Dependence between creative and non-creative mathematical reasoning in national physics tests

HELENA JOHANSSON

It is known from previous studies that a focus on rote learning and procedural mathematical reasoning hamper students' learning of mathematics. Since mathematics is an integral part of physics, it is assumed that mathematical reasoning also influences students' success in physics. This paper aims to study how students' ability to reason mathematically affects their success on different kinds of physics tasks. A descriptive statistical approach is adopted, which compares the ratio between conditional and unconditional probability to solve physics tasks requiring different kinds of mathematical reasoning. Tasks from eight Swedish national physics tests for upper secondary school, serve as a basis for the analysis. The result shows that if students succeed on tasks requiring creative mathematical reasoning, the probability to solve the other tasks on the same test increases. This increase is higher than if the students succeed on tasks not requiring creative mathematical reasoning. This result suggests that if students can reason mathematically creatively, they have the ability to use their knowledge in other novel situations and thus become more successful on tests.

Many scholars discuss the importance to understand how mathematics is used in physics and how students' mathematical knowledge affects their learning of physics (e.g. Basson, 2002; Bing, 2008; Nguyen & Meltzer, 2003; Redish & Gupta, 2009). Basson (2002), for example, mentions how difficulties in learning physics not only stem from the complexity of the subject but also from insufficient mathematical knowledge. diSessa (1993) notices how students, who have studied physics, can solve a quantitative task in physics and still give an inconsistent qualitative analysis of the same task. A *quantitative* task refers to when the task is posed in explicitly quantitative terms and the solution can be attained through calculations using appropriate physics laws. A *qualitative* task on the other

Helena Johansson
Mid Sweden University

Johansson, H. (2017). Dependence between creative and non-creative mathematical reasoning in national physics tests. *Nordic Studies in Mathematics Education*, 22 (2), 93–119.

hand refers to when the solution requires an analysis of the posed physical situation i.e. what will occur and/or why. It is shown that students' intuitive understanding of the physical world is quite robust and that their solutions to qualitative problems often contradict the basic physics principles (ibid.). Redish (2003) states that practice, in the meaning that students just solve various tasks, is necessary but not enough to get a deeper understanding of the underlying physics concepts. Students must learn both *how* to use the knowledge and *when* to use it. The same conclusion holds for learning mathematics, shown by e.g. Schoenfeld (1985) in his study of how students become good problem solvers in mathematics; as well as by Lesh and Zawojewski (2007), who discuss how working with mathematical modelling develops students' understanding and learning in mathematics. Michelsen (2005) also addresses benefits from modelling activities. He discusses how interdisciplinary modelling activities can help students to understand how to use mathematics in physics and discover the connections between the two subjects.

During studies of how students are engaging in different mathematical activities, Lithner (2008) has gradually developed a framework for characterising students' mathematical reasoning. The framework distinguishes between *creative mathematical founded reasoning* (CR) and *imitative reasoning* (IR). The former one refers to a reasoning that is anchored in intrinsic mathematical properties and that includes some novelty to the reasoner. If instead the anchoring is in surface properties and the reasoning consists of remembering an answer or following a process step by step, it is IR. Mathematical reasoning is one aspect of mathematical knowledge, and thus assumed to be one competence that influences students' learning of physics. Johansson (2015) shows e.g. that to pass Swedish national physics tests, students have to reason mathematically; and to fully master the Swedish physics curricula students have to be able to use CR. To examine further how students' ability to reason mathematically influences how they succeed in physics, the aim in this study is to analyse if there are any dependencies between students' success on different physics tasks, different with respect to which type of mathematical reasoning that is required to solve the tasks.

Conceptual framework

Lithner define *reasoning* as "the line of thought adopted to produce assertions and reach conclusions in task solving" (Lithner, 2008, p. 257). *Mathematical reasoning* is used as an extension of a strict mathematical proof to justify a solution and is seen as a product of separate reasoning sequences. Each sequence includes a choice that defines the next sequence, and the reasoning is the justification for the choice that is made. The mathematical foundation of the reasoning can either be *superficial* or *intrinsic*. The accepted mathematical properties of an object are of different relevance in different situations. This leads to a distinction between *surface* properties and *intrinsic* properties, where the former have little relevance in the actual context and lead to superficial reasoning and the latter are central and have to be taken into consideration in the given context (Lithner, 2008). As mentioned in the introduction, this framework was developed during empirical studies of how students engage in various mathematical activities. A strength of the framework is that it is not restricted to any specific context. As long as the students have to use some kind of mathematics to come up with a solution, they are assumed to reason mathematically. Therefore, Lithner's framework is considered suitable for categorising the kinds of mathematical reasoning that are required in the physics tests.

Creative mathematically founded reasoning

Creativity is an expression often used in different contexts and without an unequivocal definition (for a discussion see Lithner, 2008, p. 267–268). For the definitions of the different kinds of reasoning, the perspective of Haylock (1997) and Silver (1997) is adopted. This implies that creativity is seen as a thinking process that is novel, flexible and fluent (Lithner, 2008). CR fulfils all of the following criteria:

> i. Novelty. A new reasoning sequence is created or a forgotten one is recreated.
>
> ii. Plausibility. There are arguments supporting the strategy choice and/or strategy implementation motivating why the conclusions are true or plausible.
>
> iii. Mathematical foundation. The arguments made during the reasoning process are anchored in the intrinsic mathematical properties of the components involved in the reasoning.
>
> (Lithner, 2008, p. 266).

Imitative reasoning

The arguments that motivate the chosen solution method (i.e. the reasoning) could be anchored in surface mathematical properties. Reasoning categorised as IR fulfils

> i. The strategy choice is founded on recalling a complete answer.
>
> ii. The strategy implementation consists only of writing it down.
>
> (Lithner, 2008, p. 258)

or

> i. The strategy choice is to recall a solution algorithm. The predicted argumentation may be of different kind, but there is no need to create a new solution.
>
> ii. The remaining parts of the strategy implementation are trivial for the reasoned, only a careless mistake can lead to failure.
>
> (Lithner, 2008, p. 259)

Local and global creative mathematical reasoning

Lithner (2008) introduces a refinement of the category CR into *local CR* (LCR) and *global CR* (GCR) that captures some significant differences between tasks categorised as CR. This subdivision has been further elaborated by other scholars, e.g. Boesen, Lithner and Palm (2010) and Palm, Boesen and Lithner (2011). In LCR, the reasoning is mainly IR but contains a minor step that requires CR. If instead there is a need for CR in several steps, it is called GCR.

Non-mathematical reasoning

In the application of the framework, an additional category, defined in Johansson (2015), is used. This category consists of those tasks that can be solved by only using physics knowledge; and this category is called *non-mathematical reasoning* (NMR). Physics knowledge is here referred to as relations and facts that are discussed in the physics courses and not in the courses for mathematics, according to the syllabuses and textbooks, e.g. that angle of incidence equals angle of reflection.

Related concept

There has been a discussion in the mathematical educational research society whether procedural knowledge should be considered only as superficial and rote learned or viewed from a wider perspective (Baroody, Feil & Johnson, 2007; Star, 2007). Hiebert and Lefevre (1986) defined procedural knowledge as consisting of the formal language of mathematics, as well as of the algorithms and rules for completing mathematical tasks. There is an agreement that procedural knowledge is important, but not enough, when students learn mathematics (e.g. Baroody et al., 2007; Gray & Tall, 1994; Hiebert & Lefevre, 1986; Star, 2007). However, there is also an argumentation about whether *deep* procedural knowledge could exist without involvement of conceptual knowledge (Baroody, Feil & Johnson 2007; Star, 2005, 2007). In the description of the framework used for characterising required mathematical reasoning, Lithner (2008) discusses different aspects of procedures and concepts. Although the definitions of the reasoning categories do not include references to procedural or conceptual knowledge, one could assume some relations between CR and conceptual knowledge on one hand and IR and procedural knowledge on the other hand.

