is due to a disparity between the answer and his body of conceptions or whether it is related to the individual's search for consistency in thinking. Moreover, it should be remembered that speech can only partially reflect one's thinking. Hence, we decided to combine students' lines that could be regarded as either metacognitive or metaconceptual within a single category in order to avoid possible misinterpretations of the data.

III. RESULTS

The results are presented here in two parts. First, we pay attention to the responses to a test item made by the whole cohort of students both before and after the discussions had been introduced. Second, attention is then paid to the characteristics of the peer discussions undertaken by the recorded students and the possible connections to their answers in the post-test item.

A. Students' written responses concerning path length difference

Students' ideas concerning path length difference were evaluated with the test item seen in Fig. 1. Students answered the item before and after the tutorial assignments related to these topics.

The categorization of students' responses to the item can be seen in Fig. 2. The students' answers to these three items (the points A, B, and C) were analyzed as a single, combined item due to the fact that the students typically used the same explanations for all their answers, and we were more interested in their explanations than in the numerical values provided. The students' answers were regarded as acceptable if their explanations were adequate for the correct answer to be achieved, even if the numerical answers they provided were not precise or there were minor errors in their use of symbols. "Inadequate answers" included students' explanations where they made references to interference without further explanations, where they ignored something essential, or where they confused path length difference and path length. The final category includes the responses that did not fit into previous categories, and also the empty answers, with emphasis placed on the latter.

Thirty-seven students produced acceptable answers after the tutorial sessions, with 21 of them providing

FIG. 2. The categorized responses of the whole cohort. The numbers within the circles show the numbers of students placed in the particular category in the pre- and post-testing phases. The numbers within the rectangles indicate the frequencies of transfers between the pre- and post tests $N = 61$.

acceptable answers during the discussions. Only two students changed their acceptable answers to something else as a result of the tutorial discussions. Nineteen students did not change their inadequate, empty, or uncategorized answers at all during the tutorial, while 24 students were unable in general to provide acceptable answers in the post test.

With respect to the 19 students (S1–S19) in the seven groups (G1–G7) whose discussions were recorded, eight produced acceptable answers in the pretest. In response to the tutorial discussions, six of the students changed their inadequate or empty responses to acceptable ones. There were no groups where all of the members gave acceptable answers in both the pre- and the post tests. All of the members in four groups (G1, G4, G5, and G7) produced the desired response following participation in the tutorial sessions. Some of the students in groups G3 and G6 produced the correct response following the tutorial session, but none of the students in group G2 produced it during the actual tutorial sessions. The details of the categorized answers for the individual students and groups can be seen in Figs. 3–6.

The recorded students represent the whole cohort well, though with one minor exception: none of the students who changed their responses to inadequate or uncategorized ones during the peer discussions happened to be recorded. Nevertheless, since there were only two of these in the whole cohort, it seemed to be quite a rare phenomenon.

B. Students' peer discussions

The students' discussions were analyzed to find the main discussion types that they experienced in the course of the tutorials. The discussions were analyzed at the student level, but our findings at the group level are also

FIG. 6. A substructure of the category “metalevel discussions.”

discretionally highlighted. The numbers of lines analyzed from each of the groups can be seen in Table III, while the categorization of the four major types of discussion observed may be found in Fig. 3.

The first observation from Table III and Fig. 3 is that groups G2, G3, and G6 with the smallest number of lines uttered were those in which some students produced inadequate or uncategorized answers in the post test. This suggests that the number of lines uttered that were related to the topic at hand was connected to those that followed in the later post test.

The frequencies in the category “Creating a base for discussion” vary for individual students between 0 and 9 lines. No clear connection can be seen between the frequencies for individual students and their success in the post test. When the frequencies between the groups were evaluated, it was seen that groups G2, G3, and G6 had higher frequencies on average than the other groups (12.3 vs 9.3 lines). Those mentioned previously were also groups where some of the students produced inadequate or uncategorized answers in the post test.

