



Variable control and conceptual change: A large-scale quantitative study in elementary school



Peter A. Edelsbrunner^{a,*}, Lennart Schalk^b, Ralph Schumacher^a, Elsbeth Stern^a

^a ETH Zurich, Switzerland

^b ETH Zurich, PH Schwyz, Switzerland

ARTICLE INFO

Keywords:

Conceptual change
Elementary school
Control-of-variables strategy
Mixture modeling
Latent transition analysis

ABSTRACT

We examined the predictive value and interplay of elementary school students' understanding of the control-of-variables strategy, a domain-general experimentation skill, and their prior content knowledge for subsequent conceptual knowledge acquisition and conceptual change. Trained teachers provided $N = 1809$ first to sixth graders with 15 lessons of guided inquiry-based instruction on floating and sinking. We assessed understanding of the control-of-variables strategy before instruction, and conceptual content knowledge from before to after instruction. A mixture model analysis, specifically, a latent transition analysis, indicates that understanding of the control-of-variables strategy predicts content knowledge structure before instruction, and content knowledge development from before to after instruction. These findings corroborate lab-based research on the interplay of experimentation skills and content knowledge in inquiry settings and extend it to teacher-guided classroom instruction. We describe how learning pathways vary depending on students' understanding of the control-of-variables strategy and prior content knowledge, and discuss implications for learning and instruction.

1. Introduction

Conceptual change research has yielded many insights into students' development of conceptual knowledge. These insights have stimulated the generation of elaborate science units for kindergarten (Leuchter, Saalbach, & Hardy, 2014), elementary (Hardy, Jonen, Moeller, & Stern, 2006), and early secondary school (Smith, 2007). Often, science education in these first stages of schooling is based on inquiry. In general, inquiry-based science instruction, particularly under teacher-guidance, is an effective instructional means for developing conceptual content knowledge (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011; Furtak, Seidel, Iverson, & Briggs, 2012; Hattie, 2009; Slavin, Lake, Hanley, & Thurston, 2014). But not all students advance to the same degree. According to conceptual change theory, differences in the prior content knowledge that students bring to class can explain this interindividual variation. Students' knowledge representations differ because they have different experiences from everyday life and prior education (Carey, 1985, 2000).

Differences exist not only in content knowledge but also in students' understanding of experimentation. In inquiry-based instruction, students often engage in experimentation. Setting up and interpreting experiments requires adequate understanding of domain-general experimentation principles such as the control-of-variables strategy (CVS)

(Kuhn, Black, Keselman, & Kaplan, 2000; Kuhn, Ramsey, & Arvidsson, 2015). There are other, more and less advanced steps in the development of knowledge about experimentation, however understanding of the CVS is pivotal (Croker & Buchanan, 2011; Kuhn, Iordanou, Pease, & Wirkala, 2008; Osterhaus, Körber, & Sodian, 2016; Piekny & Maehler, 2013; Sodian, Zaitchik, & Carey, 1991). Interindividual differences in understanding this domain-general strategy can be expected to influence the acquisition of scientific concepts. In the present study, we aim at scrutinizing the predictive value and interplay of students' understanding of the CVS and their prior content knowledge for subsequent knowledge acquisition and conceptual change in inquiry-based science instruction. Building on prior lab-based research, we examine this interplay in the context of real science classrooms, with a large-scale sample of elementary school students being instructed by their classroom teachers. For data analysis, we apply an innovative kind of statistical modeling, specifically, a latent transition analysis.

1.1. Conceptual change in science

When students enter science classrooms, they bring prior conceptions about the instructed topics derived from their everyday experiences (Carey, 1985, 2009; Hardy et al., 2006). Imagine children hanging out at a river. Sitting at the riverbank, they note that stones lie

* Corresponding author at: Research on Learning and Instruction, Department of Humanities, Political and Social Sciences, Clausiusstrasse 59, RZ H16, 8092 Zurich, Switzerland.
E-mail address: peter.edelsbrunner@ivf.gess.ethz.ch (P.A. Edelsbrunner).

on the river ground but wood floats by; later they throw small wooden branches and flip yet another stone and see that the light pieces of wood float, while the stones sink to the ground. Watching a steamboat entering the port, they admire the captain whom they recognize to be essential for safe ship passage. An anchor is released and sits so firmly on the ground that it prevents any absconding of the massive iron object that floats on the water. Talking about their experiences, they come up with some explanations for their perceptions. They discuss that light things float, heavy things sink, and a captain keeps a ship floating.