Research questions

Based on Lithner's (2008) framework, physics tasks in Swedish national tests have been categorised with respect to mathematical reasoning requirements in Johansson (2015). The main results showed that students must use some kind of mathematical reasoning to solve three-fourth of the tasks in a test; and that one-third of the tasks require CR. From the outcome, one of the interests that arose was if there is a dependence between how students succeed on physics tasks requiring different kinds of mathematical reasoning. To study a possible dependence the following research questions are posed:

- Does the success on a physics task that requires CR affect the probability to succeed on any other task requiring either IR or CR in the same test?
- Does the success on a physics task solvable with IR affect the probability to succeed on any other task requiring either IR or CR in the same test?

The answers to both questions are intuitively yes, but has to be verified in order to answer the following two research questions:

- How strong is the dependence in each case?
- Are there any difference in effects on tasks requiring different mathematical reasoning?

Method

Physics in the Swedish School

There are mainly two different physics courses in the Swedish upper secondary school. Physics A, which is compulsory for all natural science and technology students, and Physics B that is compulsory for natural science students and an optional continuation for technology students. In the current curricula (from 2011) the names of the courses have changed to Physics 1 and Physics 2, and some of the areas previously included in Physics B are now in Physics 1. During the last decades, there has been a gradual change towards a stronger focus on process goals, and they are present in the curriculum from 1994 (Swedish National Agency for Education, 2006). Content goals are complemented with process goals, and teaching in the subject of physics should for example aim at helping students develop knowledge of the concepts, theories, models and working methods of physics. Students should be given opportunities to develop a scientific approach to the surrounding world, as well as to analyse and solve problems through reasoning based on concepts and models. Mathematics is explicitly required when making quantitative descriptions of phenomena and implicitly required when analysing data (Swedish National Agency for Education, 2000). Similarities between the upper secondary syllabuses for physics in Norway and Sweden are identified and discussed in the TIMSS Advanced 2008 report (Lie, Angell & Rohatgi, 2010). It is further discussed by Grønmo and Onstad (2013), how students' mathematical performance in the Nordic countries form a specific *Nordic profile*, distinct from other countries.

National physics tests for both physics courses are provided by the Swedish National Agency for Education through the National Test Bank as an assessment support to accomplish equivalent assessment for upper secondary physics students throughout the country. Most of the tests are classified to not authorised users. There are a few tests open to the public, which for example students can look upon to get an idea of what is expected. After a test is used, students' results are collected and compiled via the National Test Bank.

Data

The data comprise tasks from eight physics tests from the Swedish National Test Bank. The tests are the May 2002, December 2004 and May 2005 tests for the Physics A course and the May 2002, May 2003, May 2005, February 2006, and April 2010 tests for the Physics B course. These tests were chosen because the tasks in the tests already had been categorised according to mathematical reasoning requirements, and that there were available data about students' results on each of the tasks. Student data were used by permission from Department of Applied Educational Science at Umeå University, the department in charge of the National Test Bank in Physics. No names of the students are present, instead each student has got an ID-number, and thus data could be considered anonymous. The number of students for each test varies from 996 to 3666.

There are in total 119 of the 162 physics tasks on the tests that require mathematical reasoning to be solved, and thus included in this study, i.e. no NMR tasks are included.

Statistical method

To decide whether there exists a dependence between success on a particular task R, the *reference task*, and the success on another task X, it was decided to compare the *conditional probability* $P(X=1|R=1)$ to solve X with the *unconditional probability* $P(X=1)$ to solve X. That is, the ratio

$$\frac{P(X=1|R=1)}{P(X=1)} \quad (1)$$

was estimated, where $X = 1$ and $R = 1$ denote that the tasks have been fully solved, respectively. If this ratio is larger than 1, the probability to succeed on the task X is higher if students successfully have solved the task R than if they have not. The probabilities in (1) are estimated by computing the arithmetic means from the available student data for each test. To estimate $P(X=1|R=1)$, the number of students who had solved both X and R were divided by the number of students who had solved R. The probability $P(X=1)$ was estimated by calculating the number of students who had solved X by the total number of students who had taken the test.

In order to decide if the effect of a calculated dependence is large enough to consider, *odds ratio* is used as a measure of the effect-size. Odds ratio is defined as

$$\frac{\dfrac{P(X=1|R=1)}{1-P(X=1|R=1)}}{\dfrac{P(X=1|R=0)}{1-P(X=1|R=0)}} \quad (2)$$

where $P(X=1|R=1)$ is as defined above and $P(X=1|R=0)$ denotes the conditional probability to solve X when R *is not* solved, i.e. students have only partly or not at all solved the task R. The effects are divided into small, medium and large, and associated with the calculated magnitudes of odds ratios as follows; small = 1.5, medium = 3.5 and large = 9.0. These values could be considered as a rule of thumb (e.g. Cohen, 1988; and Hopkins, 2002).

The paired sample *t*-test was used for significance testing of the difference between the means of the dependencies that are calculated when

CR-tasks are used as reference tasks and when IR-tasks are used as reference tasks. In order to decide if a significant difference is to be accounted for, Cohen's *d* is used as an index of the effect size. Effect sizes around 0.2 could be considered small, effect sizes around 0.5 are medium and sizes above 0.8 indicate large effects (Cohen, 1988).

Implementation

For each test, a CR task was first chosen as reference task. The CR tasks were chosen so that they should not be the most difficult ones, i.e. they should not require too many analysing steps to be solved. The decisions were based on the formulation of the assignments and on what solutions that were required. For most of the tests the chosen CR task is a GCR task. On two tests the tasks categorised as GCR were judged to only occur among the most difficult tasks, thus LCR tasks were chosen instead. There was now one CR task on each test and this task was used as R in (1). The ratio was then estimated for every other task that required mathematical reasoning in each test, respectively. Since it is the effect on success with respect to mathematical reasoning that is studied, tasks not requiring mathematical reasoning, i.e. NMR tasks, were excluded from the analysis. To analyse the same ratio (1), but with an IR task as the reference task, it was decided to choose an IR task with a position approximately in the middle of the tests and relatively close to the already chosen CR task. This choice are based on that a task's positon in a test indicates how difficult the task is supposed to be to the students. The ratio (1) was estimated in the same way as above for every task requiring mathematical reasoning in each test, respectively, with the chosen IR tasks as R in (1). Below follows two examples of different tasks used as reference tasks in the Physics A 2002 test. Task 6 (figure 1) has previously been categorised as solvable with IR, and task 12 (figure 2) as requiring LCR. The method for the categorisation is thoroughly described in Johansson (2015). Two examples of the categorisation process are provided in appendix, example A-1 and example A-2.

To account for possible effects due to the positions of the tasks, it was decided to do some more calculations of the ratio (1) by choosing a GCR task that occurs earlier in the test than the IR task previously used as reference task. If there were no such GCR task, an IR task that came later in the test than the previously used CR task should be used as reference task. If there were no such IR task, an IR task positioned as close as possible to the previously used CR task was used as reference task. There are now three different ratios, i.e. with different reference tasks, for every test. To be able to statistically compare the ratios, the measure of effect size

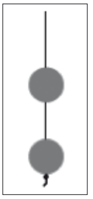
Read the press cutting below.