The category “discussions related to the physics content” was clearly the largest, and all of the students’ apart from S17 had their highest frequencies in this category. The substructure of this category reveals differences in the lines uttered by students, as can be seen in Fig. 4. On frequent

occasions, certain students, such as S2 and S8, posed numerous questions related to the content. These groups, G1 and G4, also contained students (S1 and S7) who answered numerous questions and taught their peers in a way that exceeds the necessity simply to answer their questions. One interesting finding is that answering the tutorial questions was essentially less common than answering questions posed by peers. With respect to the subcategory of “dialogue,” the frequencies between the groups varied greatly, but they were quite similar to each other within the groups. It appears that dialogues can play a significant role in discussions, but they may not guarantee success in the post test, as group G2, containing two students with inadequate answers in both the pre- and post-test items, shows.

When the groups are compared, those that were successful in the post test, namely, G1, G4, G5, and G7, had larger frequencies than the one where only partially inadequate answers were produced, on average 77 versus 45 lines. This suggests that discussing the actual physics content may well predict success in the post test.

Figure 5 shows the frequencies in the category “discussions related to practical issues.” Practical issues included how to write an answer down, how to draw something, and finding out what the other group members were doing. The substructure of this category can be seen in Fig. 5. The first observation that can be made concerns the disparity between asking and answering questions. In each of the groups apart from G7 there is one student who asks more questions than the other students, while in some groups there is also one student who answers most of the questions. However, asking or answering questions related to the practical issues does not seem to be connected with the students’ answers in either the pre- or the post tests. For example, student S5 posed a considerable number of questions concerning the practical issues, but his content knowledge was solid, as may be seen from his answers to the test item and discussions.

TABLE III. The numbers of lines analyzed for each group.

Group	Number of lines analyzed
G1	110
G2	86
G3	69
G4	121
G5	121
G6	75
G7	119

The frequencies between the successful and unsuccessful groups in this category were similar to such a degree that it offers no adequate evidence in any direction concerning their connections to the accuracy of the students' answers in the post test.

The substructure of the category of metalevel discussions may be seen in Fig. 6. The frequencies of the subcategories are rather small, and it is difficult to say whether some of the characteristics describe typical metalevel discussions. There is considerable variation between the groups and the students. For example, students S8, S11, S15, and S18 have more than 10 lines categorized as metalevel discussions, while the frequencies are visibly low, and almost absent, for students S9, S12, S13, and S17. These frequencies do not seem to correlate with the students' answers in the post test. For example, student S17 remained almost silent throughout the discussions but used his few lines to point out the mistakes that others had made, which stimulated the other group members to reevaluate their answers, and they also produced desired responses in the post test.

Analysis at group level shows that the successful groups had greater frequencies in this category than the unsuccessful ones, 25 vs 14 lines. Another observation that can be made is that the groups with more than two students, namely, groups G4–G7, used significantly more lines for metalevel discussions than the pair of groups G1–G3, 26 as opposed to 13 lines on average. This suggests that larger groups might stimulate more metalevel discussions and hence also contribute to the success in the post test.

IV. DISCUSSION

In the course of this study we have examined the types of discussions occurring in the course of a lecture-based tutorial session. In addition, we have evaluated students' answers to the test-item before and after the discussions and analyzed the similarities and differences between the discussions held by groups and students with different types of answers in the post test.

With respect to the answer to the first research question "*What types of peer discussions take place during tutorial sessions in a lecture setting?*", four main categories were constructed based on the data and previous literature. The first category includes the parts where students read tutorial assignments aloud. This category is rather small, yet it appears to be important since the lines in this category seem to function as discussion openers. Another role played by this category may be that students read their assignments aloud in order to achieve a consensus of opinion about the exact aim of the assignment.

The second category involves discussions related to the physics content. This is by far the largest category for all of the groups. This category includes questions, answers, dialogue, and parts where one discussant teaches the others

in a manner that exceeds solely answering questions. Great differences were observed in the substructure between the students and the groups, which indicates that discussions may proceed differently inside this category. One important finding is that answering questions posed by peers was more common than answering questions from the tutorial worksheet. This indicates that tutorial assignments are likely to be insufficient for teaching the content without discussions, but they also seem to raise new questions that in turn stimulate learning.

The category of practical issues is concentrated in the lines where students discuss issues such as how to formulate their answers to the tutorial assignments or how to draw pictures. It is probable that the discussions in this category cannot contribute much to the actual learning, but they nevertheless play an important role in, for example, causing students to think about how to write down something that they already seem to understand. However, we also consider it somewhat surprising that university students might use almost one-third (G3) of the lines uttered to discuss practical issues while the questions in the tutorial worksheets may appear to instructors to be rather straightforward.