The usefulness of conceptions arising from such everyday experiences is often limited for explaining scientific phenomena. The captain is not the decisive characteristic for a ship's floating ability and not all wooden things float. But these conceptions are not generally useless. They serve sufficiently well for explaining some occurrences of floating ability. However, these conceptions reach their limits when more and more phenomena are experienced. The conceptions are wrong from a scientific point of view, because they cannot explain all occurrences of floating ability, and therefore they are called misconceptions (Chi & Ohlsson, 2005). An important aim of science education is to help students to develop an understanding of scientific concepts. For phenomena of floating and sinking, these are the concepts of object density and buoyancy force. The step from misconceptions to scientific concepts is far. Intermediate conceptions can bridge the gap (Carey, 1992; Hardy et al., 2006). These conceptions typically develop when students blend information given in instruction and their prior conceptions (Hardy et al., 2006).

Intermediate conceptions are also sometimes deliberately introduced by teachers in order to simplify content, but still to prepare their students' future science learning. For example, when children think about floating and sinking, they often give explanations such as "light things float, heavy ones sink" or "small objects float, large ones sink". They do not yet understand that weight and size interact as density and thus see these conceptions as independent of each other (Maclin, Grosslight, & Davis, 1997; Smith, Carey, & Wisner, 1985). A more elaborate but still limited intermediate conception would be "things made of wood float, while stones sink". This material-based conception can explain more floating ability phenomena than the conceptions of weight and size, but it is still limited in its explanatory power. When learning science, students show diverse developmental patterns in how they change from misconceptions via intermediate conceptions to scientific concepts. To support this development, it is necessary to understand how these learning patterns are structured and constrained, and how optimal knowledge development can be supported.

1.2. Conceptual change in the science classroom

Powerful processes of knowledge restructuring have to be triggered to enrich students' initial stock of misconceptions with scientific concepts or first with intermediate conceptions. These processes are referred to as conceptual change (Chi, 2008; Chi & Ohlsson, 2005; Ohlsson, 2009). For example, novices often have difficulties in recognizing deep and meaningful relations between prior knowledge and newly acquired knowledge (diSessa, 2008). In such cases, newly acquired knowledge is not connected with prior knowledge, leading to fragmented knowledge elements that are stored independently of each other. Knowledge fragmentation decreases when students gain sufficient conceptual understanding of a domain to integrate knowledge pieces into more coherent and general knowledge structures (Linn, Eylon, & Davis, 2004). This and similar processes of conceptual change allow integrated knowledge structures to be built up, for example by learning that different phenomena can be explained by a single principle, concept, or theory (Ohlsson, 2009).

One effective educational intervention for promoting conceptual change in science is inquiry-based learning, in which students engage in the thinking processes and activities of scientists (American Association

for the Advancement of Science, 1993). This often includes social, procedural, and epistemic activities such as arguing scientific ideas, engaging in experimentation, and interpreting evidence (Furtak et al., 2012). Inquiry-based learning is a successful method for teaching science across various topics and educational levels (Anderson, 2002; Bennett, Lubben, & Hogarth, 2007; Flick, 1995; Furtak et al., 2012; Minner, Levy, & Century, 2010; Shymansky, Hedges, & Woodworth, 1990). Particularly in combination with strong teacher guidance, students' learning benefits in comparison to other traditional instructional methods, such as direct instruction (Furtak et al., 2012). However, learning differs not only between traditional and inquiry-based instructional conditions. Also within similar inquiry-based instructional settings (e.g., within one classroom), students learn to different degrees. These different learning gains on the one hand reflect differences in students' prior content knowledge, but it has also been pointed out that specific domain-general experimentation skills influence students' knowledge development (Bryant, Nunes, Hillier, Gilroy, & Barros, 2015; Chen & Klahr, 1999).

