

Metacognitive activity in the physics student laboratory: is increased metacognition necessarily better?

Rebecca Lippmann Kung · Cedric Linder

Received: 4 November 2005 / Accepted: 22 June 2007 /
Published online: 20 July 2007
© Springer Science + Business Media, LLC 2007

Abstract In this study natural-in-action metacognitive activity during the student laboratory in university physics is explored, with an aim towards quantifying the amount of metacognition used by the students. The study investigates whether quantifying natural-in-action metacognition is possible and valuable for examining teaching and learning in these contexts. Video recordings of student groups working during three types of introductory physics laboratories were transcribed and then coded using a coding scheme developed from related research on mathematical problem solving. This scheme identifies a group's general behaviour and metacognitive activity. The study recognizes that reliably identifying metacognition is challenging, and steps are taken to improve reliability. Results suggest that a greater amount of metacognition does not appear to improve students' success in the laboratory—what appears to matter is whether the metacognition causes students to change behaviour. This study indicates that it is important to consider the outcome of metacognition, not just the amount.

Keywords Metacognition · Student laboratory · Physics teaching · Sense-making · Group work

Consider a university student halfway through a 4-h introductory physics laboratory exercise. She has managed, with the help of the teaching assistant and her laboratory partner, to measure how the rotational speed of an object changes when a torque is applied to the object. She is supposed to calculate the inertia of the object, but is unsure how to do

R. Lippmann Kung (✉) · C. Linder
Department of Physics, Uppsala University,
Box 530, Uppsala 75121, Sweden
e-mail: Rebecca.kung@fysik.uu.se

C. Linder
Department of Physics, University of the Western Cape,
7535 Cape Town, South Africa

this. Her partner suggests that they move on, because they can figure that out at home. But she would prefer to complete a rough calculation while in the laboratory¹ on the grounds that they would be able to be sure they have gathered all the necessary data and they could together check to see if their results are close to the accepted value. However, they have other measurements to complete, and might run out of time.

The type of laboratory exemplified by this short episode, as well as other types of university student laboratories, are educationally significant situations for studying the use of metacognition for the teaching and learning of science. By metacognition, we mean “cognition about cognition...understanding, monitoring, and controlling one’s knowledge and strategies” (Butterfield et al. 1995). Physics student laboratories at the university level commonly aim to increase students’ ability to think and act like scientists—including how to interpret data, troubleshoot methods, and calibrate, fix, or set up equipment. University physics teachers might imagine that students who frequently use metacognition are more successful in the laboratory than students who rarely use metacognition. In this paper the pedagogical value of studying metacognition in the university physics laboratory is investigated.

Introduction

Much research has focused on science students’ learning in the student laboratory (for example, see review articles Hofstein and Lunetta 1982, 2004; Lazarowitz and Tamir 1994; Trumper 2003). Studies have been performed to investigate how well the laboratory reaches certain educational goals, such as helping students to acquire skills or concepts, learn about the nature of science, or improve students’ attitudes toward science. Studies have also looked at how to improve the student laboratory through different course structures, equipment, student assessment techniques, etc. Additional studies focus on actual teacher or student behaviour during the student laboratory. Research linking metacognition and learning in related areas, such as reading comprehension of physics texts (Koch 2001), scientific inquiry skills (Zion et al. 2005), and mathematical problem solving (Mevarech and Fridkin 2006) has been productive. This indicates that using metacognition to research the student laboratory could productively inform teaching and learning, and contribute toward all of these research areas—learning outcomes, student laboratory design, and student behaviour.

Though a large amount of metacognitive research exists, research on students’ *natural context* metacognitive activity is rare, though authoritatively recommended, (e.g. Winne 1996; Borkowski et al. 2000) and research on metacognitive activity in science, specifically during the student laboratory, is also rare and recommended (e.g. Georgiades 2004). Most studies use one or more of the following methods. Some metacognitive studies manage to avoid measuring metacognition altogether—such studies train students in certain aspects of metacognition and measure specific learning outcomes (e.g. Kramarski and Zeichner 2001). Studies that do measure metacognition tend to use self-reports, think-aloud protocols (see Veenman et al. 2003, for a comparison of the two) and/or carefully designed research laboratory measurements (e.g. Thiede et al. 2003). However, actually measuring students’ metacognitive activity as it happens *in the wild* during a *natural-in-action* classroom setting

¹ The word “laboratory” in this paper can mean either the student laboratory (the course-associated activity) or the research laboratory. To be clear, we have tried to use the modifiers student or research—when not specified, it should be clear from the context.

raises a set of challenges that are quite different from these methods. In this paper we will discuss those challenges in an effort to determine if such research is potentially productive for informing the development of teaching and learning in the student laboratory. Thus, the overall research question for this study can be delineated as follows:

Is metacognition a useful construct for examining student learning in university-level student laboratories that teach science subjects such as physics?

This question is broken down into two sub-topics concerning the specific coding scheme developed:

Does the coding scheme developed for examining metacognition in student learning in the laboratory have a reasonable reliability?

Does the coding scheme make it possible to examine interesting differences in how students interact in the laboratory?

In the next section we describe previous schemes used to quantify the amount of metacognition used in a natural context. We then outline the research context and the basic data collection for this study. The development of the coding scheme for the laboratory context follows. Finally, we describe the results of the new coding scheme, and elaborate on the challenges found when attempting to quantify metacognition in the student laboratory.

