Journal of Educational Psychology
2018, Vol. 110, No. 8, 1175–1191

0022-0663/18/$12.00 http://dx.doi.org/10.1037/edu0000266

Enhancing Physics Learning With Cognitively Activating Instruction: A Quasi-Experimental Classroom Intervention Study

Sarah I. Hofer
Technische Universität München

Ralph Schumacher and Herbert Rubin
MINT-Learning Center, ETH Zurich, Switzerland

Elsbeth Stern
Chair of Learning and Instruction, ETH Zurich, Switzerland

Physics educators today face two major challenges: supporting the acquisition of a solid base of conceptual knowledge and reducing the persisting gender gap. In the present quasi-experimental study, we investigated the potential of physics instruction that is enriched with evidence-based cognitively activating methods, such as inventing with contrasting cases or metacognitive questions, to overcome both of these challenges. Four physics teachers in charge of two parallel classes each applied our cognitively activating instruction in one of their classes (CogAct classes). The other classes received regular physics lessons (regular classes) on the same content. The sample consisted of 172 individuals from the advanced track of Swiss secondary school. Controlling for several individual student variables, CogAct classes ($N = 87$) outperformed regular classes ($N = 85$) in conceptual understanding at posttest ($p < .01$, $\beta = 0.19$, 95% CI [0.07, 0.32]) and three months later ($p < .05$, $\beta = 0.13$, 95% CI [0.00, 0.26]). The CogAct classes' advantage in conceptual understanding was not at the expense of their quantitative problem-solving performance, which even exceeded the quantitative problem-solving performance of the regular classes at posttest ($p < .05$, $\beta = 0.14$, 95% CI [0.00, 0.28]). In addition, female students with above-average intelligence (PR >75) particularly benefited from CogAct instruction, as indicated by descriptive statistics and the interaction between intelligence and condition in the group of the female students for posttest conceptual understanding ($p < .05$, $\beta = 0.88$, 95% CI [0.06, 1.69]). We conclude that teachers can successfully be supported in implementing cognitively activating methods that improve their students' conceptual understanding and reduce the gender gap in physics.

Educational Impact and Implications Statement

Physics instruction has not yet adapted appropriately to the difficulties that otherwise capable students may have with understanding concepts like force or inertia. We integrated several methods of classroom practice, all of which have the potential to help students acquire meaningful knowledge, into a comprehensive teaching unit on Newtonian mechanics. Students taught according to this unit developed a better understanding of central concepts in mechanics than students who were taught by the same teachers in a traditional way. In particular, the group of intelligent female students, who often do not exploit their potential in traditional physics instruction, profited from the teaching unit that focused on conceptual understanding. Our findings show that this method of physics instruction can be successfully implemented by regular in-service teachers, supports the acquisition of meaningful knowledge, and reduces the gender gap in physics.

Keywords: physics instruction, cognitively activating, conceptual knowledge, gender, science education

Physics is considered a difficult subject even for otherwise capable learners. This has been extensively shown in the past four decades in the area of mechanics. Although this topic is part of the early physics curriculum in secondary schools all over the world, instruction has rarely met with much success. Clement (1982) was among the first to demonstrate the limited effects of instruction even in a highly selective group of engineering students. It turned out that even after one year of college instruction, the majority of

---

Sarah I. Hofer, TUM School of Education, Centre for International Student Assessment (ZIB), Technische Universität München; Ralph Schumacher and Herbert Rubin, MINT-Learning Center, ETH Zurich, Switzerland; and Elsbeth Stern, Chair of Learning and Instruction, ETH Zurich, Switzerland.

We express our gratitude to Bahar Behzadi, Monica Vogel-Stalder, Conradin Beeli, André van der Graaff, and Mark Heinz. Moreover, we thank Pál Molnár, Samuel Nuesch, Sebastian Seehars, and Andreas Vaterlaus for their support and Jessica Büetiger for her great assistance throughout the project. We also like to thank Bruno Rütsche and Peter Edelsbrunner for their help with technical and software issues.

Correspondence concerning this article should be addressed to Sarah I. Hofer, TUM School of Education, Centre for International Student Assessment (ZIB), Technische Universität München, Arcisstraße 21, 80333 Munich, Germany. E-mail: sarah.hofer@tum.de

learners held severe misunderstandings about Newton's laws of mechanics, although they had no difficulties reciting them. In the following years, the failure of physics instruction to generate understanding was repeatedly documented by a large number of studies on physics education (e.g., Duit & Treagust, 2003; Hake, 1998; Halloun & Hestenes, 1985; Labudde, Reif, & Quinn, 1988; McCloskey, Washburn, & Felch, 1983; McDermott, 1984). In many of these studies, the Force Concept Inventory (FCI; Hestenes, Wells, & Swackhamer, 1992) was applied, a multiple-choice test featuring correct and incorrect statements about the concepts of velocity, acceleration, and force.

A problem resembling those on the FCI is Problem 2, depicted in Figure 1. Figure 1 presents two inherently different problems that could be used for formative and summative assessment in physics classes addressing Newton's laws of mechanics. Problem 1 requires access to the formula "force $=$ mass $*$ acceleration," which is central to Newton's mechanics. Together with a set of additional formulae (e.g., $v = a * t$, $x = [1/2] * a * t^2$), learners are well equipped for solving a broad range of word problems addressing the quantitative relationship between central concepts of mechanics. Learners' quantitative problem-solving performance is likely to improve when they are explicitly trained. Algorithms can be applied to solve similar problems even without conceptual understanding (Leppävirta, Kettunen, & Sihvola, 2011). However, as soon as learners are confronted with quantitative problems that are different from the ones they have practiced, deficits in conceptual knowledge become apparent (Redish, Saul, & Steinberg, 1998). Without conceptual understanding, students may be able to solve Problem 1 but fail to solve Problem 2 because they hold beliefs about force and acceleration that are in line with many everyday experiences but incompatible with theories of physics (Hake, 1998; Mazur, 2015). The third option in Problem 2 (a), for instance, may reflect an intuitive but wrong understanding, according to which only the actively pulling skater exerts a force. Students who have already had instruction on action and reaction forces might nevertheless tick the second option for Problem 2 (a) because their understanding has remained superficial. Only those who tick the correct fifth option in Problem 2 (a) and the correct third option in Problem 2 (b) can be expected to have really understood how action and reaction forces come into play in real-world situations.

The studies on conceptual understanding in physics published in the 1980s and 1990s have created an awareness of the importance of prior knowledge for future learning, as demonstrated by a quote from Carey (2000, p. 13): "Now we understand that the main barrier to learning the curricular materials we so painstakingly developed is not what the student lacks, but what the student has, namely, alternative conceptual frameworks for understanding the phenomena covered by the theories we are trying to teach. Often these conceptual frameworks work well for children, so we face a

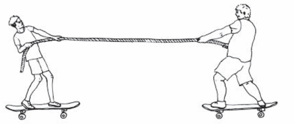
1. Quantitative Problem	**2. Conceptual Problem**
The rocket Ariane 5 of the European Space Agency (ESA) serves to launch communication satellites into geostationary orbit. At take-off, the engines produce a force of $12 * 10^6$ N. The weight of the rocket is $8 * 10^6$ N. Calculate the acceleration of the rocket at take-off.	a) Two skaters with clearly different body weights stand opposite each other, each on their own skateboard, and are connected by a rope under tension. The lighter skater on the left pulls actively on the rope, while the heavier skater on the right just holds on to it. Which of the following statements are true?  ☐ They will meet at a point that is closer to the starting position of the lighter skater. ☐ Nothing happens, because the force of the pull results in an equally large counter-force, such that both forces cancel each other out. ☐ The lighter skater remains stationary and the heavier skater rolls towards him. ☐ Both move towards the middle at the same speed. ☐ They meet at a point that is closer to the starting position of the heavier skater. b) Which of the following explanations for your answer(s) is correct? **Please tick only one answer.** ☐ As the left skater pulls on the right skater, and not the other way round, the right skater moves. ☐ Because the right skater holds on to the rope, the rope also transmits some pulling force to the left skater. ☐ The right skater has to hold on to the rope with a force that is equivalent to the force with which the left skater is pulling. Therefore, the same amount of force acts on both. ☐ The left skater is affected by his own force plus the force with which the right skater is holding on to the rope. Therefore the left skater moves faster than the right. ☐ The pulling force of the left skater is divided, through the rope, half to the left and half to the right skater.

Figure 1. Two physics problems that address the relationship between force, mass, and acceleration.

problem of trying to change theories and concepts." In past decades, science educators and psychologists have made good progress in understanding how to foster this type of conceptual change in the classroom: students must become aware of the limits of their everyday concepts and become convinced by the explanations offered during the instruction. This approach requires a classroom culture in which questioning and respect for initially diverse beliefs prevail, as was realized, for instance, in the benchmark lessons by DiSessa and Minstrell (1998).

There is evidence that instruction focusing on the acquisition of qualitative conceptual knowledge is also beneficial for the development of quantitative problem-solving skills. Ploetzner, Fehse, Kneser, and Spada (1999), for instance, showed in 10th Grade mechanics classes that conceptual understanding and quantitative problem-solving skills could be successfully taught by means of concept maps (i.e., diagrams representing conceptual knowledge by depicting the relations between concepts in a specific subject area). Students who first learned about concepts gained more from subsequent instruction on quantitative problem solving than students with a reversed order of instruction. In their meta-analysis on teaching science problem solving, Taconis, Ferguson-Hessler, and Broekkamp (2001) concluded that instruction focusing on the underlying concepts seems to be effective, whereas practicing problem solving was of little importance. Hake (1998) compared traditional teaching with student-centered methods focusing on conceptual understanding of mechanics at the high school, college, and university level. He reported that the student-centered methods focusing on conceptual understanding outperformed traditional instruction on both a conceptual knowledge measure and a more quantitative problem-solving test.

Despite such findings, physics instruction at school still predominantly follows traditional procedures of demonstrating experiments and introducing the laws of physics by referring to and practicing equations rather than focusing on students' naïve concepts and beliefs, as shown in recent studies conducted in different countries (e.g., Fischer, Labudde, Neumann, & Viiri, 2014; Nieminen, Savinainen, & Viiri, 2010; Seidel & Prenzel, 2006). This failure to activate and work on the students' existing concepts may explain the findings from cross-sectional studies with the FCI that did not reveal noticeable progress in students' conceptual understanding of mechanics (Fulmer, Liang, & Liu, 2014; Kim & Pak, 2002).