1.3. Experimentation and learning from inquiry

A precondition for beneficial engagement in inquiry is a thorough understanding of experimental designs (Kuhn, 2002; Kuhn et al., 2000). A crucial facet of experimentation concerns varying the focal variable while keeping all other factors constant. This strategy is referred to as the control-of-variables strategy (CVS), or as vary-one-thing-at-a-time (VOTAT). Following this strategy allows making unambiguous causal inferences (Strand-Cary & Klahr, 2008). CVS predicts academic performance and science learning above and beyond general reasoning abilities (Bryant et al., 2015; Wüstenberg, Greiff, & Funke, 2012). Most but not all children typically develop some understanding of the CVS at ages 6–10, depending on task context and the number of variables that have to be controlled (Sodian & Bullock, 2008; Zimmerman, 2007). Development of the understanding of the correct variation of the focal variable initiates in early childhood (Pieky & Maehler, 2013; Sodian et al., 1991). Then, around age 10, development of broader understanding of the CVS emerges in many children (Penner & Klahr, 1996). The understanding of the CVS and its development are moderately related to children's verbal reasoning and vocabulary, and to their general science content knowledge (Wagensveld, Segers, Kleemans, & Verhoeven, 2015). However, thoroughly understanding and being able to apply the CVS is challenging, and even some undergraduates lack these competencies (Lin & Lehman, 1999).

The development of conceptions about scientific phenomena and understanding of experimentation are probably not independent from each other, but exhibit mutual influence. In observational lab studies, Schauble (1990, 1996) found evidence for this interplay when she studied belief revision about causal mechanisms in observational lab studies. Students' knowledge about causal relations influenced experimentation strategies, while students' experimentation strategies in turn influenced the acquisition of content knowledge about causal relations. Based on these studies, it has been widely acknowledged that experimentation skills and content knowledge interplay in inquiry settings (Zimmerman, 2007). Taking these lab-based findings as a starting point, we aimed to scrutinize the generality and potential of this interrelation in classroom education.

We do not know from prior research whether students' understanding of experimentation influences their development of domain-specific conceptual knowledge in a teacher-guided inquiry-based curriculum unit. Does teacher guidance level out or enlarge the impact of students' understanding of experimentation on further learning? There are arguments for both sides. When teachers guide students in setting up experiments and engage them in argumentation about the outcomes, this might sufficiently support inferences and knowledge development even for students who entered the curriculum with poor understanding of experimentation. Put differently, teachers might take the

responsibility of students by explicating critical design features of experiments and how these constrain valid inferences. On the other hand, according to the widely accepted cognitive constructivist view of learning, conceptual change requires active learning processes. The role of teacher guidance first and foremost lies in guiding students' attention towards critical features (Ziegler & Stern, 2016). While teachers can trigger students' reasoning and point them towards key concepts, students have to figure out and grasp key inferences on their own. In the instructional setting used in the present study, the students for example immerse objects of different shapes, sizes, and materials into water, to examine how these variables influence floating ability. Having a more advanced understanding of the CVS can be expected to be beneficial for grasping which characteristics determine floating ability. If only one of the object characteristics changes between trials, students who understand the importance of CVS should quickly comprehend that this characteristic is relevant. When none or several main characteristics change, they should accordingly comprehend that a conclusive inference is not possible. In the present study, we examined whether students' understanding of the CVS matters and how it predicts trajectories of concept development in a teacher-guided inquiry-based curriculum unit on floating and sinking.

1.4. Innovative statistical approaches to modeling conceptual change

The type of assessment and the statistical model used to analyze data represent substantial factors in studies on conceptual change (Frède et al., 2011; Straatemeier, van der Maas, & Jansen, 2008). Typical approaches to assessing conceptual change include interviews (Christou & Vosniadou, 2012; Nussbaum & Novak, 1976), drawings (Vosniadou & Brewer, 1992), concept mapping (Liu, 2004), and questionnaires with multiple-choice questions (Hardy et al., 2006; Straatemeier et al., 2008) or open questions (Christou & Vosniadou, 2012). These methods can either be interpreted qualitatively, or students' answers can be quantified and analyzed using statistical models. Qualitative interpretations have the advantage of revealing unexpected aspects of students' beliefs. They can highlight interindividual differences in children's content knowledge development. Quantitative analyses with larger sample sizes allow for generalizations beyond the assessed student sample. This advantage of quantification, however, usually requires treating all students as stemming from the same population. That is, all differences between students are considered random and the same parameter estimates are used to describe the content knowledge structure and its development across all analyzed students. Potential interindividual differences in knowledge structure and development are usually neglected when analyzing scores from assessment instruments, for example by simply comparing mean pretest and posttest scores across large groups of students. This approach is not in line with conceptual change theory in which qualitative knowledge differences between children are emphasized as explanation for interindividual differences in content knowledge development (Kleickmann, Hardy, Pollmeier, & Möller, 2011; Schneider & Hardy, 2013).