> Watch out!
> A study of cause of death and injuries among the population in the South Pacific shows that most accidents are caused by falling coconuts and overturned palm trees. This is nothing to laugh about. A four kilos coconut that comes loose from a 25 meter high palm tree, reaches a speed of 80 km/h and hits the ground – or an unfortunately placed head – with a pressure that corresponds to one ton. The study is performed by Doctor Herman Oberli at the hospital in Honiara, Solomon Islands. (TT-DPA)

Is it true that a coconut can reach 80 km/h after a 25 m high fall?

Figure 1. *Task 6*

In order to determine the charge on two small, light silver balls, the following experiment was conducted. The balls, which were alike, weighed 26 mg each. The balls were threaded onto a nylon thread and were charged in a way that gave them equal charges. The upper ball levitated freely a little distance above the other ball. There was no friction between the balls and the nylon thread. The distance between the centres of the balls was measured to 2.9 cm.



What was the charge on each of the balls?

Figure 2. *Task 12*

(2) is calculated for respective ratio. The two calculated ratios (1) with an IR task and a CR task as close as possible were furthermore used in the paired sample *t*-test in order to decide if the means differ significantly.

Analysis and result

In table 1 the calculations for respective task on the Physics A 2002 test are displayed. The estimated values for the ratios in the first row are the results when the CR task, task number 12 (figure 2), is used as reference. Similarly, the estimated values for the ratios in the second row are the results with the IR task, task number 6 (figure 1), used as reference. By comparing the two estimated ratios for each task it was noticed that the ratio in most cases was larger when the CR-task was the reference, compared to when the IR-task was used in the calculations. For example, the estimated values on row 1 and 2 for task number 1a in table 1 are 1.21 and 1.14, respectively, which shows that it is more likely to succeed on task 1a if you solve the CR task than if you solve the IR task.

Table 1. *Ratios according to (1) for the tasks in the Physics A 2002 test with the tasks 12 and 6 as reference tasks*

Task	1a (IR)	1b (IR)	2a (NMR)	2b (IR)	3 (IR)	4 (LCR)	5 (LCR)	**6 (IR)**	7 (NMR)	8 (NMR)
12	1.21	1.47		1.20	1.44	1.08	1.36	1.49		
6	1.14	1.21		1.12	1.19	1.07	1.15	R		

Task	9 (NMR)	10 (LCR)	11 (IR)	**12 (LCR)**	13 (GCR)	14 G (LCR)	14 VG (GCR)	15 (GCR)	16a (GCR)	16b (GCR)
12		1.29	1.87	R	1.27	1.95	4.76	2.28	1.68	2.81
6		1,19	1.24	1.49	1.17	1.31	1.69	1.37	1.22	1.43

Corresponding values for e.g. task 13 in table 1 are 1.27 and 1.17, respectively, which indicates the same result. The corresponding tables containing the estimated values for the ratios for the rest of the physics tests are provided in appendix, table A-1 to table A-7. Furthermore, by comparing the values in each entry for respective row in table 1 and in the rest of the tables in appendix, the results indicate that the dependence between success on a specific task and on the rest of the tasks on the test increases the later the tasks are positioned in the tests. This tendency seems to be the same for both IR tasks and CR tasks used as reference tasks. For example, consider the values in the first row in table 1, all values are less than 1.5 for the first 10 tasks, and for task 11 and the following six tasks, only one value is less than 1.5. Similar result holds for the values in the second row, where the values for the first 10 tasks except one, are less than 1.20, and all values are larger than 1.20 for all but one of the rest of the tasks.

To analyse further whether there are any differences in the ratios with respect to mathematical reasoning categories, the mean ratio ($\bar{r}$) were calculated for respective category (IR, LCR and GCR) for each reference task in each test (table 2). The first row for every test in table 2 shows the values when the chosen CR task is used as reference task, and the second row shows the values when the IR task is used as reference task. So the top two rows are the respective means of the values in table 1. The values indicate that the success on other tasks is more dependent on students' success on a CR task than on their success on an IR task.

Table 2. *The mean ratio, for each of the mathematical reasoning categories with respect to the different reference tasks in each test*

	R	$\bar{r}$		
		IR	LCR	GCR
Physics A VT 02	12 (LCR)	1.45	1.43	2.56
	6 (IR)	1.18	1.24	1.38
Physics A HT 04	11 (GCR)	1.42	1.68	2.03
	8a (IR)	1.26	1.39	1.31
Physics A VT 05	10 (GCR)	1.28	1.53	1.45
	8a (IR)	1.02	1.03	1.03
Physics B VT 02	10 (GCR)	1.44	1.84	2.21
	7 (IR)	1.27	1.51	1.59
Physics B VT 03	8 (LCR)	1.27	1.44	1.65
	7 (IR)	1.13	1.20	1.27
Physics B VT 05	12b (GCR)	1.17	1.21	1.22
	8b (IR)	1.25	1.46	1.42
Physics B VT 06	12a (GCR)	1.40	1.73	1.94
	10b (IR)	1.38	1.32	1.29
Physics B VT 10	11b (GCR)	1.39	1.52	2.03
	9b (IR)	1.19	1.24	1.49

The tendency discussed above, that the dependence increase with the tasks position in the test, suggests that position has an effect on the dependence. As described in the Method section, a new reference task with another position was chosen for each test and values for the ratios according to (1) were estimated with this new task as reference. The new values for the Physics A 2002 test are displayed in the last row in table 3. The first two rows in the table are the previously calculated and displayed ratios in table 1, with the two other tasks used as reference tasks. As previously, tables for the other physics tests are available in appendix, table A-8 to table A-13. No new task were chosen for the Physics B 2003 test since the two previously used tasks already were positioned next to each other in the test (table A-4).

Comparing the values in row 1 and 3 for each task in table 3 shows that the ratio in all except one case, was larger when the CR task (task 12) was the reference, compared to when the IR task (task 11) was used in the calculations. Consider for example the values in row 1 and 3 for task 1b, these are 1.47 and 1.35, respectively; and corresponding values for task 15 are 2.28 and 2.08, respectively. Similar results were obtained for the rest of the calculations, which indicate that success on a GCR task, even when

Table 3. *Ratios according to (1) for the tasks in the Physics A 2002 test with the new task 11 together with tasks 12 and 6 as reference tasks*

Task	1a (IR)	1b (IR)	2a (NMR)	2b (IR)	3 (IR)	4 (LCR)	5 (LCR)	**6 (IR)**	7 (NMR)	8 (NMR)
12	1.21	1.47		1.20	1.44	1.08	1.36	1.49		
6	1.14	1.21		1.12	1.19	1.07	1.15	R		
11	1.18	1.35		1.16	1.35	1.09	1.33	1.40		

Task	9 (NMR)	10 (LCR)	**11 (IR)**	**12 (LCR)**	13 (GCR)	14 G (LCR)	14 VG (GCR)	15 (GCR)	16a (GCR)	16b (GCR)
12		1.29	1.87	R	1.27	1.95	4.76	2.28	1.68	2.81
6		1.19	1.24	1.49	1.17	1.31	1.69	1.37	1.22	1.43
11		1.23	R	2.11	1.20	1.77	2.13	2.08	1.45	1.23

account for position in the test is taken, in most cases has a larger effect on the success on the rest of the tasks, than success on an IR task has.

The means of the ratios for respective reasoning categories with respect to the latest reference tasks in each physics tasks are displayed in table 4. Comparing the means with the previous calculated ones (table 2), shows that in six out of eight tests the effect on success is higher for all three reasoning categories when students succeed on a CR task compared to if they have succeeded on an IR task. The means of the estimated ratios for the Physics A 2002 test are for example 1.45 for IR tasks, 1.43 for LCR tasks and 2.56 for GCR tasks when the LCR task (task 12) was used as the reference task (table 2). Corresponding means are 1.29, 1.51 and 1.62, respectively (table 4), when the IR task (task 11) was used as the reference task in equation (1).