Although the gap between scientifically proven instructional means and their successful implementation in regular classrooms is still unsatisfactory, the results from large-scale school studies on mathematics (Krauss et al., 2008; Staub & Stern, 2002) and physics (Keller, Neumann, & Fischer, 2017) point in a promising direction: a significant amount of between-classroom differences in learning outcomes could be traced back to the teachers' competence in providing cognitively activating learning opportunities. This was mainly realized by presenting conceptually demanding problems and by offering opportunities for meaningful learning during classroom discussions.

Psychologists and science educators have developed and evaluated a number of cognitively activating instructional methods that are intended to help students activate relevant prior knowledge, change existing knowledge, and construct new conceptual knowledge (Berthold & Renkl, 2010; Dole & Sinatra, 1998; Schneider & Stern, 2010). These methods of cognitive activation include, among others, comparing and contrasting or self-explanations, and their effectiveness has been proven in experimental studies. However, to routinely implement methods of cognitive activation in the classroom, teachers need concrete suggestions for how to initiate and structure learning activities when addressing a particular type of content in the curriculum (Guskey, 2002). We hence developed an instructional unit on Newton's mechanics that fits the content addressed in the traditional secondary school curriculum but that is also enriched with evidence-based means of cognitive activation. In a quasi-experimental study, we wanted to determine whether regular in-service teachers are able to implement cognitively activating instructional methods under realistic classroom conditions to the benefit of their students' conceptual understanding, without hampering their quantitative problem-solving performance.

The unsatisfactory situation in physics education also becomes obvious in the huge and persisting achievement differences between male and female students, to the disadvantage of the latter (see, e.g., Ceci, Williams, & Barnett, 2009; Halpern, 2014). In particular, female students' ongoing difficulties in understanding the concepts of mechanics have been demonstrated by Madsen, McKagan, and Sayre (2013). With our study, we also address the question of whether classroom instruction that focuses more on qualitative conceptual understanding than on quantitative problem solving is especially beneficial for female students.

Applying Cognitively Activating Instructional Methods to Mechanics Instruction

In the instructional unit on Newton's mechanics, we implemented five cognitively activating methods that have each proven successful in boosting learners' conceptual understanding in experimental studies. We brought them together to form a tool kit that can be applied by regular in-service teachers within a comprehensive instructional unit, hereafter referred to as CogAct instruction. In what follows, the scientific rationale and the empirical evidence are presented for each method. Each method is illustrated by an example from our instructional unit.

Generating Solutions to Novel Problems Prior to Instruction (Productive Failure)

Knowledge construction and reorganization starts with the learner's insight that a given problem cannot be solved by referring to one's preexisting knowledge. To involve students in active knowledge construction, they must be confronted with interesting phenomena they cannot explain. This situation activates related prior knowledge, raises interest, and makes students aware of their knowledge deficits. Failing in this context can be productive (*productive failure*) because it reveals the limits of the students' existing knowledge and can hence initiate conceptual change (see Chinn & Brewer, 1993; Kapur, 2014; Sanchez, Garcia-Rodicio, & Acuna, 2009; Sinatra & Pintrich, 2003).

The CogAct instructional unit suggests starting the lesson on active and reactive forces (Newton's third axiom) with the problem presented in Figure 2. The vast majority of the students believe that the force meter in the situation on the right will display 20 N. However, the correct answer would be that the force meter still displays 10 N.

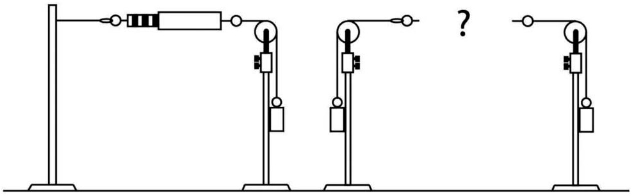


Figure 2. An example of a novel problem intended to induce a productive failure. Students are shown the depicted situations, which are introduced by the teacher in the following way: First, we consider the situation on the left. If we hang a weight of 1 kg on the rope that is connected to the force meter, it displays a force of 10 N. Second, we turn to the situation on the right. What occurs if we hang an additional weight of 1 kg on the other side of the force meter? Will it now display the same or a different amount of Newtons?

Inventing With Contrasting Cases

Learning can be promoted by instructing students to invent a principle before the scientific concept is introduced. Learners are presented with several contrasting cases illustrating a specific underlying concept (e.g., linear graphs with different slopes) and are instructed to discover the concept (e.g., to invent a common index that can be used to describe the slopes of these linear graphs). After the completion of the invention task, the scientific explanation is introduced. This instructional method requires students to actively address a given problem and activate relevant prior knowledge. It helps the students process and understand subsequent instruction (e.g., Schalk, Schumacher, Barth, & Stern, 2017; Schwartz, Chase, Oppezzo, & Chin, 2011; Schwartz & Martin, 2004).

In the CogAct instructional unit, students, for instance, are encouraged to come up with a parallelogram of forces to explain why a sagging thin thread can carry an umbrella that is suspended on the thread, whereas a tense thin thread cannot carry the umbrella and breaks immediately (depicted in Figure 3a). By drawing the parallelogram, as depicted in Figure 3b, one recognizes the increase in the resulting forces in the case of the tense thin thread.

Comparing and Contrasting

To differentiate between superficially similar concepts, two or more situations that instantiate the concepts can be juxtaposed and contrasted. The direct confrontation supports the extraction of the specific features of the single underlying concepts, and differences between the concepts are emphasized. Likewise, situations that are superficially different but represent the same underlying concept can be compared to derive the general principle that connects all cases. In both approaches, learners are instructed to describe relevant commonalities and differences between the situations (see, e.g., Gick & Holyoak, 1983; Schalk, Saalbach, & Stern, 2016; Schwartz & Bransford, 1998; Ziegler & Stern, 2014, 2016).

In mechanics, students often confuse active and reactive forces on the one hand, and equilibrium forces on the other hand (Camp et al., 1994), although both concepts explain entirely different situations. The CogAct instructional unit suggests increasing awareness of the difference between both concepts by presenting situations like the ones depicted in Figure 4.

Self-Explanation Prompts

Self-explanations are explanations that are constructed for and addressed to oneself in order to clarify and rethink concepts. Self-explanation prompts ask students to deliberate on central points of the learning content and draw connections to preexisting knowledge. There is broad evidence that prompting self-explanations by specific questions is an effective way of enhancing students' understanding (e.g., Chi, De Leeuw, Chiu, & LaVancher, 1994; Rittle-Johnson, 2006; Schworm & Renkl, 2007).

A wide range of content-specific self-explanation prompts is included in the CogAct instructional unit at the end of each of the main chapters. For instance, there are prompts for self-explanations addressing potential false beliefs, such as "Someone believes that mass and weight are the same. This is wrong. Which arguments would you apply to convince the person that this is not true?"

Metacognitive Questions

These questions prompt students to reflect on their state of knowledge and their learning progress. Students may thereby become aware of contradictions and shortcomings of their conceptual knowledge (e.g., Mevarech & Fridkin, 2006; White & Frederiksen, 1998; Zepeda, Elizabeth, Ronevich, & Nokes-Malach, 2015). In the example from the CogAct instructional unit presented

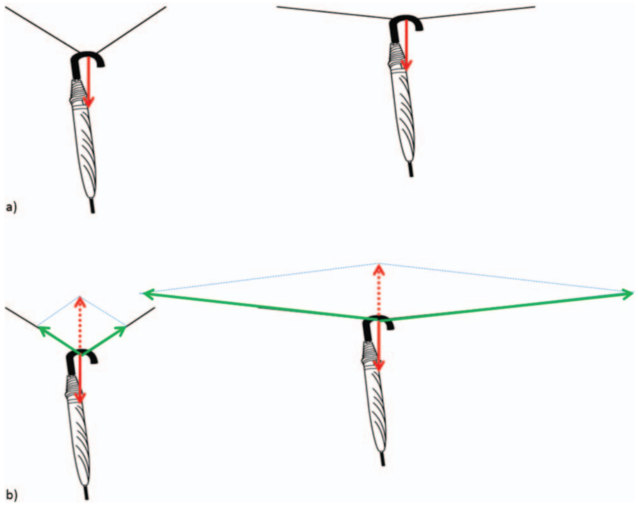


Figure 3. An example of the method inventing with contrasting cases. On the basis of the illustrations shown in the upper part of the figure (a), students are stimulated to come up with a parallelogram of forces to explain why a sagging thin thread can carry an umbrella that is suspended from the thread, whereas a tense thin thread cannot carry the same umbrella and breaks immediately. The correct solution is presented in the lower part of the figure (b). The weight of the umbrella (red arrow) has to be compensated by the forces in the thread (two green arrows). The resulting force of these two forces, which has to be the same size as the weight of the umbrella, is indicated by the dotted red arrow. The different parallelograms of forces show that the forces in the thread significantly increase as the angle between them gets larger (i.e., as the thread is tensioned). These increased forces in the tensioned thread explain why the tense thread cannot carry the umbrella and breaks immediately. See the online article for the color version of this figure.

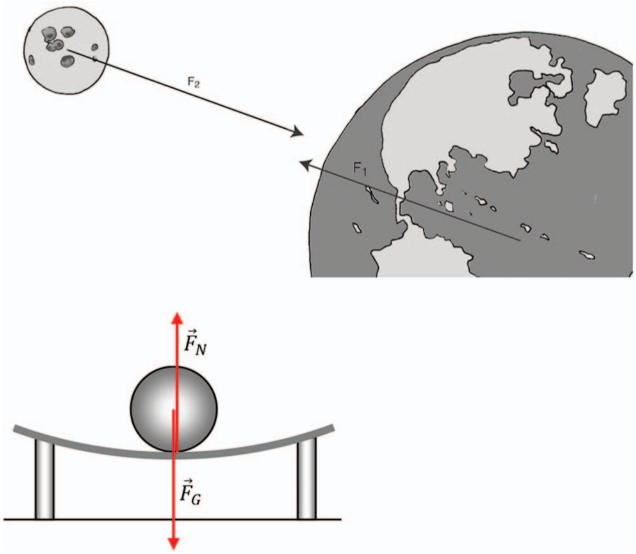


Figure 4. An example of the method comparing and contrasting. The situations depicted illustrate either action and reaction forces or equilibrium forces. To help students discern these two concepts, they are asked, for instance, whether the gravitational forces between the earth and the moon are the same size or different sizes, and whether the forces (red arrows) in the case of the bowl on the elastic surface are the same size or different sizes. They are prompted to explain their answers with the concepts of either active and reactive forces or equilibrium forces. See the online article for the color version of this figure.

in Figure 5, students should make sure whether they have understood that without friction and its corresponding reactive force, movement would not be possible.