It is possible to combine the advantages of qualitative and quantitative methods in the framework of mixture modeling. Mixture modeling is a quantitative approach that allows to model interindividual differences between students. This becomes possible by modeling students' scores to stem from a finite number of different populations. In conceptual change research, these populations describe different knowledge states. For example, in mathematics, knowledge states can reveal themselves in different strategies students apply for solving particular mathematics problems (Fagginger Auer, Hickendorff, Van Putten, Béguin, & Heiser, 2016; McMullen, Laakkonen, Hannula-Sormunen, & Lehtinen, 2015). In geography, knowledge states can be represented in the consistency of answers indicating different cognitive models of the shape of the earth (Straatemeier et al., 2008). In physics, knowledge states can refer to groups of students with different numbers of misconceptions, intermediate conceptions, and scientific concepts

about the floating ability of objects in water (Schneider & Hardy, 2013).

The present study builds on the method of the latter study. Schneider and Hardy (2013) applied mixture modeling, specifically, a latent transition analysis, in order to depict the structure and development of third-graders' knowledge development in a teacher-guided inquiry-based basic physics curriculum unit. Specifically, they examined the development of students' conceptions about floating and sinking. Latent transition analysis proved useful and informative in describing students' structural knowledge development. In the present study, we used the same curriculum as in Schneider and Hardy (2013), and a similar statistical modeling technique, but on a larger sample and additionally including covariates.

1.5. The present study

We investigated the interrelation between experimentation skills and content knowledge development using longitudinal mixture modeling. Specifically, on a large-scale sample, we applied a latent transition analysis with covariates (CVS scores and school grade) to examine students' content knowledge development and how it is influenced by their understanding of the CVS. This approach allowed us to model that students differ in their prior content knowledge and thus combined the strengths of traditional qualitative and quantitative analysis. As mentioned, teacher-guided, inquiry-based instruction is a fruitful instructional approach for science education, and in lab-based inquiry settings students' experimentation skills and their knowledge development are intertwined (Furtak et al., 2012; Schauble, 1990, 1996). But it is yet unknown whether and to which degree this interplay matters in classroom-based instruction. We provided a large number of first- to sixth-graders with teacher-guided, inquiry-based instruction on the topic *floating and sinking of objects in water* in their real school environment. In this setting, we examined whether students' understanding of the CVS predicts their content knowledge development, particularly in interaction with their prior content knowledge.

2. Method

2.1. Participants

The sample consisted of $N = 1809$ students from 108 first- to sixth-grade school classes of 50 elementary schools from German speaking cantons (kind of federal states) of Switzerland.

Mean age was 9.29 years ($SD = 1.36$), with an age range of 6 to 13 years. 48.6% ($n = 879$) of the students were female. The school classes were recruited to participate in the *Swiss MINT Study*.

The Swiss MINT Study is a large-scale study in which outcomes of early-initiated science education are investigated longitudinally. The study was initiated at ETH Zurich in 2010 and the present sample consists of all school classes that participated in the relevant parts of the first phase of the study. The study focused on third and fourth grade but there were more than 100 students from each grade. Specifically, there were $n = 109$ (6.0%) first grader, $n = 206$ (11.4%) second graders, $n = 514$ (28.4%) third graders, $n = 604$ (33.4%) fourth graders, $n = 258$ (14.3%) fifth graders, and $n = 118$ (6.3%) sixth graders. Active parental consent was sought for the participating students and parents were informed that they could withdraw their consent at any time.