The significance testing was performed to test the hypothesis that if students solve a CR task they have a higher chance to solve other physics tasks requiring mathematical reasoning, than if they solve an IR task. The hypothesis was tested on the pair of calculated values for the conditional probabilities (1) with the CR task and the IR task as close as possible as reference tasks, or with a CR task occurring earlier in the test than the used IR task. For the Physics A 2002 test this means that it is the values in the first and the last rows for every task (not used as R) that are included (table 3). Since the difference is assumed to be general and not restricted to specific tasks or tests, the *t*-test was performed on all physics tests together. In total 103 pairs of values were included. The result showed that there is a significant difference of the means of the conditional probabilities with CR as R and with IR as R ($\bar{x}_D = 0.15$, $p = 0.000017 < 0.05$). The effect of the significant difference, $d = 0.41$, is considered to be around 0.5 and thus in the lower range of medium.

Table 4. *The mean ratio for each of the mathematical reasoning categories with respect to the added reference tasks in each test*

	R	$\bar{r}$		
		IR	LCR	GCR
Physics A VT 02	11 (IR)	1.29	1.51	1.62
Physics A HT 04	7a (GCR)	1.19	1.31	1.43
Physics A VT 05	12a (IR)	1.19	1.37	1.33
Physics B VT 02	9b (IR)	1.39	1.67	1.97
Physics B VT 05	7 (GCR)	1.27	1.45	1.42
Physics B VT 06	4 (GCR)	1.50	1.45	1.65
Physics B VT 10	7b (GCR)	1.30	1.31	1.67

As described in Method, odds ratio is used as a measure of effect size of the separately calculated dependencies. Estimations of the odds ratio were calculated according to (2) for all tasks in every physics test with the three different reference tasks, respectively. The values of effect sizes for the tasks in the Physics A 2002 test are displayed in table 5, and the effect sizes for the rest of the physics tests are provided in appendix, table A-14 to table A-20. The effect sizes that are interesting to compare for each task are at first-hand the two that are calculated with reference tasks as close as possible to each other, or when the reference task is a CR task occurring earlier in the test than the used IR task. Then, as described above, possible influences of the tasks position in the tests are considered. Thus, for the Physics A 2002 test, the values in the first and the last rows are the one most interesting to compare. For example, the value of the effect size for having solved the CR task 12 (row 1, table 5) is 7.13 for task 1a, and 3.95 for task 1b. Both these could be considered medium effects according to the rule of thumb outlined in the Statistical method section, i.e. the value is between 3.5 and 9. It is also noticed that the effect is larger on task 1a than on 1b. Corresponding values of the effect for having solved the IR task 11 (row 3, table 5) are 6.91 for task 1a and 4.04 for task 1b, which also can be considered as a medium effect, and larger for task 1a than for task 1b.

When analysing the values in row 1 and row 3 for every task, it is noticed that there is a medium effect on 8 of the 14 tasks that are included in the analysis and not used as reference tasks, and a large effect on 2 of the 14 tasks. On the rest of the four tasks the effect is small. When only effect sizes that can be considered at least medium is taken into account, and pair-wise compared for each task in the Physics A 2002 test, it is noticed that the effect is larger on five tasks if the students have solved the

Table 5. *Odds ratios according to (2) for the tasks in the Physics A 2002 test with tasks 12, 6 and 11 as reference tasks*

Task	1a (IR)	1b (IR)	2a (NMR)	2b (IR)	3 (IR)	4 (LCR)	5 (LCR)	**6 (IR)**	7 (NMR)	8 (NMR)
12	7.13	3.95		6.76	4.11	1.56	3.16	4.70		
6	5.06	2.84		4.05	2.72	1.92	2.44	R		
11	6.91	4.04		5.18	3.98	1.96	3.47	5.54		

Task	9 (NMR)	10 (LCR)	**11 (IR)**	**12 (LCR)**	13 (GCR)	14 G (LCR)	14 VG (GCR)	15 (GCR)	16a (GCR)	16b (GCR)
12		2.67	6.74	R	2.04	3.09	18.9	3.82	4.75	5.44
6		3.42	4.54	4.70	2.31	3.06		3.21	2.50	3.50
11		2.98	R	8.43	2.14	4.93	10.6	6.77	4.11	11.9

CR task, compared to if they have solved the IR task. This holds for task 1a, 2b, 3, 14VG and 16a (table 5). It is also noticed that the effect is higher on five of the tasks if the student instead solves the IR task compared to the CR task, see task 1b, 6, 14G, 15 and 16b in table 5. The values in the table also shows that the effect in most cases are larger for the five tasks with a larger effect for solving the CR task, than on the five task with a larger effect for solving the IR task.

Compiling the above result with the result from the analysis of the effect sizes for the tasks in the rest of the physics tests shows that there are 50 out of 103 tasks that have a medium effect. Of these 50 tasks the effect is larger on 26 of the tasks if the students have solved the CR task used as reference task compared to if they have solved the IR task (table 6). There are four tasks (of the 103 tasks) that have a large effect; and for three of these four tasks, the effect is larger if students solve the CR task than if they solve the IR task. Furthermore, there are one task that no students manage to solve without also solving the CR task used as

Table 6. *Number of tasks with the different effect sizes*

Effect size	Number of tasks
Small effect	48
Medium effect, larger with CR as reference task	26
Medium effect, larger with IR as reference task	24
Large effect, larger with CR as reference task	3
Large effect, larger with IR as reference task	1
Infinite effect with CR as reference	1

reference task, thus the nominator in (2) is 0 and the odds-ratio turns to infinity (task 14VG, table A-14).

When only effects considered at least medium (and the one with infinity is excluded), there are in total 29 tasks for which the effects are higher when a CR task is used as reference task, and 26 tasks for which the effects are higher when an IR task is used as reference. Comparing this with the previous result, that there is a significant difference between the general effect of solving a CR task and an IR task, shows that although the general effect is large enough to consider, the effects on individual tasks do not differ so much.

Discussion and implications

The outcome of the present study shows that mastering the ability to reasoning mathematically creatively has a positive effect on the success on other physics tasks. It is shown that the effect generally is higher for tasks requiring CR compared to tasks solvable with IR. Going back to the definitions of the reasoning categories, CR tasks require that students can use their mathematical knowledge in novel situations, which in turn implies an intrinsic understanding of the mathematics that is involved. At the same time, when students are able to use their knowledge in novel situations, they have also developed another approach to the task solving process. Their strategy is based on the judgement of plausibility, which means that they analyse the task/assignment and have an idea of plausible conclusions. This ability to reason mathematically creatively is thought to be generalizable to various mathematical areas. IR tasks, on the other hand, could be solved by remembering an algorithm, and no intrinsic understanding of the mathematics is required. Therefore it is reasonable that the effect between success on CR tasks is higher than the effect of success on a CR task and on an IR task. Nevertheless, there is still a positive effect on IR tasks from the success on CR tasks, which suggests that students who have developed the ability to reason mathematically creatively also have a better chance to succeed on tasks of a more procedural character. That no intrinsic understanding is required in order to solve IR tasks does not exclude the possibility that students could have developed some conceptual understanding; and thus, success on IR tasks positively affects the success on CR tasks. At the same time, the only characteristics different IR tasks have in common, at least theoretically, are that they should be possible to solve by remembering an answer or a procedure and implement this. Therefore it is reasonable to expect that the effect on the success on other IR tasks varies depending whether they are solved by similar procedures.