The last category includes the discussions related to metalevel issues, such as agreeing or disagreeing with peers, or reflecting on understanding something. This is an important category because it should reflect students' tendency to evaluate the conceptions and to think critically about themselves and their peers. We interpret the frequencies in this category as rather modest because one might assume that no learning can occur without some of the processes related to this category. Alternatively, we should remember that this category may also involve expressions and gestures that cannot be observed with the aid of audio recordings, and hence our numbers may be somewhat dismissive. One of the important findings was that the groups consisting of three or four students held a larger number of metalevel discussions than did the pairs of students, which in turn suggests that students' ideas are challenged more frequently and effectively in larger groups than in pairs. This may be a matter of probabilities; the larger the group, the larger are the chances that there will be a student with adequate skills to ignite metalevel discussions.

The answer to the second research question "*Which types of discussions, if any, appear to predict success in the post test?*" is more descriptive than declaratory. One general finding is that the successful groups uttered many more lines related to the topic than did the groups where some of the students provided inaccurate answers in the post test. With respect to the categories *per se*, a large number of lines used in reading tutorial assignments aloud seems to correlate with poorer success in the post-test assignment when evaluated at the group level. Besides, we observed that the frequencies of lines in categories "discussions

related to the physics content” and “metalevel discussions” seem to be connected with the accurate answers in the post test, especially while evaluated at the group level. With respect to the differences between individual students, one finding is that the students who answered numerous questions concerning the content-related issues and taught the others seemed to produce the desired responses in the post test. It seems likely that this was the result of their good content knowledge and self-efficacy, which were quite noticeable in their discussions, especially in the case of student S7.

These findings suggest that there are some similarities between the successful groups but also that the findings shown here should be interpreted with caution. Even if connections have been found between the students’ discussions and their responses in the post test, no specific causalities have been proven. It is not claimed that certain types of discussions cause better learning but that there seems to be a connection between them and success in the post test, although not necessarily a causal one. Perhaps students with better content knowledge or cognitive abilities in general have similarities in their way of discussing, such as answering content-related questions accurately or challenging others with adequate comments, for example. We can claim that the discussion types referred to above indicate success, but it cannot be claimed that they represent the underlying reasons for it.

The predictors for greater success in the post test were more readily observable at the group level than at the student level. Social constructivism as the underlying learning theory for the benefits of discussions supports this finding. Learning is seen as a process taking place in social interactions between students, so it is not connected solely to individual students but to social interactions. These interactions can be better observed at the group level. Hence, we suggest that analyzing discussions at the group level may be a better predictor of success than analyzing such discussions at the student level.

Our results tend to support previous findings, but there are also some evident disparities. Naturally, all comparisons should be treated with healthy skepticism as our data might not be totally representative for the whole class due to the fact that only some students and some parts of discussions were analyzed. Besides this, there are differences in data gathering and analyzing methods between different studies which might also cause some differences. That being said, we present some remarks concerning our results in light of previous studies.

It has been claimed that groups with the most even and active participation have the best learning outcome [15,16], but this is not supported by our results; the students in groups G4, G5, and G7 displayed considerable differences (see Fig. 3) in the numbers of lines uttered, even if the numbers of lines uttered at the group level were similar

(Table III). Yet all of the students in these groups managed to reach the desired outcome.

It has been claimed that individual student characteristics have more impact on learning than group characteristics [15]. This notion is partially challenged by our results; it was observed that the similarities in the discussions held by the groups succeeding in the post test are greater than they are for the individual successful students. However, we do not have enough evidence to claim that the differences between the groups impact on their success in the post test but that they are easier to observe at the group level. One should also remember that some individual students in our study would probably have succeeded in the assignments without the benefit of any discussions.

The claim regarding the challenges posed by large classrooms for functional discussions to take place is not supported by our results [3]. We did not observe essential problems while implementing peer discussions in a lecture hall; even teachers could participate in discussions easily when every second row was devoted to maximizing their mobility.

Our results support the finding that in groups with more than two students, one of the students may to some extent be excluded from the discussion [11,17]. This does not necessarily mean that these students are not actively involved in the progress of the task at hand; rather, they may be following from the side and commenting on topics whenever they consider it necessary, such as correcting others’ answers.