Previous research quantifying the amount of metacognition

As mentioned previously, research on metacognition in the nature-in-action classroom setting is currently limited. One reason for this may be the inherent difficulty in classifying activity as cognition or metacognition. This is partly due to the varied definitions of metacognition and the number of conceivable dimensions encompassed by the term (Georghiades 2004). Many studies have been successful quantifying metacognition in verbal or written protocols gathered in a research laboratory setting. (Pressley 2000) However, such a setting differs greatly from a natural-in-action student-laboratory setting. The activities the subjects engage in, the number of subjects involved, and the subjects' framing and expectations about the activity all vary. One can easily imagine that the more varied the activity, the more difficult it will be to analyse.

Of the research in natural-in-action classroom settings that exists, a large proportion has focused on mathematical problem solving. This work has resulted in the development of a coding scheme for quantifying and classifying the metacognition of a group of students working on mathematical problems during class. Though not directly applicable to the student laboratory, this work provides a basic foundation to the work described in this paper, and the results from these two areas of research may be compared. The development of this scheme, as well as results from its application are described below.

Schoenfeld's classic work on the use of metacognition in problem solving (1985) classified university students' mathematical problem solving behaviour into 6 different types of episodes: *reading, analysis, exploration, planning, implementation, and verification*. During any of these episodes, students could engage in managerial behaviour—meaning decisions that guide the problem solving steps. The agreement between coders for this scheme was reported to be high, but no numbers were given (Schoenfeld 1985). By using this analysis to examine maps of students' behaviour while working in pairs, Schoenfeld identified various patterns of problem-solving behaviour. He found that student novices frequently dedicated large quantities of time to the implementation of unproductive

ideas, and insufficient time to metacognition such as planning and analysis. He consequently argued that teaching novices to use more metacognition while problem-solving would increase their success.

Artzt and Armour-Thomas (1992) built on this 6-episode analysis to code the behaviour of a single student working as part of a group of 4–5 students. They characterised each of Schoenfeld's episode types as predominately 'metacognitive' or 'cognitive', and added two additional types of episodes: *understanding the problem* (characterised as metacognitive) and *watching and listening* (not characterised as metacognitive or cognitive). Interrater reliability was 93% between the two authors and 91% between the research assistant and an author. Their results suggested that both substantial quantities of metacognition and substantial quantities of watching and listening were beneficial for arriving at a solution to a problem.

This research was then expanded by Goos et al. (2002) who focused on the social behaviour of a group of students—thus eliminating the watching and listening episode type. They divided the transcript into episodes using the same coding as Artzt and Armour-Thomas, but added a second level of coding by identifying instances where students identified new information or approaches (called a *new idea*) and places where students made an *assessment* of a strategy, result, or knowledge (similar to Schoenfeld's managerial behaviour). These two types were designated as *metacognitive acts* by the researchers. These statements were further characterised as directed toward the speaker (the self), or directed toward another group member. There was no mention made of interrater reliability in this study or their related studies. They found little difference in the total amount of metacognitive activity between successful and unsuccessful student groups. The difference was in the number of statements each group member made to another group member—both metacognitive and cognitive. They conclude that “the discussion *around*, and generated by, individual metacognitive acts is crucial to the success of the mathematical enterprise” (p. 213, italics in original).

In summary, these attempts to quantify the amount of metacognition in the natural-in-action classroom setting have found it useful to characterise students' activity as belonging to different types of episodes and then identify specific metacognitive statements occurring during these episodes. We have used this general method in this study. However, to do so, specific details of the episodes and metacognitive statements must be modified to extend beyond the context of mathematical problem solving to the student laboratory context. In the next section, we describe how we adapt the above methods to develop a new coding scheme.

The development of a coding scheme for the student laboratory

There are several important differences between the mathematical problem solving context and the student laboratory context that can impinge upon the success of a coding scheme. These differences influence the method for coding episode type and metacognitive statements. We begin with the simpler differences and end with the more complicated differences.

Episode types

First of all, because of the collaborative nature of the student laboratory activity, with individual group members interrupting and building on each other's ideas, an individual rarely performs alone or fully completes his or her reasoning or statement. An individual's

activity can seldom be determined without considering the behaviour of the rest of the group. This led us to code the majority group behaviour, making the episode-type watching and listening irrelevant, similar to Goos et al. (2002).

In addition, laboratory activities typically take a longer time than mathematical problem solving activities—particularly those videotaped for the mentioned studies. This, and the fact that significant time in the laboratory might be spent on low-level *automated* data-gathering, makes it likely that a group of students will go *off-topic*. Students were observed discussing their eye infection, the scandalous behaviour of a friend, the repair occurring on their automobile, and many other topics not related to the physics laboratory. This required another type of episode for laboratory coding, one frequently found in the analysis of face-to-face or computer-supported interactive learning, that is, *off-task* (e.g. Clark and Sampson 2007; van Boxtel et al. 2000).

Furthermore, students working in the laboratory perform different activities than students solving mathematical problems, for example gathering data with laboratory equipment that may or may not function. This means that students in the laboratory devote significantly more time to physical activity (pushing buttons, moving equipment, etc) than to problem-solving, compared to students in a pure problem-solving environment. Coding all such activity as implementation would lead to simple manipulation (an automated “push the button, write down the number sequence) being classified identically to more sophisticated troubleshooting. However, since we are interested in students’ ways of thinking, simply carrying out a planned calculation and carrying-out a data measurement plan should be classified as similar activities, as opposed to discussing the benefit of a calculation or measurement.