In the analysis, Cohen's *d* was used as a measure of the general effect of the difference in success. This is a recognized effect size for the comparison between means for two different groups. Odds-ratio, was also used as an effect size. Not to compare means, but to determine the association between two variables, which is a common use of odds-ratio. The *t*-test showed significant difference between the successes with respect to the different reasoning types, and Cohen's *d* suggested that the effect of this difference is large enough to consider. Odds-ratios showed that the effects due to the different reasoning types varied quite a lot on the different tasks. Although the effect seemed a bit larger due to CR tasks than due to IR tasks, this was not clearly determined. Further studies of the relation between the various effect sizes and the size of the effect of the dependence are required.

The present analysis of dependence between success on CR tasks and on IR tasks has been conducted on physics tasks, which is a limitation of the study. In order to deepen and generalise the results, continued studies of the dependence should be performed on mathematics tasks. Then, account is taken for possible influence students' understanding of physics has on the result. During the analysis, results indicated that the position of the tasks could influence the dependence of success, and this was accounted for by choosing different reference tasks. It is common in the Swedish test system that tests start with easier tasks and that the difficultness increases with later position. The scores on each task in physics and mathematics tests are labelled to indicate which grades they correspond to. Thus, accounting for the scoring of the tasks may reduce the influence of tasks' difficultness on the result even more.

Another limitation of the study is that it is conducted in a Swedish context. As discussed in the Physics in the Swedish School section, there are alignments with the Norwegian syllabus and a Nordic profile has been identified. Thus the results can be considered interesting to a Scandinavian context. Furthermore, the goals and subject description in the Swedish curriculum are quite rich and highly in accordance with the TIMSS Assessment framework (Garden et al. 2006; Swedish National Agency for Education, 2009). This suggests that the results also are relevant to an international context.

As discussed, there are additional factors to consider in the analysis of the dependence between success on CR tasks and on IR tasks and further studies should be performed in order to make general conclusions. The present results though, give reliable indications of the positive effect of creative mathematical reasoning on task solving. These results might contribute to the discussion about the effect mathematical reasoning has on students' development of knowledge of mathematics as well as of physics.

References

Baroody, A. J., Feil, Y. & Johnson, A. R. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, 38 (2), 115–131. Retrieved from http://www.jstor.org/stable/30034952

Basson, I. (2002). Physics and mathematics as interrelated fields of thought development using acceleration as an example. *International Journal of Mathematical Education in Science and Technology*, 33 (5), 679–690. doi: 10.1080/00207390210146023

Bing, T. (2008). *An epistemic framing analysis of upper level physics students' use of mathematics* (Doctoral dissertation). University of Maryland. Retrieved from http://bit.ly/Bing2008

Boesen, J., Lithner, J. & Palm, T. (2010). The relation between types of assessment tasks and the mathematical reasoning student use. *Educational studies in mathematics*, 75 (1), 89–105. doi: 10.1007/s10649-010-9242-9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

diSessa, A. (1993). Towards an epistemology of physics. *Cognition and Instruction*, 10 (2-3), 105–225. doi: 10.1080/07370008.1985.9649008

Garden, R. A., Lie, S., Robitaille, D. F., Angell, C., Martin, M. O. et al. (2006). *TIMSS advanced 2008 assessment frameworks*. Boston: TIMSS & PIRLS International Study Center.

Gray, E. M. & Tall, D. O. (1994). Duality, ambiguity and flexibility: a proceptual view of simple arithmetic. *The Journal for Research in Mathematics Education*, 26 (2), 115–141. Retrieved from http://www.jstor.org/stable/749505

Grønmo, L. S. & Onstad. T. (Eds.) (2013). *The significance of TIMSS and TIMSS Advanced mathematics education in Norway, Slovenia and Sweden*. Retrieved from http://bit.ly/TIMSSAdvNo2013

Haylock, D. (1997). Recognising mathematical creativity in schoolchildren. *ZDM*, 29 (3), 68–74. doi: 10.1007/S11858-997-0002-Y

Hiebert, J. & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: an introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: the case of mathematics* (pp. 1–27). Hillsdale: Erlbaum.

Hopkins, W. G (2002). *A scale of magnitudes for effect statistics*. Retrieved from http://www.sportsci.org/resource/stats/ [A new view of statistics]

Johansson, H. (2015). Mathematical reasoning requirements in Swedish national physics tests. *International Journal of Science and Mathematics Education*, 14 (6), 1133–1152. doi: 10.1007/s10763-015-9636-3

Lesh, R. & Zawojewski, J. (2007). Problem solving and modeling. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 763–804). Charlotte: Information Age Publishing.

Lie, S., Angell, C. & Rohatgi, A. (2010). *Fysikk i fritt fall? TIMSS Advanced 2008 i vidaregående skole* [Physics in free fall? TIMSS Advanced 2008 in upper secondary school]. Retrieved from http://bit.ly/TIMSSAdv2008no

Lithner, J. (2008). A research framework for creative and imitative reasoning. *Educational Studies in Mathematics*, 67 (3), 255–276. doi: 10.1007/s10649-007-9104-2

Michelsen, C. (2005). Expanding the domain – variables and functions in an interdisciplinary context between mathematics and physics. In A. Beckmann, C. Michelsen & B. Sriraman (Eds.), *Proceedings of the 1st international symposium of mathematics and its connections to the arts and sciences* (pp. 201–214). Hildesheim: Verlag Franzbecker.

Nguyen, N.-L. & Meltzer, D. (2003). Initial understanding of vector concepts among students in introductory physics courses. *American Journal of Physics*, 71 (6), 630–638. doi: 10.1119/1.1571831

Palm, T., Boesen, J. & Lithner, J. (2011). Mathematical reasoning requirements in upper secondary level assessments. *Mathematical Thinking and Learning*, 13 (3), 221–246. doi: 10.1080/10986065.2011.564994

Redish, E. F. & Gupta, A. (2009, August). *Making meaning with math in physics: a semantic analysis*. Presented at GIREP 2009, Leicester. Retrieved from https://arxiv.org/pdf/1002.0472.pdf

Redish, E. F. (2003). *Teaching physics with the physics suite*. Hoboken: John Wiley & Sons.

Schoenfeld, A. H. (1985). *Mathematical problem solving.* Orlando: Academic Press.

Silver, E. (1997). *Fostering creativity through instruction rich in mathematical problem solving and problem posing.* ZDM, 29 (3), 75–80. doi: 10.1007/s11858-997-0003-x

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education*, 36 (5), 404–411.

Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education*, 38 (2), 132–135.

Swedish National Agency for Education (2000). *Kursplan för gymnasieskolan – Fysik* [Syllabuses for upper secondary Physics]. Retrieved from http://bit.ly/Fysik2000

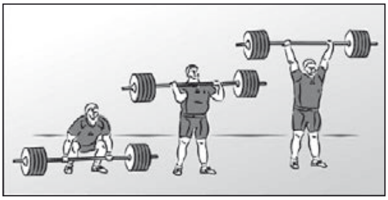
Swedish National Agency for Education (2006). *Curriculum for the non-compulsory school system Lpf 94.* Stockholm: Fritzes.

Swedish National Agency for Education (2009). *Hur samstämmiga är svenska styrdokument och nationella prov med ramverk och uppgifter i TIMSS Advanced 2008?* [How aligned are the Swedish policy documents and national tests with the framework and the tasks in TIMSS Advanced 2008?]. Stockholm: Fritzes.