2.2. Learning materials

The students received instruction on the topic *floating and sinking of objects in water*. The instructional materials for this curriculum unit were developed and extensively tested at the University of Munster (see Hardy et al., 2006, for details on the materials). The materials comprise 15 lessons of teacher-guided, inquiry-based classroom instruction. Across the lessons, students engage in many hands-on experiments in which they, for example, compare the floating ability of objects of

different size, shape, or kind. In a stepwise manner, more sophisticated explanatory models are introduced, tested, and discussed as explanations for students' assumptions and observations. Before and after experiments, the teachers initiate and lead discussions on the students' prior assumptions and observations, prompting justifications for their assumptions and for their explanations of outcomes. Thus, the instructional principles used in the curriculum unit encompassed prior knowledge activation, self-explanations, and compare and contrast activities.

Instruction starts with directing learners' attention to the size of objects as well as to the material they are made of. Concepts of object density, water displacement, and buoyancy force are then introduced by engaging the students in experiments that show the limits of object size, weight, and material as explanatory factors for floating ability. For example, the materials contain cubes and other objects known from everyday life of different material but the same volumes, of different volumes but the same materials, and further objects differing in shape and in hollowness.


These variations in object characteristics allow testing experimentally - by applying the CVS - which object features influence floating ability. Thus, the CVS can be directly applied in finding out about floating and sinking phenomena. Students experience that object weight and size interact, and how water is displaced and pushes back against objects. The key concepts are then deduced in argumentative discussions together with the teacher, without referring to scientific definitions or more scientific terms such as mass, volume, or density. Rather, the concepts are discussed in students' everyday language (e.g., "it matters how heavy something is in comparison to how big it is") to better align with students' knowledge. Taken together, the curriculum unit is thus aimed at developing students' conceptual understanding of the phenomena, rather than building up fact knowledge about scientific definitions, to help students acquire conceptual content knowledge that can be built on in more advanced future education.

The teachers received one day of training in small groups in which the study authors introduced them to the study materials, experiments, and instructional principles. The students typically received either one lesson per week instead of their usual science lessons (which in Swiss elementary schools usually encompass topics from Geography and Biology), or they received the whole curriculum within a week that was devoted to special projects. The curriculum can be used adaptively from first to sixth grade classrooms. In most first and second grade classrooms the most advanced lessons were omitted and instead some basic lessons were treated more intensively. The implementation of the curriculum unit took the teachers a median time of 2 months. There was no evaluation of teachers' implementation of the curriculum unit, however comparisons with control groups and further prior studies have shown that teachers can successfully implement the unit after receiving a short training (Schalk, Edelsbrunner, Schumacher, & Stern, submitted; Möller, Hardy, Jonen, Kleickmann, & Blumberg, 2006; Stern, Schumacher, Edelsbrunner, Schalk, & Deiglmayr, 2016).

2.3. Assessments

To assess students' content knowledge development about floating and sinking, they answered a multiple-choice questionnaire before (pretest) and after (posttest) the instruction. The questionnaire assesses misconceptions (incorrect from a scientific view), intermediate conceptions (partially correct), and scientific concepts (fully correct) about the floating ability of objects in water. For each question, the students could choose multiple answers that represented the different types of conceptions. The questionnaire has been developed in multiple pilot studies and is a reliable indicator of knowledge development triggered by the floating and sinking curriculum (see Hardy et al., 2006). An example item that encompasses answers reflecting all three types of conceptions is provided in Fig. 1. Students' answers on the questionnaire yielded three scores, indicating their numbers of

This metal plate is immersed into water. What do you expect to happen?



The metal plate...

sinks floats

Why?

...because it lies on the water.

...because it is metal.

...because its weight is evenly spread out.

...because the displaced water weighs less than the plate.

...because it is not pushed up strongly enough by the water.

...because the plate is so heavy.

Fig. 1. An example item from the assessment of students' knowledge about floating and sinking. The students had to decide whether the metal plate floats or sinks and they could choose one or more rationales for their choice. The first, third, and sixth rationales each yielded one point for their misconceptions-score, the second for their intermediate conceptions-score, and the fourth and fifth for their scientific concepts-score.


misconceptions, intermediate conceptions, and scientific concepts. We used these three scores as indicators of students' conceptual content knowledge about floating and sinking in our main analysis. The students could obtain maximum scores of 45 misconceptions, 12 intermediate conceptions, and 19 scientific concepts. The questions covered typical conceptions found in children that were identified based on prior literature and in-depth interviews with primary school students. Misconceptions encompassed for example the dependency of water displacement on mass, and object mass, volume, or form as active forces.