Why Cognitively Activating Instruction May Reduce the Gender Gap in Physics

Male students outperform female students in physics beginning in secondary school, with the gap increasing in the following years (see Ceci et al., 2009). Such a gap is also found at the upper end of the intelligence scale: a smaller proportion of females than males perform on a high level in the field of physics (e.g., Lubinski & Benbow, 1992; Seidel, 2006). Accordingly, the proportion of physics underachievers (i.e., low performance despite high intelligence) in secondary school turns out to be much higher among female students than among male students (Hofer & Stern, 2016). Reasons for these gender disparities are manifold (Halpern, 2014), and among many other factors, instructional practice may have an impact. There is evidence that the conventional way of teaching, with a focus on formalization may prevent even more female than male students from developing their potential. Zohar and Sela (2003) found that female students in Israeli advanced placement physics classes, in particular, suffered from a lack of teaching for understanding. They felt particularly uncomfortable with formulae if they did not understand what the variables stood for and requested more time for discussing the concepts. Similar preferences were found for female students in physics classes in secondary schools in German-speaking countries (Labudde, Herzog, Neuenschwander, Violi, & Gerber, 2000; Stadler, Duit, & Benke, 2000). Studies conducted in German Gymnasium classes (secondary schools that prepare their students for higher education at a university) highlighted differences between male and female students in their learning behavior in physics classes. Male students not only showed more verbal engagement during instruction (Jurik, Gröschner, & Seidel, 2013) but also reported more deep learning strategies than female students (Jurik, Gröschner, & Seidel, 2014). The latter was measured with a questionnaire that, for instance, referred to the use of self-explanations (e.g., "During the last two lessons, I explained the content in my own words"). The results suggest that in traditional physics lessons more male than female students made spontaneous use of means of cognitive activation.

One conclusion to be drawn from the aforementioned literature is that an instruction focusing on a deeper conceptual understanding should particularly benefit female students in unfolding their unused potential. Evidence for this claim comes from Lorenzo, Crouch, and Mazur (2006), who compared the learning gains of a traditional nonmajor mechanics course at Harvard University to the gains of courses that focused on conceptual understanding by interactive engagement. They found that both genders benefited from interactive methods, but female students improved their performance more, which decreased the gender gap. The interactive engagement methods used at Harvard shared many aspects with the methods used in CogAct instruction, particularly the opportunity to articulate thoughts about concepts verbally. However, attempts to replicate the reduction of the gender gap by interactive engagement at a less-selective university than Harvard failed (Pollock, Finkelstein, & Kost, 2007). Additionally, attempts in German-speaking countries to provide more cognitively activating interactive physics lessons by group work (Hänze & Berger, 2007) or by a more communicative classroom culture (Labudde et al., 2000) did not reduce the gender gap. These mixed findings suggest that further research is needed to determine under what conditions and to what extent cognitively activating instruction can benefit females. The reduction of the gender gap found in the highly selective sample by Lorenzo et al. (2006) may indicate that females with above-average intelligence in particular are able to make use of cognitively activating interactive engagement methods to realize their potential. The design of our study allows us to address these questions.

The Present Study

To facilitate the integration of physics instruction, which focuses on conceptual understanding, into everyday school life, we

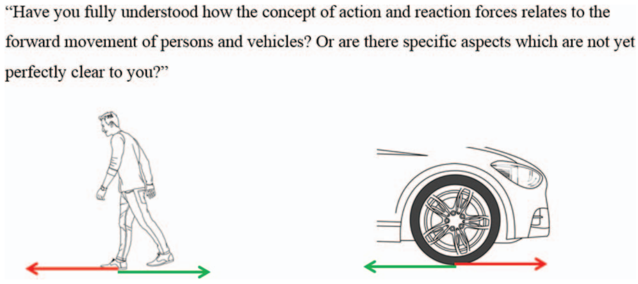

"Have you fully understood how the concept of action and reaction forces relates to the forward movement of persons and vehicles? Or are there specific aspects which are not yet perfectly clear to you?"

Figure 5. An example of a metacognitive question. See the online article for the color version of this figure.

constructed a comprehensive unit on Newtonian mechanics that combines different cognitively activating methods and hence provides practitioners with a tool kit for teaching for conceptual understanding. The unit was developed together with an experienced in-service physics teacher, repeatedly pretested in classrooms, and accompanied by a teacher training designed to bridge the theory-practice gap. In a quasi-experimental classroom study, we wanted to determine whether students who received instruction on the basis of this unit (CogAct instruction) from their teachers showed a better understanding of Newtonian mechanics than those who received regular instruction.

We analyzed immediate (posttest) effects and long-term (follow-up) effects after three months on (1) conceptual understanding of core concepts of mechanics and on (2) quantitative problem solving (physics word problems requiring the application of the formulae that model the quantitative relationship between the concepts). We hypothesized that learners with CogAct instruction would outperform those with regular instruction on the test of conceptual understanding at both measurement points. In line with existing research, which showed that a better conceptual understanding also improves quantitative problem solving, we hypothesized that students in the CogAct classes would catch up with students in the regular classes in the quantitative problem-solving test, even though less time is spent on practicing quantitative problem solving in CogAct instruction. We hence did not expect differences between the two instructional conditions in quantitative problem solving at any time of measurement.

Moreover, we hypothesized that female students in general benefit more from CogAct instruction than male students, resulting in a smaller gender gap after CogAct instruction than after regular instruction. As the results by Lorenzo et al. (2006) suggest, we particularly expected a boost in the performance of female students with above-average intelligence in CogAct classes.

Method

Altogether, eight 10th Grade physics classes and four teachers from four Gymnasium schools in German-speaking Switzerland participated in our quasi-experimental study, which was approved by the ethics committee of the ETH Zurich. The Gymnasium is the highest track of the public school system that provides secondary and high school education to students ages 12 to 19 who show good to very good performance in primary school. Depending on the region, the Gymnasium starts at Grade 7 or 9 and ends in Grade 13. Approximately 20% to 25% of all Swiss students attend the Gymnasium, and the final diploma (Matura) allows access to universities. Gymnasium students are comparable to U.S. high school students attending college preparatory classes.

The four teachers involved in our study were each in charge of two classes of the same age group. One of these parallel classes was randomly assigned to the CogAct instruction condition, whereas the other class was taught mechanics by the same teacher in his or her traditional manner (regular instruction). This design allowed us to control for the specific influence of each teacher. Moreover, the parallel classes not only shared the physics teacher but also learned in a highly comparable environment. With our quasi-experimental design, we implemented a conservative method of testing the added value of CogAct instruction.

The Teachers

The four Gymnasium physics teachers were recruited from a pool of schools that were already cooperating with us. We selected teachers who were designated to teach an introductory course in Newton's mechanics in two 10th Grade parallel classes. The teachers agreed to participate in a CogAct training program and to teach one randomly assigned class according to the CogAct instructional unit and the other class in their traditional manner. All of the four teachers hold a master's degree in physics, and three of them also hold a PhD degree. Their age varied between 40 and 56 years, and three of them were male. In addition to their physics degrees, all four teachers hold a university degree from teacher education programs, and they had been teaching physics at the Gymnasium for 3 to 17 years.

Student Sample

With one student excluded who did not receive parental written consent to participate in the study, the final sample consisted of 172 (92 females, 54%) students (mean age $M = 15.96$ years, $SD = 0.96$ years). Eight students missed two and 34 students missed one of the pre, post, or follow-up assessments. Their data were treated as missing values. Eighty-seven of the students (mean age $M = 16.00$ years, $SD = 0.81$ years; 48 females, 55%) received CogAct instruction, and 85 students (mean age $M = 15.92$ years, $SD = 1.10$ years; 44 females, 52%) received regular instruction. All students and their parents were informed about the study, and the parents' written consent was obtained.

Swiss Gymnasium students can decide whether they want to have a stronger focus on science, technology, engineering, and mathematics (STEM) subjects or language and social science subjects (in the following, referred to as non-STEM focus) in Grades 10 to 13 of the Gymnasium. However, irrespective of the focus, the same core subjects (among them physics) are taught, leading to the same final high school diploma (Matura). The focus affects the range of topics that are covered and the number of subject-specific courses. During this study, however, all participating students received the same predefined number of physics lessons. Because Newtonian mechanics is dealt with early in the physics curriculum, the focus that had just been chosen had not yet influenced the students' physics literacy. Independent of the focus, the participating students entered the study with a highly comparable background regarding their prior formal educational experiences in physics. We nonetheless considered the students' focus in our analyses to control for self-selection effects.

Sixty-five percent of all regularly taught students and 70% of all students in the CogAct instruction condition specialized in non-STEM subjects. Seventy-seven percent of all female students and 51% of all male students had chosen a non-STEM focus in the regular instruction condition. In the CogAct instruction condition, this was the case for 81% of all female students and 56% of all male students.

Measures

Demographic variables. The students' age, gender, and focus at school on non-STEM subjects or STEM subjects were assessed with an online questionnaire administered before the intervention started.

Intelligence. Intelligence was measured with the Set II score of Raven's Advanced Progressive Matrices (maximum score = 36; split-half reliability Guttman's λ 4 = .84; Raven, Raven, & Court, 1992), which is one of the most common nonverbal intelligence tests. For each of the 36 problems in the test, an incomplete graphical pattern must be completed by choosing one of eight alternative segments. Set I was used as training set, and the time on Set II was limited to 40 min. The test was administered by Sarah I. Hofer as a group test in the classrooms following the instructions described in the test's manual.

Conceptual understanding. Conceptual understanding in Newtonian mechanics was measured with the test of basic Mechanics Conceptual Understanding (bMCU; Hofer, Schumacher, & Rubin, 2017), a Rasch-scaled multiple-choice test (for a sample item, see Problem 2 of Figure 1). The test, which resembles the FCI, assesses the conceptual knowledge covered in the CogAct instructional unit (see the chapter "The CogAct Instruction"). An item is scored one point only if all correct answer alternatives are checked and no wrong answer alternatives are checked (for more details on the test, see Hofer et al., 2017). A short version of the test consisting of 11 items (maximum score = 11; λ 4 = .67) was used to assess the students' conceptual understanding prior to the intervention at pretest. At the posttest and follow-up test, the bMCU test was augmented by six additional multiple-choice items that resembled the original bMCU test items. Hence, a maximum of 17 points could be achieved in the resulting bMCU test plus (λ 4 = .60). The six new items required the students to transfer their knowledge to another knowledge domain (e.g., transfer the concept of action and reaction forces from mechanics to magnetism) or to combine what they had learned in the context of complex problem situations with several forces operating (e.g., elevator ride or tug-of-war). These items could be considered impossible to solve correctly without instruction. The problem contexts implemented in the bMCU test (plus) were purposely not discussed during instruction in either the CogAct or the regular classes. Therefore, all items required the students to transfer their conceptual understanding to new situations.