Appendix

Example A-1

A weightlifter is lifting a barbell that weighs 219 kg. The barbell is lifted 2.1 m up from the floor in 5.0 s.



What is the average power the weightlifter develops on the barbell during the lift?

Analysis

I. Analysis of the assessment task – answers and solutions

A typical solution from an average student could be derived by the relation between power and the change of energy over a specific period of time. In this task, the change of energy is the same as the change of potential energy for the barbell. Multiply the mass of the barbell by the acceleration of gravity and the height of the lift and then divide by the time to get the power asked for. The mathematical subject area is identified as algebra, in this case working with formulas. The identification of the situation to lift a barbell can trigger the student to use a certain solution method and is, therefore, included in this analysis as an identified "real-life" situation.

II. Analysis of the assessment task – task variables

The assignment is to calculate the average power during the lift. The mass of the barbell, the height of the lift, and the time for the lift are all considered as mathematical objects. In this example, all of the objects are given explicitly in the assignment in numerical form. In the presentation of the task, there is also an illustrative figure of the lift.

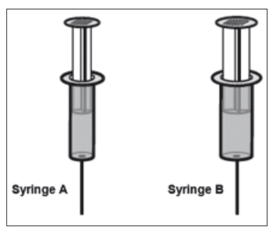
III. Analysis of the textbooks and handbook – answers and solutions

Handbook: Formulas for power, $P = \Delta W/\Delta t$, with the explanation "ΔW = the change in energy during time Δt"; for "work during lift", $W_l = mg \cdot h$, with the explanatory text, "A body with weight mg is lifted to a height h. The lifting work is ..."; and for potential energy with the text "A body with mass m at a height h over the zero level has the potential energy $W_p = mg \cdot h$". *Mathematics book*: Numerous examples and exercises of how to use formulas, e.g. on pages 28–30. *Physics book*: Power is presented as work divided by time, and in on example work is exemplified as lifting a barbell. An identical example is found on page 130. An example of calculating work during a lift in relation to change in potential energy is found on page 136. Exercises 5.05 and 5.10 are solved by a similar algorithm.

IV. Argumentation for the requirement of reasoning
The analysis of the textbooks shows that there are more than three tasks similar to the task being categorised with respect to the task variables, and these tasks can be solved with a similar algorithm. If the students have seen tasks solvable by a similar algorithm at least three times, it is assumed that they will remember the solution procedure. This task is then categorised as solvable using IR.

Example A-2

A patient is going to get an injection. The medical staffs are reading in the instructions that they are supposed to use a syringe that gives the lowest pressure as possible in the body tissue. Which of the syringes A or B shall the staff choose if the same force, F, is applied and the injection needles have the same dimensions? *Argue for the answer*



Analysis
I. Analysis of the assessment task – answers and solutions
To solve this task, the student can use the relation between pressure, force, and area ($p=F/A$). Neglecting the hydrostatic pressure from the injection fluid, if the force applied to the syringe is the same then it is the area of the bottom that affects the pressure. The larger the area, the lower the pressure. The staff should choose syringe B. The mathematical subject area is identified as algebra, such as to work with formulas and proportionality.

II. Analysis of the assessment task – task variables
The assignment is to choose which syringe that gives the minimum pressure and to provide an argument for this choice. Only the force is given as a variable, and this is represented by a letter. Key words for the students

can be *force* and *pressure*. The situation is illustrated by a figure in which it appears that syringe B has a greater diameter than syringe A.

III. Analysis of the textbooks and handbook – answers and solutions
Handbook: The relation p=F/A is defined. *Mathematics book*: Proportionalities are discussed and exemplified but are not used for general comparisons. *Physics book*: One example about how different areas affect the pressure and one exercise that is solved in a similar way by using a general comparison between different areas and pressure.

IV. Argumentation for the requirement of reasoning
There is only one example and one exercise that can be considered similar with regard to the task variables and the solution algorithm. The formula is in the handbook, but there has to be some understanding of the intrinsic properties in order to be able to use the formula in the solution. This task is, therefore, considered to require some CR, in this case GCR, in order to be solved.

Table A-1. *Ratios according to (1) for the tasks in the Physics A 2004 test with the tasks 11 and 8a as reference tasks*

Task	1 (IR)	2 (IR)	3 (IR)	4 (LCR)	5a (NMR)	5b (IR)	6a (NMR)	6b (NMR)	7a (GCR)	7b (NMR)	8a (IR)	8b (IR)	9 (IR)	10 (NMR)	11 (GCR)	12 (GCR)	13 (LCR)	14 G (GCR)	14 VG (GCR)
11	1.17	1.24	1.39	1.45		1.59			1.27		1.38	1.32	1.83		R	2.28	1.91	1.49	3.06
8a	1.12	1.14	1.23	1.24		1.35			1.19		R	1.37	1.37		1.38	1.18	1.53	1.23	1.59

Table A-2. *Ratios according to (1) for the tasks in the Physics A 2005 test with the tasks 10 and 8a as reference tasks*

Task	1 (IR)	2a (NMR)	2b (IR)	3a (IR)	3b (NMR)	4 (NMR)	5 (LCR)	6a (LCR)	6b (LCR)	7 (NMR)	8a (IR)	8b (IR)	9 (IR)	10 (GCR)	11 (IR)	12a (IR)	12b (LCR)	13 (GCR)	14 (LCR)	15 (NMR)
10	1.20		1.24	1.18			1.20		1.53		1.03	1.16	1.43	R	1.65	1.32	1.50	1.45	1.91	
8a	1.02		1.02	1.02			1.03		1.02		R	1.03	1.03	1.03	1.03	1.02	1.03	1.03	1.03	

Table A-3. *Ratios according to (1) for the tasks in the Physics B 2002 test with the tasks 10 and 7 as reference tasks*

Task	1 (NMR)	2 (IR)	3 (IR)	4 (IR)	5 (IR)	6a (LCR)	6b (LCR)	7 (IR)	8 (LCR)	9a (IR)	9b (IR)	10 (GCR)	11 (LCR)	12a (NMR)	12b (LCR)	13 (GCR)	14 (GCR)	15 (GCR)
10		1.36	1.31	1.23	1.69	1.61	1.67	1.54	1.83	1.33	1.59	R	1.69		2.38	2.12	2.21	2.29
7		1.25	1.20	1.18	1.48	1.39	1.42	R	1.38	1.13	1.39	1.49	1.42		1.81	1.51	1.65	1.61

Table A-4. *Ratios according to (1) for the tasks in the Physics B 2003 test with the tasks 8 and 7 as reference tasks*

Task	1 (IR)	2 (IR)	3 (NMR)	4 (IR)	5 (IR)	6 (NMR)	7 (IR)	8 (LCR)	9 (LCR)	10a (LCR)	10b (NMR)	11a (NMR)	11b (IR)	12 (IR)	13 (IR)	14a (NMR)	14b (IR)	15 (GCR)	16 (GCR)
8	1.25	1.11		0.98	1.28		1.19	R	1.53	1.34			1.27	1.47	1.51		1.34	1.47	1.83
7	1.14	1.07		0.99	1.14		R	1.22	1.22	1.17			1.13	1.21	1.23		1.16	1.24	1.30

Table A-5. *Ratios according to (1) for the tasks in the Physics B 2005 test with the tasks 12b and 8b as reference tasks*