Laboratory behaviour often involves several activities occurring simultaneously. Consider the following transcript from a group of 4 students measuring the acceleration of two identical-appearing cylinders, labelled 9 and 11, as they roll down a ramp. The transcript begins about 40 min from the start of the laboratory, when the students have already timed how long it takes each cylinder to roll 1 m starting from rest, five trials each. From these times they determined that the accelerations were the same, and thus that the two cylinders were identical inside as well as outside. Student A stands up to return the cylinders. (Clarifying comments are set in square brackets.)

- C: Wait, we have to race them first.
- A: We have to race them?
- C: Are we going to?
- D: Yeah, race them now. [They set the two cylinders side by side and let them roll simultaneously down the ramp]
- A: Don’t let them touch.
- C: Oh, [cylinder] nine wins!
- A: Let’s do it one more time.
- C: Yeah, we’re going to do it five times.
- B: That doesn’t make sense. The average is –
- D: Yeah it does, because the times are so small between them.
- C: I’m going to be crazy and say nine is going to win all five times. [sarcastic prediction]
- A: By a little. [sarcasm]
- C: If we measured how much one wins by, do we win anything? Can we figure anything out? [They finish the 5 races.] So nine won all five times.
- D: The reason we didn’t notice it in timing is because the differences are so slow [small].
- A: Nine is heavier [holding a can in each hand and shaking them]

- C: No, if you switch it, if you switch hands, it doesn't [feel heavier]. They definitely are different. Why would they, if they were all the same then don't you think someone wouldn't have wasted their money and they would have bought all the same cylinders? But they added some, like they made these, so they're different without a doubt. They're different. Shut up, you're wrong.
- D: Every single number [cylinder] is the same.

During this 3 min transcript, the students take some data (race the cylinders and shake the cylinders), interpret the data, compare their race data interpretation to previous results, suggest other data that might be useful, and consider the effort made by the laboratory designer. Is this behaviour analysis, exploration, or implementation? It is very difficult to classify such mixed behaviour as precisely as the described 6-episode coding scheme requires. However, it is possible to distinguish between a fewer number of more inclusive categories. We thus decided to merge reading, understanding, analysis, exploration, planning, implementation, and verification into two episodes: *logistical*, and *sense-making*. With these wider episode types we felt it became more appropriate to call them modes.

The *logistical* mode includes data taking, writing, reading, gathering equipment, reporting to the teaching assistant, and listening to the teaching assistant—activities that must be accomplished through the course of the laboratory, but that do not involve the students explicitly puzzling through or discussing an issue. In the problem-solving research, this would typically involve the episode-types reading, analysis, and implementation. The *sense-making* mode involves discussions about physics formulas or concepts, the group's experimental design, their data, the laboratory question, etc. Typically, during sense-making students are responding to each other, making progress toward an answer, and holding a coherent conversation. The problem-solving episode-types exploration, planning, and assessment are often involved in sense-making. A group was considered to be in a certain mode when the majority of the students appeared to be participating in the behaviour. Of course, this requires that at least 2–3 conversation turns have taken place in this mode, depending on group size.

In summary, we argue that the activity of students in the student laboratory is different from mathematical problem solving in four ways critical to the development of a coding scheme:

- The laboratory involves the manipulation of both physical objects and ideas, including activities varying from now-automatic data taking to troubleshooting to designing.
- The laboratory involves several different activities occurring simultaneously.
- The laboratory is likely to involve students spending some time not discussing the actual laboratory exercise.
- The laboratory activity is collaborative and can rarely be attributed to a single person.

These differences led to the new 3-mode coding scheme consisting of the modes *off-task*, *logistics*, and *sense-making*.

Metacognitive statements

During any of the three modes students might make an overt metacognitive statement, similar to Goos et al.'s (2002) *assessment statements* and Schoenfeld's (1985) *managerial behaviour*. This study is directed towards quantifying the amount of metacognition in the student laboratory by coding students' natural-in-action verbalizations – thus we are limited

to coding specific evidences of students applying metacognitive skills or knowledge. For this setting, the definition for metacognition that we found most useful was based upon Georgiades (2004, p. 371): a “critical revisiting” of one’s cognition, requiring “an element of judgement.” Using the description of metacognition found in Pintrich et al. (2000) we can say more specifically that these metacognitive verbalizations contain students’ metacognitive judgements and monitoring: “individuals becoming aware that they do not understand something they just read or heard or becoming aware that they are reading too quickly or slowly given the text and their goals” (Pintrich et al. 2000, p. 48–9).

As an example, consider the statement “I don’t get this” frequently heard from students throughout the laboratory exercises. Such students are revisiting their current thinking, and admitting that they do not understand. Or consider a student who just measured a distance and a force to calculate the work done on a object, and then exclaims “Whoa, I got a huge number”. In this statement, the student is judging her result, and considering that perhaps she has done something wrong. Another example, “That might not be right, though” comes after a student has explained how he changed the units on three measurements simultaneously. Again, this shows evidence that the student is monitoring reasoning—performing judgement on cognition. Statements coded as metacognitive included the evaluation of reasoning about plans, measurements, actions, data analysis, or data interpretation.