Quantitative problem solving. An experienced physics teacher not involved in the study was asked to create a test along with an evaluation schema that closely resembled standard mechanics examinations. The test included five quantitative problems on the Newtonian mechanics topics covered in the study (see the chapter "The CogAct Instruction") and required the students to read graphs, apply formulae, and make calculations (for a sample test question, see Problem 1 of Figure 1). The specific problems used in the test were again purposely not discussed during instruction in either the CogAct or the regular classes. The test was scored according to an evaluation schema that considered different aspects of the solution (maximum score = 11.25; λ 4 = .81). The students' solutions were not merely categorized as correct or wrong. Each single step in the solution process was scored with 0.25 points. Two independent raters coded 32 pilot tests with five quantitative physics problems according to the evaluation schema. The intraclass correlation coefficient (ICC) confirmed high interrater agreement (ICC = .91). Hence, one of the two raters coded all tests according to the evaluation schema.

Procedure

After the teacher training, which is described later, and one week before each of the eight participating classes started with Newtonian mechanics, the teachers received a link to the online questionnaire containing the demographic variables, which had to be forwarded to the students. The students were instructed to complete the questionnaire within one week. Immediately before each of the eight participating classes started with Newtonian mechanics, Sarah I. Hofer presented the short version of the test assessing students' conceptual understanding in introductory Newtonian mechanics, which had to be completed in 30 min. Afterward, the classes received 18 lessons (45 min each, two lessons per week, spread over 10 to 12 weeks due to holidays) of CogAct instruction or regular instruction, respectively.

Immediately after the 18 lessons, the students' conceptual understanding and quantitative problem-solving performance were assessed at posttest by Sarah I. Hofer. The quantitative problem-solving test immediately followed the assessment of conceptual understanding (i.e., the bMCU Test plus). Both assessments had to be completed in 45 min (i.e., one lesson). Sarah I. Hofer made sure that the students started to work on the quantitative problem-solving test at least 10 min before the end of the lesson. Three months after the completion of the 18 lessons, at the follow-up test, Sarah I. Hofer once again elicited the students' conceptual understanding and quantitative problem-solving performance. At any time between the posttest and follow-up test (each teacher could choose a convenient date), Sarah I. Hofer administered the intelligence test in each of the eight classes. There was no overlap in the type of problems presented in the intelligence test and the performance tests, ruling out mutual learning effects. All teachers were requested to schedule their main regular exam as close to the study's posttest as possible to ensure comparable external learning conditions at posttest across classes.

The CogAct Instruction

Adding to the already high demands of regular classroom interaction, teachers and students may struggle with the adaptation to unfamiliar methods and to a different structuring of the content. It is thus far from certain that instructional elements that have been successful in controlled studies also result in learning benefits when implemented by real teachers during instruction in real classroom situations (see, e.g., Guskey, 2002; Murphy & Cromley, 2015; Newcombe et al., 2009; Remillard, Herbel-Eisenmann, & Lloyd, 2011).

The CogAct instruction and the corresponding teacher training were thus developed together with an experienced in-service physics teacher (third author of this article). In the sense of design experiments (e.g., Brown, 1992; Cobb, Jackson, Smith, Sorum, & Henrick, 2013), individual elements and the entire instructional unit were repeatedly implemented in his and his colleagues' classrooms. On the basis of these experiences, elements of the unit were revised and adapted to optimally support student understanding and facilitate implementation on the part of the teachers by integrating the cognitively activating methods into the instructional routine. In response to experiences in the classrooms, for example, the order of activities during a lesson, the wording of an instructional text, or the time specifications in the lesson plans were modified. Altogether, the development of the instructional unit took more than 3 years.

The resulting CogAct instructional unit consists of six parts encompassing a total of 16 lessons (15 content-related lessons and one summary lesson). Because we added two extra lessons as buffer time, the unit examined in this study encompasses 18 lessons. The six parts ("inertia and motion," "mass and weight," "force and acceleration," "balance of forces," "reciprocal action," and "Newton's axioms") are arranged in a way such that each topic follows naturally from the preceding topic to help the students build meaningful knowledge. A CogAct instructional manual, which includes all necessary teaching materials, serves as a guideline for the teachers and structures the implementation of the CogAct instruction.

The five evidence-based cognitively activating methods are supposed to help the learners activate relevant prior knowledge (all methods), overcome unfavorable prior knowledge (generating solutions to novel problems prior to instruction, comparing and contrasting, and self-explanation prompts), build new conceptual knowledge (inventing with contrasting cases, comparing and contrasting), and rework and elaborate their knowledge (self-explanation prompts and metacognitive questions). By concentrating instructional efforts and time on developing a conceptual understanding of the contents, teachers encourage CogAct instruction students to use their conceptual knowledge to understand how concepts translate into formulae and quantitative problem-solving routines. This is realized by introducing formulae only after the underlying conceptual knowledge has been acquired. To give an example, students are instructed to discover the concept of velocity and its unit by describing and comparing two situations that illustrate the movements of two objects (inventing with contrasting cases). Only afterwards is the formula $v = d/t$ given.

Teachers can often choose from several proposed methods or can modify the suggested methods, as long as the idea behind the respective method is retained. For instance, the method of generating solutions to novel problems is recommended as a useful method to start each of the 15 content-related lessons. However, teachers are also free to choose another cognitively activating method (e.g., inventing with contrasting cases) to introduce a new topic. The methods of inventing with contrasting cases and comparing and contrasting are exemplified in nine lessons, and it is suggested to implement at least one of them in each of these lessons. A large choice of self-explanation prompts and metacognitive questions is included in the last lesson of each of the five content-related parts. However, teachers are encouraged to use these two methods, which are straightforward to adapt and to implement, whenever appropriate. There are a few suggestions for practicing quantitative problem-solving routines in the manual that can be implemented as required. Only little time is spent on such activities.

The CogAct Teacher Training

A 2-day training was carried out by all of the authors, including the in-service physics teacher. Although the CogAct instruction and its theoretical foundations were presented on the first day (6 hr), the second day (6 hr) required the teachers to contemplate the implementation of the CogAct instruction in their classrooms. Consequently, on the first day, the teachers were informed about the rationale underlying the different cognitively activating instructional methods and were introduced to the CogAct instructional manual. The structure and usage of the manual, including the attached additional worksheets and PowerPoint slides, were described. We emphasized that the purpose of all five cognitively activating methods was to help learners activate, rethink, and adapt existing knowledge and thereby construct new knowledge. Throughout the training, it was clearly communicated that the cognitively activating instructional methods are the "active ingredients" of the CogAct instruction and must not be omitted. The training was supposed to prepare the teachers to adjust the implementation of the CogAct instruction to their own teaching preferences while remaining in keeping with its core ideas. Accordingly, on the second day of the training, the teachers could discuss their interpretation of the manual, including the cognitively activating methods, and solutions to several important questions concerning the implementation of the CogAct instruction were developed together to ease later implementation in the classroom. These questions included, for instance, what elements of the CogAct instruction can and cannot be omitted, or how much leeway is necessary to adapt the teaching to the students' needs. The teachers were informed that they were free to choose from several proposed methods or modify the suggested methods, as long as the idea behind the respective method was retained. Their suggestions for modification were shared and discussed. The design of our training was in line with the results of a large review on the effectiveness of teacher professional development by Yoon, Duncan, Lee, Scarloss, and Shapley (2007). According to the review, teacher workshops that result in a successful theory-practice transfer address evidence-based practices, provide active-learning experiences, and allow teachers to adapt the instructional methods to their specific classroom needs.

To increase the teachers' cooperation and commitment, we clearly communicated their role in the research project and emphasized that we wanted to test the added value of the means of cognitive activation under realistic classroom conditions by comparing CogAct instruction to a serious competitor, their regular physics instruction. By stressing the potential of the cognitively activating methods, on the one hand, and informing them about the function of their regular mechanics instruction in the study design, on the other hand, we wanted to avoid any conflicts of interest or feelings of threat that the teachers' previous way of teaching might be called into question.

Because the specific sequencing of the topics was intended to promote the active incremental construction of meaningful knowledge, the teachers were further requested to stick to the given order. All teachers received a protocol documenting the results of the discussions that occurred during the training.

The Regular Instruction

In the teacher training, we presented all mechanics topics covered in the CogAct instruction that had to be taught during the regular instruction as well. In terms of the regular instruction, teachers were told that they should teach introductory Newtonian mechanics as always, with the only restriction that all topics presented had to be covered within the study's time frame of 18 lessons, in an individual order and with individual prioritization.

Implementation Fidelity of the Study Design

To make our research design work, we had to ascertain that the same teacher delivered a different type of instruction in the two parallel classes, depending on the condition to which the class was assigned. Under the CogAct instruction condition, teachers were supposed to use the CogAct instructional manual and implement means of cognitive activation presented in the manual. In contrast, under the regular instruction condition, teachers should refrain from applying any of the examples presented in the manual. Although we were confident that the teachers had understood their role in the research project and were well prepared to implement the CogAct instruction in their classrooms, we gained additional information on the actual implementation of the mechanics classes in both conditions from three sources: (a) classroom observations, (b) semistructured interviews, and (c) teachers' reports.

Classroom observations. Sarah I. Hofer visited two lessons in each of the four CogAct classes and two lessons in each of the four regular instruction classes without prior notice. One visit took place in the first and the other in the second half of the 18 lessons. Each teacher was thus observed four times. We used protocol forms to record the general didactic phases of the lesson, including repetition, introduction, teaching of new content, elaboration of the content, or practice of procedures (see Seidel, Prenzel, Duit, & Lehrke, 2003), together with content descriptions and activities that referred to these phases. For each of these phases, we also noted any cognitively activating instructional method that was used by the teacher. For instance, the phase "repetition" with the content description "on parallelogram of forces" could be followed by the phase "introduction" on "barycenter of a clothes hanger" using the cognitively activating method "generating solutions to a novel problem prior to instruction." All of the lessons protocolled in the CogAct classrooms reflected specific lessons described in the CogAct instructional manual, including suggested cognitively activating methods. On the other hand, the lessons recorded in the regular instruction classrooms did not correspond to any lesson described in the CogAct instructional manual. Furthermore, the teachers did not use cognitively activating methods in the observed regular instruction lessons, with the exception of the method of generating solutions to novel problems prior to instruction, which was implemented to introduce a new topic in two of the eight regular lessons that were observed, and the method of comparing and contrasting, which was also implemented in two classes in the phase "elaboration of the content." In all cases, these cognitively activating methods were not based on examples from the CogAct instructional manual. Two uninvolved and uninformed research assistants who received the CogAct instructional manual for comparison could correctly assign all protocols to the corresponding instructional condition.