Task	1 (NMR)	2 (NMR)	3 (LCR)	4a (NMR)	4b (NMR)	5 (NMR)	6 (IR)	7 (GCR)	8a (NMR)	8b (IR)	9a (IR)	9b (NMR)	10 (GCR)	11 (IR)	12a (IR)	12b (GCR)	12c (IR)	13 (LCR)	14 (GCR)	15a (NMR)	15b (IR)	15c (LCR)	16 (LCR)
12b			1.11				1.07	1.16		1.11	1.10		1.18	1.31	1.17	R	1.35	1.16	1.32		1.11	1.30	1.27
8b			1.34				1.11	1.30		R	1.21		1.45	1.61	1.11	1.11	1.25	1.31	1.54		1.19	1.57	1.61

Table A-6. *Ratios according to (1) for the tasks in the Physics B 2006 test with the tasks 12a and 10b as reference tasks*

Task	1 (NMR)	2 (IR)	3 (NMR)	4 (GCR)	5 (LCR)	6 (NMR)	7 (NMR)	8a (IR)	8b (IR)	9 (IR)	10a (NMR)	10b (IR)	11 (NMR)	12a (GCR)	12b (GCR)	12c (LCR)	13 (IR)	14a (NMR)	14b (IR)	15 (NMR)	16 G (IR)	16 VG (NMR)
12a		1.21		1.67	1.39			1.36	1.38	1.35		1.27		R	2.21	2.07	1.66		1.64		1.29	
10b		1.08		1.24	1.22			1.21	1.32	1.34		R		1.27	1.35	1.42	1.44		1.38		1.91	

Table A-7. *Ratios according to (1) for the tasks in the Physics B 2010 test with the tasks 11b and 9b as reference tasks*

Task	1 (NMR)	2 (NMR)	3 (IR)	4a (IR)	4b (NMR)	5 (IR)	6 (IR)	7a (IR)	7b (GCR)	8 (LCR)	9a (NMR)	9b (IR)	9c (GCR)	10 (GCR)	11a (LCR)	11b (GCR)	11c (GCR)	12 (GCR)	13 (IR)	14 G (LCR)	14 VG (GCR)
11b			1.24	1.25		1.25	1.50	1.23	1.45	1.39		1.43	1.70	1.93	1.36	R	2.69	1.91	1.85	1.80	2.47
9b			1.10	1.08		1.22	1.16	1.14	1.25	1.25		R	1.59	1.51	1.14	1.43	1.52	1.45	1.44	1.33	1.70

Table A-8. *Ratios according to (1) for the Physics A 2004 test with the new task 7a together with tasks 11 and 8a as reference tasks*

Task	1 (IR)	2 (IR)	3 (IR)	4 (LCR)	5a (NMR)	5b (IR)	6a (NMR)	6b (NMR)	7a (GCR)	7b (NMR)	8a (IR)	8b (IR)	9 (IR)	10 (NMR)	11 (GCR)	12 (GCR)	13 (LCR)	14 G (GCR)	14 VG (GCR)
11	1.17	1.24	1.39	1.45		1.59			1.27		1.38	1.32	1.83		R	2.28	1.91	1.49	3.06
8a	1.12	1.14	1.23	1.24		1.35			1.19		R	1.37	1.37		1.38	1.18	1.53	1.23	1.59
7a	1.10	2.22	1.17	1.24		1.24			R		1.19	1.18	1.31		1.27	1.45	1.37	1.26	

Table A-9. *Ratios according to (1) for the Physics A 2005 test with the new task 12a together with tasks 10 and 8a as reference tasks*

Task	1 (IR)	2a (NMR)	2b (IR)	3a (IR)	3b (NMR)	4 (NMR)	5 (LCR)	6a (LCR)	6b (LCR)	7 (NMR)	8a (IR)	8b (IR)	9 (IR)	10 (GCR)	11 (IR)	12a (IR)	12b (LCR)	13 (GCR)	14 (LCR)	15 (NMR)
10	1.20		1.24	1.18			1.20		1.53		1.03	1.16	1.43	R	1.65	1.32	1.50	1.45	1.91	
8a	1.02		1.02	1.02			1.03		1.02		R	1.03	1.03	1.03	1.03	1.02	1.03	1.03	1.03	
12a	1.15		1.17	1.14			1.14		1.35		1.02	1.12	1.33	1.32	1.42	R	1.41	1.33	1.59	

Table A-10. *Ratios according to (1) for the Physics B 2002 test with the new task 9b together with tasks 10 and 7 as reference tasks*

Task	1 (NMR)	2 (IR)	3 (IR)	4 (IR)	5 (IR)	6a (LCR)	6b (LCR)	7 (IR)	8 (LCR)	9a (IR)	9b (IR)	10 (GCR)	11 (LCR)	12a (NMR)	12b (LCR)	13 (GCR)	14 (GCR)	15 (GCR)
10		1.36	1.31	1.23	1.69	1.61	1.67	1.54	1.83	1.33	1.59	R	1.69		2.38	2.12	2.21	2.29
7		1.25	1.20	1.18	1.48	1.39	1.42	R	1.38	1.13	1.39	1.49	1.42		1.81	1.51	1.65	1.61
9b		1.30	1.25	1.25	1.56	1.53	1.51	1.53	1.51	1.42	R	1.70	1.55		2.25	2.21	2.07	1.68

Table A-11. *Ratios according to (1) for the Physics B 2005 test with the new task 7 together with tasks 12b and 8b as reference tasks*

Task	1 (NMR)	2 (NMR)	3 (LCR)	4a (NMR)	4b (NMR)	5 (NMR)	6 (IR)	7 (GCR)	8a (NMR)	8b (IR)	9a (IR)	9b (NMR)	10 (GCR)	11 (IR)	12a (IR)	12b (GCR)	12c (IR)	13 (LCR)	14 (GCR)	15a (NMR)	15b (IR)	15c (LCR)	16 (LCR)
12b			1.11				1.07	1.16		1.11	1.10		1.18	1.31	1.17	R	1.35	1.16	1.32		1.11	1.30	1.27
8b			1.34				1.11	1.30		R	1.21		1.45	1.61	1.11	1.11	1.25	1.31	1.54		1.19	1.57	1.61
7			1.24				1.14	R		1.30	1.21		1.47	1.66	1.11	1.16	1.28	1.35	1.62		1.19	1.59	1.60

Table A-12. *Ratios according to (1) for the Physics B 2006 test with the new task 4 together with tasks 12a and 10b as reference tasks*

Task	1 (NMR)	2 (IR)	3 (NMR)	4 (GCR)	5 (LCR)	6 (NMR)	7 (NMR)	8a (IR)	8b (IR)	9 (IR)	10a (NMR)	10b (IR)	11 (NMR)	12a (GCR)	12b (GCR)	12c (LCR)	13 (IR)	14a (NMR)	14b (IR)	15 (NMR)	16 G (IR)	16 VG (NMR)
12a		1.21		1.67	1.39			1.36	1.38	1.35		1.27		R	2.21	2.07	1.66		1.64		1.29	
10b		1.08		1.24	1.22			1.21	1.32	1.34		R		1.27	1.35	1.42	1.44		1.38		1.91	
4		1.14		R	1.3			1.25	1.32	1.27		1.24		1.67	1.62	1.54	1.62		1.81		2.37	

Table A-13. *Ratios according to (1) for the Physics B 2010 test with the new task 7b together with tasks 11b and 9b as reference tasks*