Research context

The purpose of this study becomes now to determine whether this three-mode coding is useful for looking at student learning in university-level student laboratories in physics. To investigate this question fully, we have studied a variety of student laboratory designs. The research occurred at a large state university in the United States, where the physics department was in the process of changing their introductory laboratory (taught parallel to a shared lecture component). This allowed the unique opportunity to simultaneously examine 3 different types of 2-h laboratories. In the first type, the typical *cookbook* laboratory, where students are given detailed instructions to follow, two groups of two students each were analysed. This was the same for the second type, the *cookbook + explanations* laboratory, where ‘explain your reasoning’ questions were added to the instructions. The third type was a more open-ended laboratory, where students were given only a question to answer, and defended their method and results to the other groups in the class. This design attempted to mimic the *scientific community* in which research is performed. In this type four groups of three or four students each were analysed. These variations in amount of guiding and size of group were chosen to provide a wider sample for investigating metacognition.

All 8 laboratory activities were videotaped, transcribed fully, and then coded using our scheme. First, the groups’ behaviour was coded into the three modes of off-task, logistics, and sense-making. Then, specific metacognitive statements were identified. The results were then analysed to determine first the reliability of the coding scheme, and then, whether results from the scheme were useful and/or interesting as related to research on learning.

Coding scheme reliability

For this coding scheme to be useable, it must have a reasonable inter-rater agreement. Three people individually analysed each laboratory, using the transcript and video, and then

discussed their coding to determine the final results. Before discussion, inter-rater agreement for the 3 modes ranged from 77 to 94%, with an average of 86%. Inter-rater agreement for the metacognitive statements ranged from 83 to 94% (before discussion), calculating the ratio of agreed coding (metacognitive and non-metacognitive) to total statements.² This level is well above 80%, an established lower bound of acceptability (Gardner 1995; Bakeman et al. 1997). If we compare these levels to the reported agreement for the previously discussed mathematical studies, this coding scheme appears to have a similar level of agreement.

The agreement for metacognitive statements is somewhat misleading, however, because of the large number of non-metacognitive statements in each laboratory. For this situation, some claim that Cohen's Kappa, which shows the amount of agreement after correcting for chance alone, is more appropriate (Hayes and Hatch 1999). Cohen's Kappa for the metacognitive coding ranged from 0.29 to 0.40 with an average of 0.33. This has been interpreted as a fair level of agreement (Landis and Koch 1977). However, the interpretation of kappa values "depends on the number of codes and, especially when the codes are few, on the variability of their simple probabilities" (Bakeman et al. 1997, p. 360). In this case, there are only 2 codes and their simple probabilities are highly variable, making this an extremely difficult case to interpret.

As discussed earlier, distinguishing cognitive statements from metacognitive statements can be quite difficult. Consider whether the following is metacognitive or not.

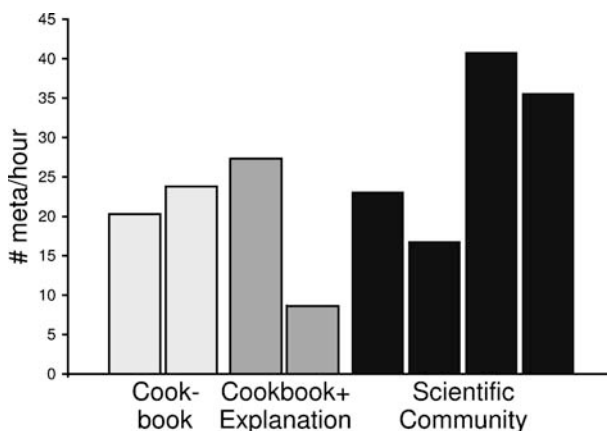
- A: Do any of you have a watch with a stopper [stopwatch timer] on it?
 B: They have timing things there.
 A: The timers here are really bad. Like, they just don't work. You click on them, they just do whatever they want.
 C: You can use this if you prefer.

Student A asks whether anyone has a digital wristwatch with a timer feature. Student B responds that stopwatches are supplied equipment. Student A replies that the supplied stopwatches do not work. This statement could be interpreted as a cognitive statement about the functionality of the stopwatch—evaluating an object, not a thought. Or it could be interpreted as a metacognitive statement, evaluating student B's suggestion that they use the lab's stopwatches. Other statements, such as "Those are really great numbers, but I don't think they're right because it was kind of, last time it hadn't really started rolling yet" are more clearly metacognitive.

Further difficulty in coding metacognition comes from determining whether a statement such as "I don't know" is a figure of speech, a social negotiation, or truly an evaluation of the speaker's knowledge. We argue that even if the initial cause of the statement is a figure of speech or social pressure, it is likely that the speaker is (at some level) doubting their knowledge, for they would not say "I don't know" when stating their name. Because of all of these issues, only those statements originally identified as metacognition by at least 2 of the 3 coders are identified as metacognition in the final results. Cohen's kappa for comparing each individual coder to the final results ranged from 0.44 to 0.76, with an average of 0.59. This may be interpreted as a moderate to substantial agreement (Landis and Koch 1977).

² For example, one laboratory activity had a total number of 761 lines, and both coders agreed on 721 lines (24 lines as metacognitive, 697 lines as non-metacognitive), and disagreed about 64 lines. This gives a percent agreement of $721/761=94\%$.

Fig. 1 The number of metacognitive statements per hour for each laboratory activity



Though the coding scheme is not as reliable as those used in the mathematical problem solving studies, we would argue that it still has a reasonable level of reliability for the purposes of our study. So now we attempt to use this scheme to investigate learning in the laboratory.

Possible coding scheme results

There are several questions which this scheme could be used to investigate. The most simple would be to look at the quantity of metacognition and explore how this relates to the laboratory design. It would also be possible to investigate how the groups in different laboratories responded to metacognitive statements—what their behaviour was before and after such statements, what kind of effect the statement had, and whether there appears to be different types of metacognition in use.