Semistructured interviews. Each teacher was interviewed three times by Sarah I. Hofer. The first two interviews took place after the first and the fourth (i.e., last) classroom visit, and the third interview took place at the very end, after the posttest had been applied. Teachers were asked whether they had encountered any difficulties in applying the CogAct instructional manual and the embedded cognitively activating methods and therefore refrained from doing so. At all interviews, all four teachers indicated that they based their teaching on the manual and that they implemented cognitively activating methods as described in the manual. Two of the four teachers once mentioned a decrease in authenticity and fluency in the teaching process (i.e., more frequent intermissions and a feeling of uncertainty in discussions of conceptual knowledge). The teachers were further asked whether they followed the sequence of topics prescribed by the CogAct instructional manual. No deviations were reported. In the last interview, the teachers were asked to provide the textbooks and other teaching resources they had used in their regular instruction classes. Four well-known textbooks in German-speaking countries were mentioned, all of which have a strong focus on quantitative problems, while cognitively activating principles to support conceptual understanding are not explicitly addressed in these books. Most of the exercises in the teaching resources named by the teachers involved practicing strategies for solving different types of quantitative problems. All of the teachers confirmed that the mechanics topics were covered in all classrooms.

Teachers' reports. Teachers were handed a checklist with a total of six activities listed: the five means of cognitive activation and, in addition, quantitative problem-solving activities (i.e., practice of procedures). By completing the list, teachers documented in how many of the 18 lessons they had applied each of the cognitively activating methods and quantitative problem-solving activities, both in their CogAct instruction and in their regular instruction. Results indicated that metacognitive questions were reported in all CogAct lessons by one teacher, in about half of all lessons by two teachers, and in two lessons by one teacher. Self-explanation prompts were reported in 12 to 15 lessons by three teachers and in seven lessons by one teacher under CogAct instruction. Both methods were not reported in the regularly instructed classes, with the only exception being one teacher who indicated the implementation of self-explanation prompts in two regular lessons. Under CogAct instruction, two teachers reported the method of generating solutions to novel problems prior to instruction in 15 and 16 lessons, and the other two teachers reported this method in about half of all CogAct lessons, it was reported in about half of all lessons and less (to not at all) under regular instruction. Two teachers reported the methods of inventing with contrasting cases and comparing and contrasting in about two thirds of their CogAct lessons each, although the other two teachers reported each of these methods in approximately one quarter of the CogAct lessons. Under regular instruction, inventing with contrasting cases was not mentioned at all by two teachers and was reported in two lessons by the other two teachers. Comparing and contrasting was mentioned in seven regular lessons by two teachers and in two regular lessons and not at all, respectively, by the remaining two teachers. The variation in the frequency of the implementation of the cognitively activating methods across teachers reflects the leeway in the choice of the methods embedded in the CogAct instructional manual. Importantly, for each individual teacher, it holds that he or she used all of the cognitively activating methods more regularly in the CogAct instruction than in the regular instruction condition.

The reported number of lessons in which quantitative problem-solving activities were used also differed between both conditions. Although the teachers reported that they implemented activities that aim at the practice of quantitative procedures in at least seven to 16 of the regular lessons, such activities were reported in two to a maximum of 11 of the CogAct lessons. Again, for each individual teacher, it holds that he or she reported quantitative problem-solving activities more often under regular than under CogAct instruction. The three sources of implementation fidelity corresponded and indicated the teachers' cooperation by carefully applying two different types of instruction, depending on the condition.

Results

Data Analysis

Mplus Version 7.11 (Muthén & Muthén, 2012) was used for all analyses. To answer our research questions, different regression models were analyzed. We conducted robust maximum likelihood estimation (i.e., MLR) to potentially correct fit statistics and all parameter estimates' standard errors for leptokurtic or platykurtic data (see, e.g., Heck & Thomas, 2015; Hox, Moerbeek, & van de Schoot, 2010). The *p* values resulting from the significance tests of the regression coefficients may be distorted since they are based on the assumption of normally distributed parameters. To obtain more stable *p* values, we performed log-likelihood tests that compare less restrictive models (i.e., the regression coefficient is estimated freely) to more restrictive (i.e., the regression coefficient is set to zero) but nested models (for detailed information on the test, see UCLA & the Statistical Consulting Group, 2014). A significant *LL* *p* value would then suggest that the regression coefficient significantly contributed to the regression model and should not be set to zero. Missing values were estimated using full information maximum likelihood (FIML; see, e.g., Johnson & Young, 2011).

Descriptive Statistics

Table 1 summarizes means and standard deviations of the intelligence test, the performance tests, and the continuous demographic variables for both instructional conditions for the total sample, as well as for female and male students separately.

General Effects of CogAct Instruction

To investigate the general effectiveness of the CogAct instruction in terms of immediate (posttest) and long-term (follow-up) effects, regression analyses were conducted. The conceptual understanding and quantitative problem-solving posttest and follow-up scores were regressed on condition (0 = regular instruction, 1 = CogAct instruction) and the five control variables: age (in years), gender (0 = female, 1 = male), focus (0 = non-STEM, 1 = STEM),

Table 1

Condition-Specific Means and Standard Deviations of Major Continuous Study Variables for the Total Sample as well as for Female and Male Students Separately

	Instructional condition						
	CogAct			Regular			
Variable	*M*	*SD*	*N*	*M*	*SD*	*N*	Scale
Intelligence (Set II score of Raven's Advanced Progressive Matrices)							
Total	27.36	4.56	87	27.41	4.05	85	0–36
Female	26.42	4.41	48	27.45	4.31	44	
Male	28.31	4.58	39	27.37	3.82	41	
Pre							
Age							
Total	16.00	.81	87	15.92	1.10	85	
Female	15.87	.76	48	16.09	.94	44	
Male	16.15	.84	39	15.73	1.25	41	
Prior conceptual understanding							
Total	2.95	1.57	79	2.77	1.59	81	0–11
Female	2.60	1.33	47	2.52	1.50	44	
Male	3.47	1.76	32	3.05	1.67	37	
Post							
Conceptual understanding							
Total	6.57	2.96	77	5.57	2.23	82	0–17
Female	5.49	2.01	41	5.09	1.94	43	
Male	7.81	3.39	36	6.10	2.44	39	
Quantitative problem solving							
Total	4.63	3.20	77	3.94	3.07	82	0–11.25
Female	3.93	2.70	41	3.72	3.03	43	
Male	5.43	3.56	36	4.18	3.14	39	
Follow-up							
Conceptual understanding							
Total	5.62	2.99	69	5.09	2.19	78	0–17
Female	4.87	2.28	39	4.16	1.48	38	
Male	6.60	3.52	30	5.98	2.39	40	
Quantitative problem solving							
Total	3.10	2.96	69	3.73	2.69	78	0–11.25
Female	2.15	2.13	39	3.68	2.65	38	
Male	4.41	3.45	30	3.78	2.76	40	

intelligence, and prior conceptual understanding (i.e., measured at pretest). Students in the CogAct and the regular classes did not differ in terms of the control variables' means or proportions, respectively (all *p*s $\geq$ .45). The five control variables were included in the regression analyses to control for variations on the individual student level that could not be attributed to the intervention but had to be considered additional predictors of learning due to the quasi-experimental setting. We took the students' prior conceptual understanding into account but did not include prior quantitative problem-solving performance in basic Newtonian mechanics as an additional control variable because, without instruction, the majority of students have no knowledge of the calculation routines necessary to solve the quantitative problems. Intuitive conceptual knowledge on basic Newtonian mechanics consisting of correct ideas, synthetic ideas (an amalgamation of correct ideas and naïve beliefs), and naïve beliefs, however, already exists before formal instruction.

We hypothesized that learners with CogAct instruction would outperform those with regular instruction on conceptual understanding, whereas no differences were expected in quantitative problem-solving skills. Accordingly, we expected significant positive regression coefficients of condition predicting the conceptual understanding posttest and follow-up scores, and nonsignificant regression coefficients of condition predicting the quantitative problem-solving posttest and follow-up scores. The results of the regression analyses for the posttest and follow-up test measures of conceptual understanding and quantitative problem solving are presented in Table 2.

At posttest, being in the CogAct condition had a significant positive effect on both conceptual understanding (β = 0.19, *SE* = 0.06, 95% CI [0.07, 0.32]) and quantitative problem solving (β = 0.14, *SE* = 0.07, 95% CI [0.00, 0.28]). These effects implied an advantage of 1.03 points (95% CI [0.32, 1.73]) in the conceptual understanding test and an advantage of 0.87 points (95% CI [−0.02, 1.77]) in the quantitative problem-solving test for the students in the CogAct classes. At the follow-up test, students benefited significantly from CogAct instruction only in terms of conceptual understanding (β = 0.13, *SE* = 0.07, 95% CI [0.00, 0.26]), indicating an advantage of 0.68 points (95% CI [−0.02, 1.39]) in the conceptual understanding test. These effects were present after controlling for the five individual student variables.

Table 2

Parameter Estimates Based on the Regression Model for Post Data and the Regression Model for Follow-Up Data

Variable	Post data			Follow-up data		
	b	*SE*	*LL p*	*b*	*SE*	*LL p*
DV = Conceptual understanding						
Condition (0 = regular, 1 = CogAct)	1.03	.36	<.01	.68	.36	<.05
Control variables						
Age	.09	.22	.67	.11	.21	.59
Gender (0 = female, 1 = male)	1.06	.37	<.01	1.24	.37	<.01
Focus (0 = non-STEM, 1 = STEM)	.52	.42	.21	.46	.45	.28
Intelligence	.07	.04	.10	.05	.03	.18
Prior conceptual understanding	.57	.14	<.001	.70	.14	<.001
DV = Quantitative problem solving						
Condition	.87	.46	<.05	−.50	.45	.28
Control variables						
Age	−.15	.26	.56	−.09	.25	.72
Gender	.33	.48	.50	.53	.45	.25
Focus	.31	.53	.56	.88	.54	.09
Intelligence	.21	.04	<.001	.05	.04	.21
Prior conceptual understanding	.36	.17	<.05	.36	.16	<.05

Note. *LL p* = *p* values that resulted from the log-likelihood tests; DV = dependent variable; STEM = science, technology, engineering, and mathematics.

Effects of CogAct Instruction as a Function of Gender and Intelligence

To investigate whether CogAct instruction reduces the gender gap, we extended the regression models described earlier by including the interaction between gender and condition. We hypothesized that CogAct instruction enables female students to better exploit their untapped cognitive potential and therefore allows them to catch up with male students. Accordingly, we expected a negative significant interaction between gender and condition.