The present study also stimulates ideas for further research projects. The structures of discussions could be studied in more depth to provide an understanding of the kinds of entities that the discussions form. At present, we have concentrated on individual lines rather than entities. In addition, further emphasis could be placed on the roles that individual students adopt in discussions. There are some signals in our data that the roles might emerge naturally as the discussions proceed, but studying this thoroughly would require different methods of analysis. Moreover, it would be interesting to see whether discussions change if the topic of tutorials were to change, or whether the topics for discussion also included some experimental parts.

Our findings also provide some suggestions for the implementation of tutorials. We would suggest that students should be encouraged to work in groups of three or four rather than in pairs since group size appears to correlate with a greater number of metalevel discussions. There is the potential disadvantage that larger groups may cause students’ participation in the discussions to become uneven, but this is also related to students’ personalities and we did not observe that quietness correlated with inaccurate answers in the post test. It would also be important to consider whether students should be allowed to choose their groups freely or whether the composition of the groups should be guided by the instructors based on

students' preknowledge. Our findings suggest that success in the post test is unlikely to occur if the preknowledge of every student in a group is inadequate, so it would seem essential that each group should have at least one student with adequate content knowledge and preferably also the skills needed to ignite metalevel discussions. Hence, if groups' compositions were based on certain students' skills diagnosed before tutorials instead of their own preferences, it might add diversity for the discussions experienced and likely enhance learning.

The detailed description of the types of discussions experienced in peer discussions during tutorial sessions has enriched the field of educational research by revealing what students' discussions are actually related to. This description has theoretical value in its own right. Despite the extensive body of research concerning the benefits of discussions during tutorials, we have found no descriptions of this type in the literature. Regarding the practical value, our findings can work as a base for developing new types of tasks that would, for example, place greater emphasis on metalevel discussion. Furthermore, our findings could offer a starting point for creating a research-based instructional framework for effective

tutorial discussions. This framework should introduce and sequence certain essential discussion elements that seem to contribute for a better learning outcome. Introducing this for students should teach them to systematically evaluate their conceptions and thinking, which does not only serve conceptual understanding and better learning in general but also one's understanding of the nature of science.

APPENDIX: EXAMPLE ASSIGNMENTS FROM THE TUTORIALWORKSHEET [1]

Describe what happens at a point on the surface of the water where

- *A crest meets a crest*
- *A trough meets a crest*

...

Consider a point on your diagram where a crest meets a crest. How would the displacement of the water surface at this point change over time?

...

For each of these points, determine the difference in distances from the point to the two sources.

-
- [1] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Pearson, Boston, 2010).
 - [2] L. Vygotsky, Interaction between learning, and development, in *Mind, and Society—The Development of Higher Psychological Processes*, edited by L. Vygotsky and M. Cole (Harvard University Press, Cambridge, MA, 1978), p. 79.
 - [3] D. J. Nicol and J. T. Boyle, Peer instruction versus class-wide discussion in large classes: A comparison of two interaction methods in the wired classroom, *Stud. Higher Educ.* **28**, 457 (2003).
 - [4] R. Leinonen, M. A. Asikainen, and P. E. Hirvonen, Overcoming students' misconceptions concerning thermal physics with the aid of hints and peer interaction during a lecture course, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020112 (2013).
 - [5] K. C. Powell and C. J. Kalina, Cognitive and social constructivism: Developing tools for an effective classroom, *Education* **130**, 241 (2009).
 - [6] P. D. Cottle and G. E. Hart, Cooperative Learning in the Tutorials of a Large Lecture Physics Class, *Res. Sci. Educ.* **26**, 219 (1996).
 - [7] C. H. Crouch and E. Mazur, Peer instruction: Ten years of experience and results, *Am. J. Phys.* **69**, 970 (2001).
 - [8] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics: Instructor's Guide* (Pearson, New Jersey, 2003).
 - [9] P. H. Pollock, K. Hamann, and B. M. Wilson, Learning through discussions: Comparing the benefits of small-group and large-class settings, *J. Political Sci. Educ.* **7**, 48 (2011).
 - [10] S. Benckert and S. Pettersson, Learning physics in small-group discussions—Three examples, *Eurasia Journal of Mathematics, Sci. Technol. Educ.* **4**, 121 (2008).
 - [11] P. Heller and M. Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
 - [12] M. Prince, Does Active Learning Work? A Review of the Research, *J. Eng. Educ.* **93**, 223 (2004).
 - [13] C. Garside, Look who's talking: A comparison of lecture and group discussion teaching strategies in developing critical thinking skills, *Commun. Educ.* **45**, 212 (1996).
 - [14] D. Laurillard, *Teaching as a Design Science: Building Pedagogical Patterns for Learning and Technology* (Routledge, London, 2013).
 - [15] T. Schellens, H. v. Keer, M. Valcke, and B. d. Wever, Learning in asynchronous discussion groups: A multilevel approach to study the influence of student, group and task characteristics, *Behav. Inf. Tech.* **26**, 55 (2007).
 - [16] S. Lindblom-Ylänne, H. Pihlajamäki, and T. Kotkas, What makes a student group successful? Student-student and student-teacher interaction in a problem-based learning environment, *Learn. Environ. Res.* **6**, 59 (2003).
 - [17] E. Alexopoulou and R. Driver, Small-group discussions in physics—Peer interaction modes in pairs and fours, *J. Residuals Sci. Technol.* **33**, 1099 (1996).
 - [18] B. S. Ambrose, P. S. Shaffer, R. N. Steinberg, and L. C. McDermott, An investigation of student understanding of