Intermediate conceptions included conceptions derived from everyday life, for example references to objects' material or hollowness as decisive factors for floating ability. Scientific concepts included for example references to the ratio of object mass to displaced water, and to water exerting buoyancy force. Further detailed descriptions of the questionnaire are available in Hardy et al. (2006). The posttest-questionnaire included an additional transfer test with seven questions that were not part of the first assessment. The transfer questions assessed whether the students could apply the instructed concepts to new situations that were not covered in the instruction. We used students' sum score on these seven questions as an indicator of knowledge transfer of the acquired concepts.


Before the instruction and the pretest on floating and sinking, the students answered a questionnaire to assess their understanding of the CVS. The questionnaire consisted of 14 multiple-choice questions (latent reliability estimate = 0.75; see Raykov, Dimitrov, & Asparouhov, 2010). The questions were analogue to classic tasks such as the ramp- (Chen & Klahr, 1999), airplane- (Bullock & Ziegler, 1999) and the mouse-task (Sodian et al., 1991). An example item is provided in Fig. 2. Seven questions dealt with the evaluation or interpretation of experiments, and seven with the creation of experimental designs (Bryant et al., 2015). In the evaluation items, students were asked whether a presented experiment was a good comparison for finding something out. The interpretation items included for example the mouse-task (Sodian et al., 1991), in which students have to interpret findings, taking into account whether the focal variable was correctly manipulated. For six items, the correct variation of the focal variable was relevant (e.g., the mouse task), and for eight items the control of confounding variables (e.g., the airplane task), two aspects which together

Which airplane needs the least fuel?

Mr. Miller builds airplanes. His airplanes should use as little fuel as possible. He has a couple of ideas about what might be important for how much fuel an airplane needs.

The form of the nose can be pointed or rounded. 

The rudder can be on the bottom or on the top. 

The wings can be double or single. 

Mr. Miller believes that airplanes with a pointed nose need less fuel than airplanes with a rounded nose. What should Mr. Miller do to find out whether the form of the nose is important for how much fuel the airplane needs?

Select the correct answer:

Mr. Miller should build a couple of planes. Then, he can compare how much fuel each one needs.

Mr. Miller should build two planes: one with a pointed and one with a round nose, but with the same rudder and wings. Then, he can compare how much fuel each one needs.

Mr. Miller should build two very different planes. They should have different noses, rudders, and wings. Then, he can compare how much fuel each one needs.

Fig. 2. An example item from the assessment of students' understanding of the CVS. In this item, the control of confounding factors is the relevant feature for the design of the experiment. The correct answer is the second choice. Item context taken from Bullock and Ziegler (1999).

represent the CVS. It should be noted that this assessment goes beyond traditional CVS measures in which often only the control of confounding variables is in focus. However, the correct variation of the focal variable is an important step in the development of CVS, and related to common misconceptions about experimentation (Klahr & Chen, 2003; Piekny & Maehler, 2013; Schwichow, Christoph, Boone, & Härtig, 2016; Siler & Klahr, 2012; Sodian et al., 1991). We therefore included these different types of questions to assess a broad construct representing the varied facets of the CVS and related experimentation skills similar to Schwichow et al. (2016). All questions treated domains not related to the curriculum unit. The students received a score of 1 for each correct answer and a score of 0 for each wrong answer. We used students' mean score on the questionnaire ranging from 0 to 1 as an indicator of their understanding of the CVS. A bi-factor analysis corroborated the use of a composite score (Reise, 2012). The understanding of the CVS was assessed when the students entered the longitudinal study. For some students, this was directly before the floating and sinking assessment while others received other instructional sessions in between that were not concerned with floating and sinking. The students took the CVS assessment with a median delay of 9 months before answering the floating and sinking pretest and starting with the instruction.