This expectation was not confirmed. The interaction between gender and condition significantly predicted only the quantitative problem-solving follow-up score (*LL p* < .05), although not in the expected direction (*b* = 2.13, *SE* = 0.91; β = 0.32, *SE* = 0.13). The significant interaction term implied an advantage of 2.13 points (95% CI [0.34, 3.92]) on the quantitative problem-solving follow-up test for male students in the CogAct condition. The interaction between gender and condition did not reach significance for any of the other performance measures (all *LL p*s $\geq$ .15).

To more specifically investigate whether CogAct instruction is particularly beneficial for female students with high intellectual potential, we ran multiple group regression analyses, with gender as the grouping variable and intelligence, condition, and the interaction between intelligence and condition as predictor variables. Again, the conceptual understanding and quantitative problem-solving posttest and follow-up scores served as dependent variables. We hypothesized a stronger boost in the performance of female students than for male students with increasing intelligence in CogAct classes, which would manifest itself in a positive significant interaction between intelligence and condition only in the group of the female students.

The results revealed a positive significant interaction term in the group of the female students for posttest conceptual understanding (*LL p* < .05; *b* = 0.13, *SE* = 0.06, 95% CI [0.01, 0.26]; β = 0.88, *SE* = 0.42, 95% CI [0.06, 1.69]). Although pointing in the expected direction, the interaction between intelligence and condition did not significantly predict posttest quantitative problem solving (*LL p* = .10; *b* = 0.16, *SE* = 0.10), follow-up conceptual understanding (*LL p* = .18; *b* = 0.11, *SE* = 0.07), or follow-up quantitative problem solving (*LL p* = .26; *b* = 0.11, *SE* = 0.10) in the group of the female students. Irrespective of the performance measure and the time point investigated, in the group of the male students, the interaction between intelligence and condition was not significant, with the lowest *LL p* value resulting for follow-up conceptual understanding (*LL p* = .05).

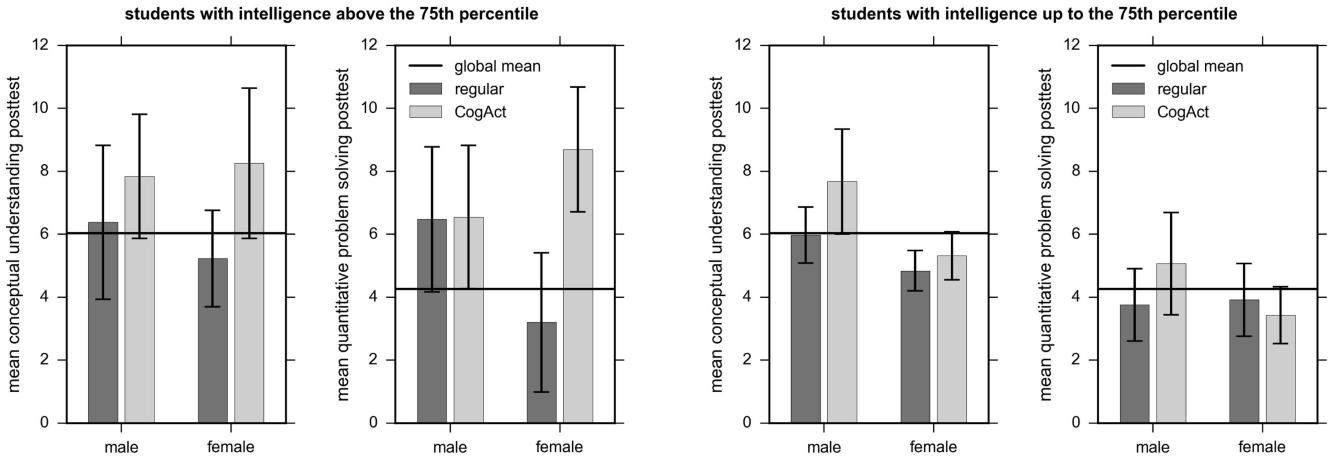


Figure 6. Mean conceptual understanding and quantitative problem-solving posttest scores for females and males with an intelligence score above the 75th percentile (left side) and all other students (right side), as a function of instructional condition. The global mean indicates the mean value of the total sample. Error bars represent the 95% confidence intervals.

To complement these results with descriptive and not linearly modeled data, the left half of Figure 6 shows the mean conceptual understanding and quantitative problem-solving posttest scores and the 95% confidence intervals only for female ($N_{\text{regular}} = 9$, $N_{\text{CogAct}} = 6$) and male students ($N_{\text{regular}} = 8$, $N_{\text{CogAct}} = 12$) scoring in the upper quartile (intelligence above the 75th percentile) of the intelligence distribution in the student sample[1] as a function of instructional condition. For comparison, the right half of Figure 6 provides the analogous information for all other students (i.e., intelligence up to the 75th percentile). Table 3 lists the corresponding condition-specific means, standard deviations, and 95% confidence intervals of both the posttest and follow-up test performance measures for female and male students scoring in the upper quartile of the intelligence test. The descriptive data of these female students in the CogAct condition indicated high conceptual understanding ($M = 8.25$, $SD = 1.50$) and quantitative problem-solving skills ($M = 8.69$, $SD = 1.25$) at posttest and a rather sharp knowledge decline ($M = 4.80$, $SD = 2.17$ and $M = 3.60$, $SD = 3.38$) at follow-up. Despite this decline, they still performed better than their female counterparts in the regular instruction condition (see Table 3).

Discussion

Research from the 1980s has demonstrated that high school and college students can pass through physics education, even with good grades, by having memorized some facts and solution algorithms, whereas their conceptual understanding remains essentially unchanged (Clement, 1982). This phenomenon has been documented for several content areas, but first and foremost for the area of mechanics, where naïve concepts of force or inertia notoriously hamper the processing and adoption of valid scientific explanations (Hake, 1998). Informed by the theoretical framework of conceptual change, concerted efforts in research on learning and instruction have been made to develop and evaluate instructional inputs that can help students become aware of the limits of their naïve concepts and to acknowledge the explanatory power of the scientific concepts offered in class.

Particularly in the case of mechanics, deliberate instructional effort is required to encourage conceptual change, because many mechanics-related concepts, such as active and reactive forces or equilibrium forces, are initially counterintuitive to nearly everyone. It is therefore uncontested among educational researchers that, especially when Newtonian mechanics is to be taught, more time must be devoted to learning activities that support learners' qualitative conceptual understanding. In the past decades, several alternatives to traditional tell-and-practice instruction have been developed that help learners activate, restructure, and extend their prior knowledge. In our study, we made use of five instructional means to enhance students' conceptual learning, which were included in a comprehensive 18-lesson unit on Newtonian mechanics.

The effectiveness of each of these five means (generating solutions to novel problems, inventing with contrasting cases, comparing and contrasting, self-explanation prompts, and metacognitive questions) has already been confirmed in controlled experiments in the lab and in classrooms. These studies typically investigate the effects of single methods in short instructional units by focusing on a single concept, a particular type of problem, or a narrow section of a broader content area. In real physics classes, however, instructional units cover a broader content area composed of various interrelated concepts, such as in the case of introductory Newtonian mechanics. To successfully teach a broader content area composed of interrelated concepts, teachers could benefit from a comprehensive instructional unit that includes various suggestions for how to help their students replace scientifically inappropriate explanations with valid ones. In this study, we thus investigated the potential of such a comprehensive instructional unit on Newtonian mechanics that combines several evidence-based cognitively activating methods (CogAct instruction).

---

[1] The pattern of results of this study did not change when, as an alternative, the cut-off was set at one standard deviation above the sample mean and higher.

Table 3
Condition-Specific Means, Standard Deviations, and 95% Confidence Intervals (CIs) of Performance Measures for the Sample of Females and Males With an Intelligence Score Above the 75th Percentile

Measure	Condition	Gender	*M*	*SD*	95% CI Lower bound	95% CI Upper bound
Conceptual understanding posttest	Regular	Female	5.22	1.99	3.70	6.75
		Male	6.38	2.92	3.93	8.82
	CogAct	Female	8.25	1.50	5.86	10.64
		Male	7.83	3.10	5.86	9.80
Conceptual understanding follow-up	Regular	Female	3.86	1.46	2.50	5.21
		Male	7.00	2.27	5.10	8.90
	CogAct	Female	4.80	2.17	2.11	7.49
		Male	6.10	3.03	3.93	8.27
Quantitative problem solving posttest	Regular	Female	3.19	2.88	.98	5.41
		Male	6.47	2.76	4.16	8.78
	CogAct	Female	8.69	1.25	6.70	10.67
		Male	6.54	3.59	4.26	8.82
Quantitative problem solving follow-up	Regular	Female	2.89	2.77	.33	5.45
		Male	5.41	3.25	2.69	8.12
	CogAct	Female	3.60	3.38	−.59	7.79
		Male	5.00	4.03	2.12	7.88

Our quasi-experimental design revealed that students who had been taught by in-service teachers according to our CogAct instructional unit showed better performance than students of a parallel class who had received regular instruction by the same teachers. Immediately after the 18 lessons of physics instruction, students with CogAct instruction outperformed students with regular instruction in terms of their conceptual understanding in basic Newtonian mechanics. Even 3 months after the instruction, the conceptual understanding of the students who had received CogAct instruction exceeded the conceptual understanding of the students with regular instruction.

In addition to advantages in conceptual understanding, students in the CogAct condition also showed significantly better performance in quantitative problem solving at posttest than students who underwent regular instruction, even though the practice of quantitative problem solving played a more significant role in the regular instruction classes than in the CogAct classes. This finding suggests that the mastery of quantitative problems is promoted by conceptual understanding, but not vice versa (e.g., Hake, 1998). Three months after instruction, the quantitative problem-solving performance had adjusted in both groups.

We also expected CogAct instruction to reduce the gender gap because it would allow female students to unfold their untapped cognitive potential for conceptual understanding in physics. This prediction was, however, only partly confirmed. With the exception of the quantitative problem-solving performance at the follow-up test, both male and female students, on average, reached higher scores in the CogAct condition than in the regular instruction condition. Unexpectedly, however, with respect to the quantitative problem-solving performance at the follow-up test, male students benefited even more from CogAct instruction than female students. CogAct instruction seems to support more male than female students in enduringly integrating qualitative and formalized quantitative aspects of concepts, perhaps partly reflecting female students' tendentially higher aversion to quantitative problem solving (Zohar, 2006; Zohar & Sela, 2003).