- single-slit diffraction and double-slit interference, *Am. J. Phys.* **67**, 146 (1999).
- [19] W.M. Christensen, D.E. Meltzer, and C.A. Ogilvie, Student ideas regarding entropy and the second law of thermodynamics in an introductory physics course, *Am. J. Phys.* **77**, 907 (2009).
- [20] D.E. Trowbridge and L.C. McDermott, Investigation of student understanding of the concept of acceleration in one dimension, *Am. J. Phys.* **49**, 242 (1981).
- [21] L.C. McDermott and P.S. Shaffer, *Tutorials in Introductory Physics: Homework* (Pearson, New Jersey, 2003).
- [22] L.C. McDermott, Oersted Medal Lecture 2001: Physics education research—The key to student learning, *Am. J. Phys.* **69**, 1127 (2001).
- [23] J. Bernhard, Does active engagement curricula give long-lived conceptual understanding?, in *Proceedings of GIREP2000: Physics Teacher Education Beyond 2000*, edited by R. Pinto and S. Surinach (Elsevier, Paris, 2001).
- [24] R.R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).
- [25] M. Kesonen, M.A. Asikainen, and P.E. Hirvonen, University students' difficulties in a tutorial featuring two source interference in *Physics Alive. Proceedings of the GIREP-EPEC 2011 Conference, Jyväskylä (Finland), 2012*, edited by A. Lindell, A.-L. Kähkönen, and J. Viiri (University of Jyväskylä, Finland, 2012).
- [26] R.D. Knight, *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 3rd ed. (Pearson, San Francisco, 2014).
- [27] S. Elo and H. Kyngäs, The qualitative content analysis process, *J. Adv. Nurs.* **62**, 107 (2008).
- [28] K. Hogan, B.K. Nastasi, and M. Pressley, Discourse patterns and collaborative scientific reasoning in peer and teacher-guided discussions, *Cognit. Instr.* **17**, 379 (1999).
- [29] A. Pritchard, *Ways of Learning: Learning Theories and Learning Styles in the Classroom* (Routledge, New York, 2013).
- [30] N. Yürük, A case study of one student's metaconceptual processes, and the changes in her alternative conceptions of force, and motion, *EJMSTE* **3**, 305 (2007).
- [31] T. Iiskala, M. Vauras, and E. Lehtinen, Socially-shared metacognition in peer learning?, *Hell. J. Psych.* **1**, 147 (2004).
- [32] S. Vosniadou, Capturing and modeling the process of conceptual change, *Learn. Instr.* **4**, 45 (1994).
- [33] E. C. Sayre and P. W. Irving, Brief, embedded, spontaneous metacognitive talk indicates thinking like a physicist, *Phys. Rev. ST Phys. Educ. Res.* **11**, 020121 (2015).