Consider first how the quantity of metacognition relates to the laboratory design. The results of the 3-mode coding show that the quantity of metacognition varies more within each type of laboratory than between the types (see Fig. 1). One could conjecture that the amount of metacognition is more dependent on the particular group members, their comfort working together, and habits of speech and thought, than on the type of laboratory. Of course, we must be careful making conclusions based on this data, since the study has no control—different students performing different laboratory exercises associated with different professors were analysed. However, the point is that such a study would be possible.

From the data, one can see that students in all types of laboratories frequently made metacognitive comments. It would also be interesting to consider these comments more closely, and to look for different types or for the results of such comments. Consider the following transcript, where the students are constantly evaluating their own understanding (shown as underlined text).

B: Ok. I really don't know how this is going to work.

A: Oohh, we're supposed to. [reads manual]

B: We have to weigh it, though. [B uses the scale, A reads the lab manual]

A: How much did it weigh?

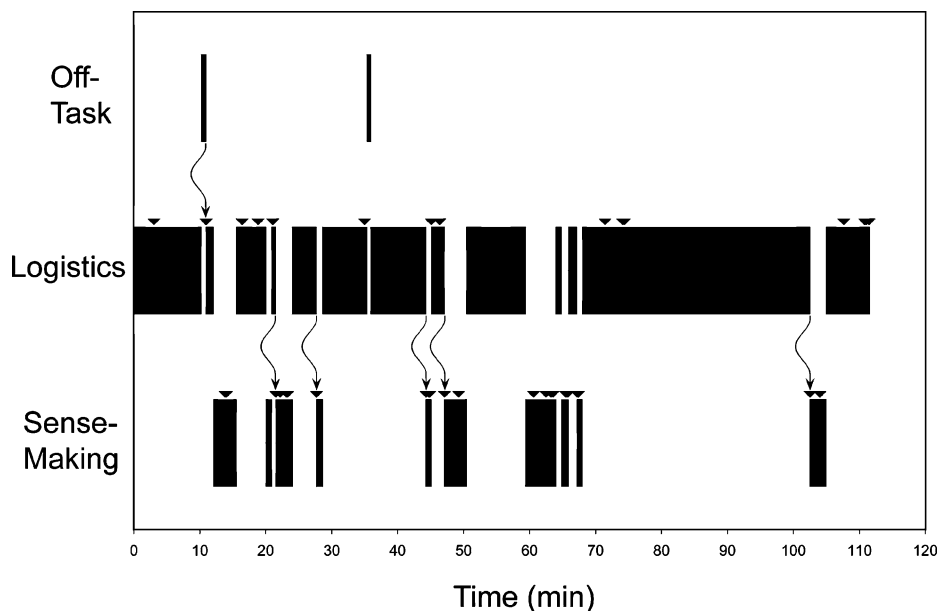


Fig. 2 The time-line for a scientific community lab. Metacognitive statements are marked by upside down triangles, and transitioning metacognition is marked by a *curvy arrow*

B: One hundred. I think. I don't know.

A: And the nylon cord?

B: Oh. I was wondering do we measure it with or without the metal thing? With?

A: I think. I don't know...How are we supposed to fill it [the calorimeter]?

B: Oh, we fill that thing? [the calorimeter]

A: I think. [asks the laboratory teaching assistant] We fill this with water?

TA: Yeah

A: I told you so [to B].

B: Well, I don't know.

[They fill the calorimeter with water, reweigh it, and then put the thermometer and top on it]

B: Ours is 19.5 [Celsius reading from thermometer]. I don't know why it's so warm.

A: We need to make it warmer.

B: Don't we need to record it before we put it in?

A: Yeah. No. We need to record what?

B: I have no idea. Ok, this thing [the calorimeter top] is nice and tight now. Ok, we're supposed to wrap it [the rope] around somehow?

While repeatedly asserting that they know nothing about what they are doing, the two students continue working, stopping a few times to read the laboratory manual and once asking the teaching assistant. At no time do they stop and attempt to gain an understanding of their actions and the reasoning behind such actions; they just continue attempting to follow the instructions. It appears that increasing or decreasing the number of metacognitive statements would have no affect on their behaviour or on their learning. This suggests that having more metacognition is not a critical differentiator—the metacognitive statements' effect on student behaviour is more important. This result agrees with that of Goos et al. (2002)—what matters is the activity evoked by the metacognition.