Although we could not confirm a reduction of the gender gap for the entire sample, we did find evidence for specific effects of CogAct instruction on the performance of female students with above-average intelligence. Interaction analyses indicated that these females had a significantly better conceptual understanding at posttest when taught by CogAct instruction than by regular instruction. When analyzing the descriptive statistics of the group of students scoring above the 75th percentile of the intelligence test in our sample, a similar advantage was also found in terms of posttest quantitative problem solving. The discrepancy between the two instructional conditions was less pronounced for male students. Female students with above-average intelligence in the CogAct condition even caught up with the male students with above-average intelligence on posttest performance measures. CogAct instruction seems to particularly allow female learners who clearly score above average in an intelligence test to exploit more of their thus-far untapped cognitive resources. In the follow-up tests, however, the performance of these female students decreased again in the CogAct condition, although never below the level of their female counterparts in the regular instruction condition. These results may help us better understand the inconsistent results concerning gender-specific effects of instruction with a focus on concepts in mechanics. Although Lorenzo and colleagues (2006) found a decreasing gender gap among Harvard students, Pollock and colleagues (2007) even found an increase in a less-selective group of learners. Analogously, though highly capable female students can rapidly make use of innovative instructional elements, more effort and time may be necessary to stimulate the entire group of female learners to exploit their thus-far untapped cognitive potential in physics.

Methodological Considerations

By comparing the performance of students who received our newly developed instructional unit to the performance of students who received the regular instruction of the same teachers, we implemented an ecologically valid control condition that can be considered a strong competitor to the CogAct instruction condition. The advantages of CogAct instruction over regular instruction could be observed despite the fact that the participating teachers were implementing the CogAct instruction for the first time. Implementation checks revealed that regular instruction only sporadically included the five cognitively activating methods and had a strong focus on quantitative problem solving. On the other hand, the same teachers made extensive use of the cognitively activating methods in their parallel classes assigned to the CogAct instruction condition.

We are aware that the teacher self-report data used to evaluate implementation fidelity might be less accurate than objective classroom observations. However, in vivo or video observations of a larger number (or all) lessons would have interfered with and presumably altered everyday classroom routines. Because in this study we aimed to investigate the benefits of CogAct instruction compared with regular physics instruction of in-service teachers in their real physics classes with high ecological validity, we decided against extensive classroom observations and instead combined self-report data and observation data and checked their correspondence.

Our study was not designed to analyze the impact of the single elements of the CogAct instruction. In our CogAct instruction, several learning activities intentionally formed a tool kit so that teachers could select from a larger pool of cognitively activating methods. By demonstrating the superiority of the CogAct instruction over the regular mode, we could highlight a promising direction for future classroom practice. However, we do not know what features of the instruction were decisive for the students' performance. During the teacher training, the five means of cognitive activation were not presented as techniques or guidelines that had to be strictly adhered to, but rather were explained in light of cognitive theories of human learning. Teachers were supposed to remain in keeping with the CogAct instructional unit's mission, but they had some leeway in how to implement it. In future studies, a larger sample of physics teachers should be observed when implementing the CogAct instruction. Systematic data revealing how each teacher implements specific elements of the CogAct instruction should be gathered, using both questionnaires and in vivo or video observations. A dataset that has the statistical power for multilevel analyses will allow us to identify the elements of the CogAct instructional unit that are decisive for the students' performance. On the basis of these results, CogAct instruction could be further optimized by finding out which features must be implemented by the teachers and which can be handled optionally. Once a pool of teachers has been recruited, it would also be worthwhile to study whether a repeated implementation by the same teachers will lead to more pronounced effects.

With our quasi-experimental study, we wanted to bridge the gap between well-controlled but narrow learning experiments and the implementation of scientifically approved means of instruction by in-service teachers in real classroom contexts. By making use of parallel classes, we could control for teacher effects and therefore run a controlled intervention study with a relatively small number of classrooms. Although access to parallel classes may not always be as easy as it is in the Swiss system, it should be feasible in other countries as well. Such quasi-experimental intervention studies can be considered an intermediate step between laboratory experiments and large-scale studies.

Implications for Future Research and Practice

Our results show that experienced in-service teachers can implement means of cognitive activation that benefit their students within the proper time frame and with limited costs and effort. Providing teachers with content-specific means of cognitive activation that have been shown to be feasible in design experiments and that are embedded into a comprehensive instructional unit that is introduced by a carefully designed teacher training seems worthwhile and can be recommended for future classroom practice. At the same time, however, a short-term change in instructional practice, as induced in our study, is unlikely to produce strong and sustainable changes in students' learning outcomes. It is worth mentioning that the superiority of the CogAct classes in conceptual understanding was significant but not very large. This result is in line with findings of Cheung, Slavin, Kim, and Lake (2017), according to which the effect sizes of secondary school science programs that are tested against a regularly taught control group rarely exceed $d = .20$. Nonetheless, stronger and more sustainable effects are definitely desirable. On average, students having received CogAct instruction gained seven of 17 points in the conceptual understanding posttest, which clearly indicates room for improvement. The gender gap was reduced only for the most intelligent female students. The superiority of CogAct classes in quantitative problem solving at posttest had vanished in the follow-up test. Further effort and research is needed to improve the effect of cognitively activating instruction and activate all learners.

Although we confirmed previous findings, according to which qualitative conceptual understanding has a positive effect on quantitative problem solving (see Chi, Feltovich, & Glaser, 1981; Hardiman, Dufresne, & Mestre, 1989; Heyworth, 1999; Ploetzner et al., 1999; Taconis et al., 2001), our results suggest that the construction of quantitative problem-solving knowledge deserves more attention in future research. It is important to note that the focus on conceptual understanding propagated by learning researchers in past decades should not at all be understood as contempt for quantitative problem solving in physics education. Physics is an exact science, which means that theories and laws are expressed using mathematical equations, and this message must be imparted in secondary school physics classes. Our results show that in the regular and CogAct instruction, the mean scores achieved in the quantitative problem-solving test were less than half of the maximum score. The explicit connections made between conceptual knowledge and the respective quantitative formulae during the CogAct instruction (corresponding to the explicit production rules as described by Anderson & Schunn, 2000 or Chi et al., 1981) boosted the students' quantitative problem-solving performance for a short time, as the posttest results indicate. The follow-up test, however, showed that this effect was not sustainable. In particular, the female students experienced a considerable drop in their quantitative problem-solving performance. What students might need are opportunities for deliberate and meaningful practice involving means of cognitive activation. Activating conceptual knowledge while practicing quantitative problem-solving routines has been successful in constructing and relating conceptual knowledge and quantitative problem-solving knowledge. Particularly promising are self-explanation prompts and metacognitive questions that refer to quantitative problems and require students to think about the concepts underlying the quantitative problems (Berthold & Renkl, 2010; Mevarech & Fridkin, 2006). A revised version of our CogAct instruction and other future instructional units should therefore pay more attention to the meaningful treatment of quantitative problems.

References

Anderson, J. R., & Schunn, C. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. 5, pp. 1–27). Mahwah, NJ: Lawrence Erlbaum.

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review, 22,* 25–40. http://dx.doi.org/10.1007/s10648-010-9124-9

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2,* 141–178. http://dx.doi.org/10.1207/s15327809jls0202_2

Camp, C. W., Clement, J. J., Brown, D., Gonzalez, K., Kudukey, J., Minstrell, J., & Zietsman, A. (1994). *Preconceptions in mechanics: Lessons dealing with students' conceptual difficulties*. Dubuque, IA: Kendall/Hunt.

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology, 21,* 13–19. http://dx.doi.org/10.1016/S0193-3973(99)00046-5

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin, 135,* 218–261. http://dx.doi.org/10.1037/a0014412218

Cheung, A., Slavin, R. E., Kim, E., & Lake, C. (2017). Effective secondary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching, 54,* 58–81. http://dx.doi.org/10.1002/tea.21338

Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152. http://dx.doi.org/10.1207/s15516709cog0502_2

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63,* 1–49. http://dx.doi.org/10.3102/00346543063001001

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50,* 66–71. http://dx.doi.org/10.1119/1.12989

Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. *National Society for the Study of Education Yearbook, 112,* 320–349.

DiSessa, A. A., & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. V. Goldman (Eds.), *Thinking practices in mathematics and science learning* (pp. 155–187). New York, NY: Routledge.

Dole, J. A., & Sinatra, G. M. (1998). Reconceptalizing change in the cognitive construction of knowledge. *Educational Psychologist, 33*(2–3), 109–128. http://dx.doi.org/10.1080/00461520.1998.9653294

Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25,* 671–688. http://dx.doi.org/10.1080/09500690305016

Fischer, H. E., Labudde, P., Neumann, K., & Viiri, J. (Eds.). (2014). *Quality of Instruction in Physics: Comparing Finland, Switzerland, and Germany*. Münster, Germany: Waxmann Verlag.

Fulmer, G. W., Liang, L. L., & Liu, X. (2014). Applying a force and motion learning progression over an extended time span using the Force Concept Inventory. *International Journal of Science Education, 36,* 2918–2936. http://dx.doi.org/10.1080/09500693.2014.939120

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1–38. http://dx.doi.org/10.1016/0010-0285(83)90002-6

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and teaching, 8,* 381–391. http://dx.doi.org/10.1080/135406002100000512

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66,* 64–74. http://dx.doi.org/10.1119/1.18809

Halloun, I. A., & Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53,* 1056–1065. http://dx.doi.org/10.1119/1.14031

Halpern, D. F. (2014). It's complicated—In fact, it's complex: Explaining the gender gap in academic achievement in science and mathematics. *Psychological Science in the Public Interest, 15,* 72–74. http://dx.doi.org/10.1177/1529100614548844

Hänze, M., & Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics: An experimental study comparing cooperative learning and direct instruction in 12th grade physics classes. *Learning and Instruction, 17,* 29–41. http://dx.doi.org/10.1016/j.learninstruc.2006.11.004

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition, 17,* 627–638. http://dx.doi.org/10.3758/BF03197085

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. New York, NY: Routledge.

Heller, K. A., Finsterwald, M., & Ziegler, A. (2010). Implicit theories of mathematics and physics teachers on gender-specific giftedness and motivation. In K. A. Heller (Ed.), *Munich studies of giftedness* (pp. 239–252). Berlin, Germany: LIT.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30,* 141–158. http://dx.doi.org/10.1119/1.2343497

Heyworth, R. M. (1999). Procedural and conceptual knowledge of expert and novice students for the solving of a basic problem in chemistry. *International Journal of Science Education, 21,* 195–211. http://dx.doi.org/10.1080/095006999290787

Hofer, S. I., Schumacher, R., & Rubin, H. (2017). The test of basic Mechanics Conceptual Understanding (bMCU): Using Rasch analysis to develop and evaluate an efficient multiple-choice test on Newton's mechanics. *International Journal of STEM Education, 4* (18). Advance online publication.