2.4. Statistical analysis

The dataset, analysis scripts and further materials are freely available from <https://osf.io/n6gev/>. The student variables encompassed their mean score on the CVS questionnaire, the three scores on the floating and sinking questionnaire (misconceptions, intermediate conceptions, and scientific concepts) at pretest and posttest, as well as the score from the transfer questions at posttest, and their age, gender, and school grade. Our statistical approach comprised three analyses. First, we estimated descriptive statistics and intercorrelations of the main study variables. Then, we estimated a basic regression-based model, to compare its results to those from the mixture modeling approach. For the basic regression-based model, we set up a change score model (McArdle, 2009, analytic details provided in the appendix) to estimate how much of the variance in students' change on the three knowledge indicator variables could be explained by their CVS scores.

In the next step, we conducted a latent transition analysis (LTA), the appropriate type of mixture model for our data. We first decided on the number of knowledge profiles to estimate based on the BIC criterion (Nylund, Asparouhov, & Muthen, 2007) and theoretical considerations. For the final model, the following parameters were estimated: a) mean and variance patterns capturing students' knowledge profiles across the three knowledge indicators at pretest and posttest, b) profile sizes that indicated how many of the children showed each of the knowledge profiles at pretest, c) knowledge profile transition patterns indicating

how likely the children were to transition from one to another profile from pre- to posttest. Afterwards, we added students' score on the CVS measure as a covariate to the model. This model indicated whether and how students' understanding of the CVS predicted their content knowledge profiles at pretest, and their transitions between the different content knowledge profiles from pretest to posttest. We decided that a difference of more than 10% in transition probabilities between students high (+1.5SD) and low (−1.5SD) on CVS would indicate a relevant predictive value of CVS. We also controlled for school grade by adding it as another covariate, to take into account that the most advanced lessons were omitted in the lower grades. Finally, based on the profiles at posttest, we also estimated students' scores on the transfer questions. That is, we compared the group means of the knowledge profiles on the transfer score. In the estimation process, we fixed the mean values on profiles to be the same at pretest and posttest, in order to be able to interpret which students stayed in the same profile, and which transitioned into a different one. We did however let the number of knowledge profiles differ between pretest and posttest. This decision reflects the assumption that through instruction, it can happen that some knowledge profiles do not exist anymore, while other, new profiles might only be reached after students have received elaborate instruction about the concepts.

We did not use null hypothesis significance testing in these analyses. Significance tests should only be used for purely confirmatory testing (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). Confirmatory testing is in principle possible in mixture modeling (see e.g., Finch & Bronk, 2011), however, it is not always appropriate. We do not deem a purely confirmatory approach appropriate for the present study, rather the mixture modeling mostly serves an exploratory aim. Therefore, we abstain from reporting and interpreting *p*-values.

Rather, we report information criteria, which can inform about the relative plausibility and predictive value of models and estimates (Wagenmakers, 2007). In addition, we report effect sizes, confidence intervals, and criteria for judging the strength or size of effects, where appropriate. This analysis strategy prevents mixing up predictive modeling with hypothesis testing, an issue that is seldom considered however can lead to unfounded inferences about parameter estimates, models, and theoretical implications (Wagenmakers et al., 2012).

There were 29 (1.6%) students missing at the first content knowledge assessment, and 40 (2.2%) at the second content knowledge assessment. These students were absent from school for unknown reasons. We applied full information maximum likelihood estimation for all models except for the change score model for which we applied Bayesian estimation (for an overview of Bayesian methods, see Edelsbrunner, 2014; Etz, Gronau, Dablander, Edelsbrunner, & Baribault, 2017; Wagenmakers, Morey, & Lee, 2016). Both estimators handle missing data so that listwise deletion was not required, which means we could use data from all students for the statistical estimations. The Mplus software version 7.11 was used for all analyses (Muthén & Muthén, 2012).