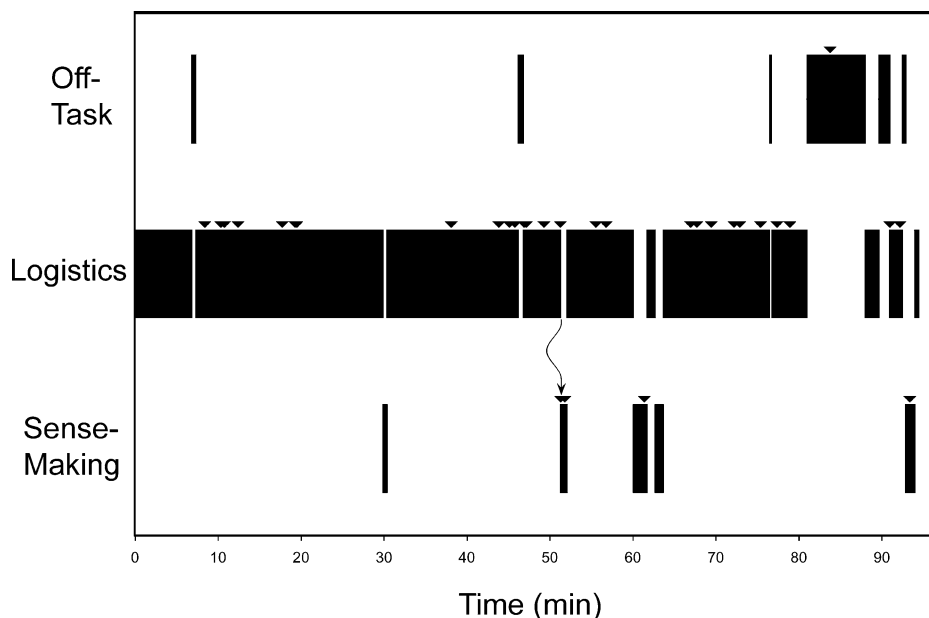


Fig. 3 The time-line for a cookbook lab. Only one metacognitive statement is transitioning

Transitioning metacognition

Using this coding scheme, it is possible to investigate how the groups in different laboratories responded to metacognitive statements. A given group could continue operating in a single mode of behaviour, as in the previous cookbook lab example, or they could switch into another mode of behaviour. By looking at a timeline of a group's behaviour mode and metacognitive statements, one can see where such transitions could occur. For example, the comment "You know, those are really great numbers, but I don't think they're right because it was, kind of, last time it hadn't really started rolling yet" caused students to move from taking data (logistical) to discussing how to improve their method (sense-making). This *transitioning* metacognition occurred at 47 min in the scientific community lab shown in Fig. 2. This and other transitioning metacognition is marked by arrows. For each of these transitions, a specific metacognitive statement was made that caused the group to change to a deeper level of behaviour.

Only one transitioning metacognitive statement occurred in the two analysed cookbook labs. At 51 min in the cookbook lab quoted earlier (see Fig. 3), the students discuss how to change the units of the result of multiplying a distance by a mass.

- A: Centimetres to metres is 2 decimal places, and grams to kilograms is three, so I divided this by one hundred thousand.
- B: That might not be right, though.
- A: Why?
- B: Because of the, I don't know.
- A: Let's try it again. [recalculates with converted numbers instead of converting final results] Huh!
- B: Is it the same?
- A: [nods yes]

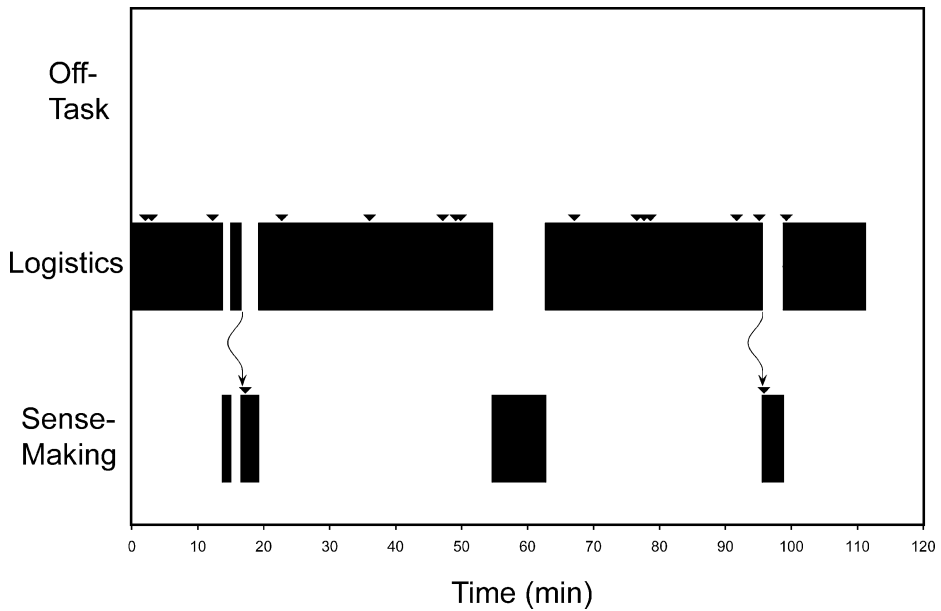


Fig. 4 The time-line for a cookbook + explanation lab. The two marked transitioning metacognitive statements correspond to questions in the instructions asking for explanation

B: Weird.

A: Why weird? You just divided by a...

For these students, the mathematical issue of changing units before multiplying or after multiplying should not be a difficult one, and perhaps this allows them to feel that sense-making would be possible and productive in this case.

Similar transitioning metacognition occurred in the *cookbook + explanation* labs, however, the transitioning metacognition precisely coincided with the ‘explain’ questions prompted in the laboratory handout. For example, one question in the laboratory handout asked students to look at the pattern formed when a laser beam passes through 2 slits and 3 slits, then predict the pattern for 4 and 5 slits, and explain why this happens. Around 95 min, the two students make their prediction and then ask why (see Fig. 4).

B: The pattern may become brighter as you increase the number of slits. Why? We know that the interference—

A: Because the pattern is being reinforced more by more light getting in. So there’s more light but the pattern doesn’t change. So it can be brighter but the same pattern.

B: Ok. What does that have to do with interference?

A: Because the interference is what makes the pattern...I guess you could have the same thing with waves in water. Like say you have two waves and you go like this...