Hofer, S. I., & Stern, E. (2016). Underachievement in physics: When intelligent girls fail. *Learning and Individual Differences, 51,* 119–131. http://dx.doi.org/10.1016/j.lindif.2016.08.006

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2010). *Multilevel analysis: Techniques and applications*. New York, NY: Routledge.

Huber, S. A., Häusler, J., Jurik, V., & Seidel, T. (2015). Self-underestimating students in physics instruction: Development over a school year and its connection to internal learning processes. *Learning and Individual Differences, 43,* 83–91. http://dx.doi.org/10.1016/j.lindif.2015.08.021

Johnson, D. R., & Young, R. (2011). Toward best practices in analyzing datasets with missing data: Comparisons and recommendations. *Journal of Marriage and the Family, 73,* 926–945. http://dx.doi.org/10.1111/j.1741-3737.2011.00861.x

Jurik, V., Gröschner, A., & Seidel, T. (2013). How student characteristics affect girls' and boys' verbal engagement in physics instruction. *Learning and Instruction, 23,* 33–42. http://dx.doi.org/10.1016/j.learninstruc.2012.09.002

Jurik, V., Gröschner, A., & Seidel, T. (2014). Predicting students' cognitive learning activity and intrinsic learning motivation: How powerful are teacher statements, student profiles, and gender? *Learning and Individual Differences, 32,* 132–139. http://dx.doi.org/10.1016/j.lindif.2014.01.005

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38,* 1008–1022. http://dx.doi.org/10.1111/cogs.12107

Keller, M. M., Neumann, K., & Fischer, H. E. (2017). The impact of physics teachers' pedagogical content knowledge and motivation on students' achievement and interest. *Journal of Research in Science Teaching, 54,* 586–614. http://dx.doi.org/10.1002/tea.21378

Kim, E., & Pak, S. J. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics, 70,* 759–765. http://dx.doi.org/10.1119/1.1484151

Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology, 100,* 716–725. http://dx.doi.org/10.1037/0022-0663.100.3.716

Labudde, P., Herzog, W., Neuenschwander, M., Violi, E., & Gerber, C. (2000). Girls and physics: Teaching and learning strategies tested by classroom interventions in grade 11. *International Journal of Science Education, 22,* 143–157. http://dx.doi.org/10.1080/095006900289921

Labudde, P., Reif, F., & Quinn, L. (1988). Facilitation of scientific concept learning by interpretation procedures and diagnosis. *International Journal of Science Education, 10,* 81–98. http://dx.doi.org/10.1080/0950069880100108

Leppävirta, J., Kettunen, H., & Sihvola, A. (2011). Complex problem exercises in developing engineering students' conceptual and procedural knowledge of electromagnetics. *IEEE Transactions on Education, 54,* 63–66. http://dx.doi.org/10.1109/TE.2010.2043531

Lorenzo, M., Crouch, C. H., & Mazur, E. (2006). Reducing the gender gap in the physics classroom. *American Journal of Physics, 74,* 118–122. http://dx.doi.org/10.1119/1.2162549

Lubinski, D., & Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math-science pipeline. *Current Directions in Psychological Science, 1,* 61–66. http://dx.doi.org/10.1111/1467-8721.ep11509746

Madsen, A., McKagan, S. B., & Sayre, E. C. (2013). Gender gap on concept inventories in physics: What is consistent, what is inconsistent, and what factors influence the gap? *Physical Review Special Topics. Physics Education Research*, *9,* 020121. http://dx.doi.org/10.1103/PhysRevSTPER.9.020121

Mazur, E. (2015). *Principles and practice of physics*. Harlow, England: Pearson.

McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 636–649. http://dx.doi.org/10.1037/0278-7393.9.4.636

McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today, 37,* 24–32. http://dx.doi.org/10.1063/1.2916318

Mevarech, Z. R., & Fridkin, S. (2006). The effects of IMPROVE on mathematical knowledge, mathematical reasoning and meta-cognition. *Metacognition and Learning, 1,* 85–97. http://dx.doi.org/10.1007/s11409-006-6584-x

Murphy, P. K., & Cromley, J. G. (Eds.). (2015). Examining innovations: Navigating the dynamic complexities of school-based intervention research. *Contemporary Educational Psychology, 40,* 1–130.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Newcombe, N. S., Ambady, N., Eccles, J., Gomez, L., Klahr, D., Linn, M., . . . Mix, K. (2009). Psychology's role in mathematics and science education. *American Psychologist, 64,* 538–550. http://dx.doi.org/10.1037/a0014813

Nieminen, P., Savinainen, A., & Viiri, J. (2010). Force Concept Inventory-based multiple-choice test for investigating students' representational consistency. *Physical Review Special Topics Physics Education Research, 6,* 1–12.

Ploetzner, R., Fehse, E., Kneser, C., & Spada, H. (1999). Learning to relate qualitative and quantitative problem representations in a model-based setting for collaborative problem solving. *Journal of the Learning Sciences, 8,* 177–214. http://dx.doi.org/10.1207/s15327809jls0802_1

Pollock, S. J., Finkelstein, N. D., & Kost, L. E. (2007). Reducing the gender gap in the physics classroom: How sufficient is interactive engagement? *Physical Review Special Topics—Physics Education Research, 3,* 010107-1–010107-4.

Raven, J. C., Raven, J., & Court, J. H. (1992). *Raven's progressive matrices und vocabulary scales. Teil 4 Advanced progressive matrices* (S. Bulheller & H. Häcker, Trans.). Frankfurt, Germany: Swets & Zeitlinger.

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1998). Student expectations in introductory physics. *American Journal of Physics, 66,* 212–224. http://dx.doi.org/10.1119/1.18847

Remillard, J. T., Herbel-Eisenmann, B. A., & Lloyd, G. M. (2011). *Mathematics teachers at work: Connecting curriculum materials and classroom instruction*. New York, NY: Routledge.

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development, 77,* 1–15. http://dx.doi.org/10.1111/j.1467-8624.2006.00852.x

Sanchez, E., Garcia-Rodicio, H., & Acuna, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science, 37,* 537–563. http://dx.doi.org/10.1007/s11251-008-9074-5&psfiisttps

Schalk, L., Saalbach, H., & Stern, E. (2016). Approaches to foster transfer of formal principles: Which route to take? *PLoS ONE, 11*(2), e0148787. http://dx.doi.org/10.1371/journal.pone.0148787

Schalk, L., Schumacher, R., Barth, A., & Stern, E. (2017). When problem-solving followed by instruction is superior to the traditional tell-and-practice sequence. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000234

Schneider, M., & Stern, E. (2010). The cognitive perspective on learning: Ten cornerstone findings. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 69–90). Paris, France: OECD Publishing. http://dx.doi.org/10.1787/9789264086487-5-en

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–5223. http://dx.doi.org/10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103,* 759–775. http://dx.doi.org/10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22,* 129–184. http://dx.doi.org/10.1207/s1532690xci2202_1

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99,* 285–296. http://dx.doi.org/10.1037/0022-0663.99.2.285

Seidel, T. (2006). The role of student characteristics in studying micro teaching–learning environments. *Learning Environments Research, 9,* 253–271. http://dx.doi.org/10.1007/s10984-006-9012-x

Seidel, T., & Prenzel, M. (2006). Stability of teaching patterns in physics instruction: Findings from a video study. *Learning and Instruction, 16,* 228–240. http://dx.doi.org/10.1016/j.learninstruc.2006.03.002

Seidel, T., Prenzel, M., Duit, R., & Lehrke, M. (Eds.). (2003). *IPN-Materialien. Technischer Bericht zur Videostudie "Lehr-Lern-Prozesse im Physikunterricht"*. Retrieved from http://archiv.ipn.uni-kiel.de/buecherarchiv/buch_videostudie2.html

Sinatra, G. M., & Pintrich, P. R. (2003). *Intentional conceptual change*. Mahwah, NJ: Lawrence Erlbaum.

Stadler, H., Duit, R., & Benke, G. (2000). Do boys and girls hold different notions of understanding in physics? *Physics Education, 35,* 417–422. http://dx.doi.org/10.1088/0031-9120/35/6/307

Staub, F. C., & Stern, E. (2002). The nature of teachers' pedagogical content beliefs matters for students' achievement gains: Quasi-experimental evidence from elementary mathematics. *Journal of Educational Psychology, 94,* 344–355. http://dx.doi.org/10.1037/0022-0663.94.2.344

Taconis, R., Ferguson-Hessler, M. G. M., & Broekkamp, H. (2001). Teaching science problem solving: An overview of experimental work. *Journal of Research in Science Teaching, 38,* 442–468. http://dx.doi.org/10.1002/tea.1013

UCLA & the Statistical Consulting Group. (2014, July 18). *Mplus FAQ. How can I compute a chi-square test for nested models with the MLR or MLM estimators?* Retrieved from http://www.ats.ucla.edu/stat/mplus/faq/s_b_chi2.htm

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16,* 3–118. http://dx.doi.org/10.1207/s1532690xci1601_2

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Zepeda, C. D., Elizabeth, J., Ronevich, P., & Nokes-Malach, T. J. (2015). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. *Journal of Educational Psychology, 107,* 954–970. http://dx.doi.org/10.1037/edu0000022

Ziegler, E., & Stern, E. (2014). Delayed benefits of learning elementary algebraic transformations through contrasted comparisons. *Learning and Instruction, 33,* 131–146. http://dx.doi.org/10.1016/j.learninstruc.2014.04.006

Ziegler, E., & Stern, E. (2016). Consistent advantages of contrasted comparisons: Algebra learning under direct instruction. *Learning and Instruction, 41,* 41–51. http://dx.doi.org/10.1016/j.learninstruc.2015.09.006

Zohar, A. (2006). Connected knowledge in science and mathematics education. *International Journal of Science Education, 28,* 1579–1599. http://dx.doi.org/10.1080/09500690500439199

Zohar, A., & Sela, D. (2003). Her physics, his physics: Gender issues in Israeli advanced placement physics classes. *International Journal of Science Education, 25,* 245–268. http://dx.doi.org/10.1080/09500690210126766

Received October 29, 2016
Revision received December 12, 2017
Accepted December 13, 2017 ■

Correction to Greene et al. (2018)

In the article "A Meta-Analytic Review of the Relationship Between Epistemic Cognition and Academic Achievement," by Jeffrey A. Greene, Brian M. Cartiff, and Rebekah F. Duke (*Journal of Educational Psychology*, Advance online publication. March 8, 2018. http://dx.doi.org/10.1037/edu0000263), Table 7 contained a production-related error. Overall *N* was listed as "1,9,319" when it should be "159,319." All versions of this article have been corrected.

http://dx.doi.org/10.1037/edu0000312