Task	1 (NMR)	2 (NMR)	3 (IR)	4a (IR)	4b (NMR)	5 (IR)	6 (IR)	7a (IR)	7b (GCR)	8 (LCR)	9a (NMR)	9b (IR)	9c (GCR)	10 (GCR)	11a (LCR)	11b (GCR)	11c (GCR)	12 (GCR)	13 (IR)	14 G (LCR)	14 VG (GCR)
11b			1.24	1.25		1.25	1.50	1.23	1.45	1.39		1.43	1.70	1.93	1.36	R	2.69	1.91	1.85	1.80	2.47
9b			1.10	1.08		1.22	1.16	1.14	1.25	1.25		R	1.59	1.51	1.14	1.43	1.52	1.45	1.44	1.33	1.70
7b			1.18	1.19		1.19	1.17	1.60	R	1.30		1.25	1.52	1.83	1.18	1.45	1.57	1.64	1.50	1.46	1.98

Table A-14. *Odds-ratios according to (2) for the tasks in the Physics A 2004 test with tasks 11, 8a and 7a as reference tasks*

Task	1 (IR)	2 (IR)	3 (IR)	4 (LCR)	5a (NMR)	5b (IR)	6a (NMR)	6b (NMR)	7a (GCR)	7b (NMR)	8a (IR)	8b (IR)	9 (IR)	10 (NMR)	11 (GCR)	12 (GCR)	13 (LCR)	14 G (GCR)	14 VG (GCR)
11	2.84	4.06	4.44	3.30		3.63			2.65		2.55	2.19	3.38		R	5.23	4.55	2.97	24.3
8a	2.23	2.44	2.63	2.19		2.65			2.31		R	3.15	2.08		2.55	1.39	3.29	1.87	2.88
7a	2.47	2.46	2.58	2.94		2.61			R		2.31	2.15	2.57		2.65	4.08	3.54	2.87	–

Table A-15. *Odds-ratios according to (2) for the tasks in the Physics A 2005 test with tasks 10, 8a and 12a as reference tasks*

Task	1 (IR)	2a (NMR)	2b (IR)	3a (IR)	3b (NMR)	4 (NMR)	5 (LCR)	6a (NMR)	6b (LCR)	7 (NMR)	8a (IR)	8b (IR)	9 (IR)	10 (GCR)	11 (IR)	12a (IR)	12b (LCR)	13 (GCR)	14 (LCR)	15 (NMR)
10	1.73		4.23	2.86			1.80		3.26		3.14	2.94	2.64	R	3.89	2.79	3.26	2.51	5.62	
8a	2.48		2.69	3.40			2.98		1.93		R	6.35	3.29	3.14	3.20	1.98	3.11	2.43	2.38	
12a	1.65		2.87	2.54			1.63		2.56		1.98	2.44	2.48	2.79	2.94	R	3.29	2.27	4.06	

Table A-16. *Odds-ratios according to (2) for the tasks in the Physics B 2002 test with tasks 10, 7 and 9b as reference tasks*

Task	1 (NMR)	2 (IR)	3 (IR)	4 (IR)	5 (IR)	6a (LCR)	6b (LCR)	7 (IR)	8 (LCR)	9a (IR)	9b (IR)	10 (GCR)	11 (LCR)	12a (NMR)	12b (LCR)	13 (GCR)	14 (GCR)	15 (GCR)
10		2.84	3.14	2.53	3.92	4.50	5.43	4.06	4.49	3.67	3.53	R	4.26		6.44	4.32	5.03	5.31
7		2.69	2.66	2.97	4.70	3.90	4.46	R	3.71	2.51	4.18	4.09	3.64		10.2	4.15	5.72	5.95
9b		2.31	2.59	2.99	3.74	4.14	4.11	4.65	3.30	4.36	R	3.91	3.66		7.94	6.72	5.47	4.27

Table A-17. *Odds-ratios according to (2) for the tasks in the Physics B 2003 test with the tasks 8 and 7 as reference tasks*

Task	1 (IR)	2 (IR)	3 (NMR)	4 (IR)	5 (IR)	6 (NMR)	7 (IR)	8 (LCR)	9 (LCR)	10a (LCR)	10b (NMR)	11a (NMR)	11b (IR)	12 (IR)	13 (IR)	14a (NMR)	14b (IR)	15 (GCR)	16 (GCR)
8	2.63	2.95		1.98	4.17		4.34	R	5.94	2.75			4.03	5.68	4.51		4.64	3.80	17.3
7	2.55	2.94		2.61	3.27		R	4.61	3.85	2.62			3.95	4.40	3.47		3.56	6.52	6.97

Table A-18. *Odds-ratios according to (2) for the tasks in the Physics B 2005 test with tasks 12b and 8b and 7 as reference tasks*

Task	1 (NMR)	2 (NMR)	3 (LCR)	4a (NMR)	4b (NMR)	5 (NMR)	6 (IR)	7 (GCR)	8a (NMR)	8b (IR)	9a (IR)	9b (NMR)	10 (GCR)	11 (IR)	12a (IR)	12b (GCR)	12c (IR)	13 (LCR)	14 (GCR)	15a (NMR)	15b (IR)	15c (LCR)	16 (LCR)
12b			1.57				2.18	2.03		1.73	1.79		1.92	2.57	4.88	R	3.41	2.11	2.67		1.77	2.50	2.23
8b			3.25				3.09	3.10		R	2.71		4.03	5.37	2.28	1.73	2.00	3.41	4.20		2.46	4.62	5.09
7			2.00				3.89	R		3.10	2.42		3.27	4.01	2.22	2.03	1.89	3.29	3.65		2.12	3.45	3.38

Table A-19. *Odds-ratios according to (2) for the tasks in the Physics B 2006 test with tasks 12a, 10b and 4 as reference tasks*

Task	1 (NMR)	2 (IR)	3 (NMR)	4 (GCR)	5 (LCR)	6 (NMR)	7 (NMR)	8a (IR)	8b (IR)	9 (IR)	10a (NMR)	10b (IR)	11 (NMR)	12a (GCR)	12b (GCR)	12c (LCR)	13 (IR)	14a (NMR)	14b (IR)	15 (NMR)	16 G (IR)	16 VG (NMR)
12a		3.12		3.40	3.39			3.76	2.58	2.59		2.34		R	19.3	8.35	3.41		3.82		1.67	
10b		1.70		1.93	2.54			2.71	3.06	3.70		R		2.34	3.03	3.42	3.53		3.19		3.66	
4		1.88		R	2.64			2.14	1.95	1.88		1.93		3.40	3.10	2.33	2.54		3.91		3.44	

Table A-20. *Odds-ratios according to (2) for the tasks in the Physics B 2010 test with tasks 11b, 9b and 7b as reference tasks*

Task	1 (NMR)	2 (NMR)	3 (IR)	4a (IR)	4b (NMR)	5 (IR)	6 (IR)	7a (IR)	7b (GCR)	8 (LCR)	9a (NMR)	9b (IR)	9c (GCR)	10 (GCR)	11a (LCR)	11b (GCR)	11c (GCR)	12 (GCR)	13 (IR)	14 G (LCR)	14 VG (GCR)
11b			8.36	6.51		3.77	3.86	2.12	3.03	3.86		4.25	3.34	4.80	16.9	R	41.8	4.46	4.84	5.68	10.4
9b			2.31	1.96		4.60	1.83	1.91	2.41	3.24		R	5.89	4.22	2.66	4.25	4.64	3.36	3.64	2.88	7.51
7b			4.01	3.94		2.86	1.61	11.8	R	3.07		2.41	2.76	5.15	3.02	3.03	3.22	3.32	2.83	2.98	5.61

Helena Johansson

Helena Johansson has a PhD in Mathematics, specialising in Educational Sciences and is a postdoc at Mid Sweden University. Her research interests concern students' mathematical reasoning and how this competence influences students' learning in mathematics and in physics; and how natural language influences students' learning of the symbolic language of mathematics.

helena.johansson@miun.se