Although initiated by the handout, asking “Why?” and “What does that have to do with interference?” are metacognitive questions that in this case cause the students to begin making sense by using water waves as a model. In the *cookbook + explanation* labs, the students appear to require a specific, directed question in the handout as a stimulus to sense-making.

Non-transitioning metacognition

In the cookbook + explanation labs, some handout questions fail to initiate sense-making. An earlier question asked students to look at the pattern formed when a hair is placed in a laser beam. The students were also asked to “explain why this happens”. In this case, at 47 min one student says “I have no idea why that happens. I don’t know how hairs work, you know?” and the students abandon the question, moving on to some unrelated calculations. Two and a half minutes later, they return briefly to the hair question, but again give up. “Alright, now let’s make up about the hair thing. Obviously there’s some kind of slits in hair. I don’t know.” About 15 min later, they ask the laboratory assistant for an explanation about the hair. In this case, an ‘explain this’ question initiated by the handout did not initiate any sense-making. Perhaps to these students this was a harder question, and they felt they could not arrive at a reasonable answer without external aid.³

In the scientific community laboratory, a metacognitive statement that failed to change modes occurred at 74 min. One student was measuring the volume of a cylinder, and said it was “724.8 centimetres squared”. A second student corrected “centimetres cubed”. The first responded “You’re right. I wasn’t paying attention.” Such an elementary correction and corresponding acknowledgement does not require a switch into a sense-making mode.

Discussion

In this study we attempt to show that investigating the natural-in-action, observable metacognition occurring during a physics student laboratory can be useful for studying learning. There are many challenges involved in this type of research, many of them likely reasons why this research has not been done before. Our main argument is that this research is possible and potentially extremely useful for informing both the development of teaching and learning in the laboratory and for further research. However, some problems remain only partly solved. One main problem is the issue of reliability.

Based on research in a related area – the natural-in-action, observable metacognition occurring when students work together on mathematical problem solving – we developed a coding scheme for classifying students’ behaviour during the student laboratory into three modes and quantifying their metacognitive activity. For this scheme to be usable, it must be reliable. This is one significant challenge with this type of research—reliably identifying instances of metacognition. Whether a statement is simply cognition or is metacognition is not straightforward to determine. As described by Flavell (1979, p. 909) “Cognitive strategies are invoked to *make* cognitive progress, metacognitive strategies to *monitor* it. However, it is possible in some cases for the same strategy to be invoked for either purpose and also, regardless of why it was invoked, for it to achieve both goals.” As one way of coping with this problem, we limited the study to explicitly verbalised metacognitive assessment—such statements were easier to reliably identify. However, even with this limitation, the inter-rater reliability could be considered to be relatively low.

One might think that requiring verbalization of metacognition imposes a significant limitation on the research (Gavelek and Raphael 1985). However, this research aims at

³ Another interpretation could be that the students judged a question about hair not worth the effort, compared to the more mainstream physics questions about slit number and slit width. However, this interpretation appears to be less likely, because their written answer to the hair question in their report is as full and detailed as their answers to similar questions about slit number and width.

informing the development of teaching and learning in situations where students work in groups, and groupwork depends on verbal communication. For this situation, as opposed to research aimed at developing metacognitive theory, we would argue that requiring verbalization is not a significant limitation. In a group interaction, metacognition that is never verbalized is unlikely to affect the rest of the group.

To further improve the reliability, we required that a metacognitive statement be identified by at least 2 out of 3 coders. This increases significantly the amount of time it takes to code a transcript; however, it also increases the reliability from an average of 0.33 to an average of 0.59. This final reliability is at a reasonable level, but it would be appropriate for future studies using this coding scheme to aim to improve the reliability. Because of this design, metacognitive statements are more likely to be unidentified than to be identified where they should not—that is, false-negatives are more likely than false-positives. This means conclusions about the number of metacognitive statements cannot be strongly made; however, conclusions about the type or usefulness of metacognitive statements are less affected by the low reliability.

The results from this coding scheme allowed us to argue that metacognitive statements, alone, are not lacking in quantity. Students are well able to recognize when something “feels wrong” or they do not understand. More critical is *how* students react to this metacognition. Changing the goal from increasing the amount of metacognition to increasing the reaction to existing metacognition has significant pedagogical implications. The laboratory should be designed so that transitions into sense-making will be productive and required, and so that the students recognise this.

Because of limitations in this study’s method, we are not able to conclude that the different laboratory designs led to different amounts of transitioning behaviour, though such a study would be possible with this type of coding scheme. We do, however, believe that the process of designing their own method and defending this method to their classmates in a verbal discussion convinced students that they could not satisfactorily complete the laboratory exercise without sense-making. By altering the information given to the students, the design of the laboratory activity, and the final product of the activity, the students can be encouraged to transition more frequently into sense-making. These, and other modifications all contribute to the framing of the laboratory class, that is, a person’s sense of what is going on (MacLachlan and Reid 1994). Research building upon this study could aim to investigate the relationship between particular laboratory design elements and the amount of transitioning metacognition. We would argue that such research would likely corroborate other studies showing that the framing of an activity significantly impacts behaviour in that activity (for example, Hammer et al. 2004; or Fong and Woodruff 2003).

Unfortunately, any study attempting to compare different laboratory designs using a well-designed, controlled study would need to contend with significant challenges. Communicating to students the expectations and goals of a particular laboratory design can be quite challenging, particularly when the designs are not traditional (Kung 2005). In our experience, attempting to change those expectations and goals by introducing the students to a different laboratory design can lead to students either failing to change and treating the new design like the old, or students may become completely disillusioned with the process and revert back to treating the laboratory like a traditional course. Either of these results would weaken the conclusions of such a study. A study attempting to appropriately sample different groups of students in different laboratory designs would require such numbers of students as to make such a study extremely time-intensive, though possible. However, even though research on quantifying the metacognition in a natural-in-

action student laboratory setting is extremely time-intensive and challenging, we have attempted to show that it can have interesting and useful results.

Conclusion

In this paper we have attempted to show that it is possible and valuable to use metacognition to examine student learning in the university-level student laboratory for science subjects such as physics. The implications of this study apply to the development of metacognitive research, and suggest a way to evaluate the design of a student laboratory. A coding scheme was developed for the student laboratory that codes the behaviour of a group of students as sense-making, logistics, or off-task and also identifies metacognition. The scheme was used to analyse three types of laboratory classroom designs varying in the amount of information given to the students and in open-endedness. The results of this scheme showed that it is important to consider the outcome of metacognition and not only the amount of metacognition.

Acknowledgements We thank Anna T. Danielsson, Edward F. Redish, and Rachel E. Scherr for substantive discussions of this research. This work was funded in part by the Swedish Research Council and by the US National Science Foundation grant REC-008 7519.

References

- Artzt, A. F., & Armour-Thomas, E. (1992). Development of a cognitive–metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition and Instruction*, 9(2), 137–175.
- Bakeman, R., McArthur, D., Quera, V., & Robinson, B. F. (1997). Detecting sequential patterns and determining their reliability with fallible observers. *Psychological Methods*, 2(4), 357–370.
- Borkowski, J. G., Chan, L. K. S., & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 1–41). Lincoln: Buros Institute of Mental Measurements.
- Butterfield, E. G., Albertson, L. R., & Johnston, J. C. (1995). On making cognitive theory more general and developmentally pertinent. In F. E. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 181–205). Mahwah: Lawrence Erlbaum.
- Clark, D. B., & Sampson, V. D. (2007). Personally-seeded discussions to scaffold online argumentation. *International Journal of Science Education*, 29(3), 253–277.
- Fong, C., & Woodruff, E. (2003). Web-based video and frame theory in the professional development of teachers: some implications for distance education. *Distance Education*, 24(2), 195–211.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911.
- Gardner, W. (1995). On the reliability of sequential data: Measurement, meaning, and correction. In J. M. Gottman (Ed.), *The analysis of change* (pp. 339–359). Mahwah: Erlbaum.
- Gavelek, J. R., & Raphael, T. E. (1985). Metacognition, instruction, and the role of questioning activities. In D. L. Forrest-Pressley, G. E. MacKinnon & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (vol. 2, pp. 103–136). Orlando: Academic Press.
- Georghiades, P. (2004). From the general to the situated: three decades of metacognition. *International Journal of Science Education*, 26(3), 365–385.
- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational Studies in Mathematics*, 49, 193–223.
- Hammer, D., Elby, A., Scherr, R., & Redish, E. F. (2004). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning: Research and perspectives*. Greenwich: Information Age Publishing.
- Hayes, J. R., & Hatch, J. A. (1999). Issues in measuring reliability. *Written Communication*, 16(3), 354–367.

- Hofstein, A., & Lunetta, V. N. (1982). The role of the laboratory in science teaching: Neglected aspects of research. *Review of Educational Research*, 52, 201–217.
- Hofstein, A., & Lunetta, V. N. (2004). The laboratory in science education: Foundations for the twenty-first century. *Science Education*, 88(1), 28–54.
- Koch, A. (2001). Training in metacognition and comprehension of physics texts. *Science Education*, 85(6), 758–768.
- Kramarski, B., & Zeichner, O. (2001). Using technology to enhance mathematical reasoning: effects of feedback and self-regulation learning. *Educational Media International*, 38(2–3), 77.
- Kung, R. L. (2005). Do students know the purpose of the laboratory? Presentation at the American Association of Physics Teachers Summer National Meeting, Salt Lake City.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.
- Lazarowitz, R., & Tamir, P. (1994). Research on using laboratory instruction in science. In D. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 94–128). New York: Macmillan.
- MacLachlan, G., & Reid, I. (1994). *Framing and Interpretation*. Portland, Or.: Melbourne University Press.
- Mevarech, Z., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and Learning*, 1(1), 85.
- Pintrich, P. R., Wolters, C., & Baxter, G. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Lincoln: Buros Institute of Mental Measurements.
- Pressley, M. (2000). Development of grounded theories of complex cognitive processing: Exhaustive within and between study analyses of think-aloud data. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 262–296). Lincoln: Buros Institute of Mental Measurements.
- Schoenfeld, A.H. (1985). *Mathematical problem solving*. New York: Academic Press.
- Thiede, K. W., Anderson, M. C. M., & Theriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, 95(1), 66–73.
- Trumper, R. (2003). The physics laboratory—a historical overview and future perspectives. *Science & Education*, 12, 645–670.
- van Boxtel, C., van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, 10, 311–330.
- Veenman, M. V. J., Prins, F. J., & Verheij, J. (2003). Learning styles: Self-reports versus thinking-aloud measures. *British Journal of Educational Psychology*, 73, 357–372.
- Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, 8(4), 327–354.
- Zion, M., Michalski, T., & Mevarech, Z. R. (2005). The effects of meta-cognitive instruction embedded within asynchronous learning network on scientific inquiry skills. *International Journal of Science Education*, 27(8), 957–983.