3. Results

Descriptive statistics of the main study variables in the different school grades are provided in Table 1. The estimated variance-covariance and correlation matrix with descriptive statistics are provided in the appendix, Table A2. In comparison to another study with the same learning materials and knowledge assessments (Hardy et al., 2006), students' achievement in third grade was slightly lower (Table 1). As a general indicator of the effectiveness of the instruction, effect sizes and 95% confidence intervals of students' change in mean values on the three indicator variables of their knowledge about floating and sinking were estimated. There was a strong reduction of Cohen's $d = -0.98$, CI95 [−1.02, −0.93] on children's number of misconceptions, an increase of $d = 0.72$, CI95 [0.68, 0.77] on their number of everyday conceptions, and an increase of $d = 0.75$, CI95 [0.71, 0.80] on their number of scientific concepts. These estimates indicate that across the whole sample, from pretest to posttest students' misconceptions decreased strongly and knowledge about the instructed concepts increased moderately to strongly. The increase on the number of intermediate conceptions was expected, because in the course of the instruction students first learned about the intermediate conceptions, and then about the scientific concepts (see also Hardy et al., 2006).

3.1. Regression-based analysis

A depiction of the change score model and analytic details are provided in Appendix A. The model estimates indicated that students' CVS score explained 2% of variance in change in their misconceptions-scores from pretest to posttest, 1% of variance in change in their intermediate conceptions-scores, and 2% of variance in change in their scientific concepts-scores. Thus, in a traditional regression-based approach, the estimated explained variance in knowledge development by students' CVS scores is low. Including students' school grade or gender in this model did not change the results.

3.2. Latent transition analysis

Our main analytic approach was the latent transition analysis. We first examined the number of knowledge profiles based on students' misconceptions, intermediate conceptions, and scientific concepts scores at pretest and posttest. We increased the number of profiles in a stepwise manner up to eight profiles. Best fit was initially obtained with seven profiles at pretest and posttest. Some profiles were very small at one of the assessment points, that is, they contained almost no students and thus not much information. We restricted one of these profiles to contain zero students and removed others until we achieved the best fitting model. This model contained four profiles at pretest, and six profiles at posttest (see Table 2). We controlled for grade by adding it as a covariate, and added students' scores on the CVS test as another covariate to examine its interrelations with students' knowledge profiles at pretest (i.e., their prior knowledge), and with their profile transitions from pretest to posttest (i.e., their knowledge development). The inclusion of students' CVS scores improved model fit from $BIC = 56,725$

Table 1
Estimates of means and standard deviations of study variables in different school grades.

Grad		1	2	3	4	5	6
Misconceptions	Pre	20.27 (6.47)	19.63 (5.70)	18.36 (5.24)	17.56 (5.42)	16.53 (5.54)	14.73 (4.98)
	Post	13.88 (7.05)	13.67 (6.58)	12.12 (6.21)	11.56 (6.02)	8.86 (5.31)	9.63 (5.29)
Interm. conceptions	Pre	6.01 (2.42)	6.21 (2.53)	6.78 (2.45)	6.90 (2.48)	6.97 (2.62)	7.11 (2.58)
	Post	8.91 (2.11)	8.83 (2.39)	9.01 (2.51)	8.68 (2.31)	8.59 (2.50)	9.38 (2.22)
Scientific concepts	Pre	6.91 (3.56)	7.17 (3.41)	6.86 (3.41)	6.72 (3.39)	6.57 (3.11)	6.52 (2.91)
	Post	8.22 (4.07)	8.64 (4.04)	9.88 (3.93)	10.83 (4.15)	10.71 (4.27)	10.76 (3.90)
CVS score		0.44 (0.16)	0.46 (0.16)	0.45 (0.19)	0.47 (0.20)	0.51 (0.23)	0.57 (0.22)
Transfer score		2.21 (1.80)	1.89 (1.32)	2.27 (1.43)	2.90 (1.57)	3.24 (1.64)	3.35 (1.34)

Table 2
Relative fit indices and entropy for the estimated latent transition analysis models.

No. of profiles at pretest	No. of profiles at posttest	AIC	BIC	aBIC	Entropy
1	1	60,773	60,806	60,787	na
2	2	58,431	58,513	58,466	0.74
3	3	57,490	57,633	57,550	0.76
4	4	56,921	57,136	57,012	0.75
5	5	56,713	57,010	56,839	0.72
6	6	56,494	56,884	56,659	0.73
7	7	56,358	56,853	56,567	0.75
8	8	56,280	56,874	56,531	0.75
6	7	56,344	56,801	56,537	0.74
5	7	56,339	56,757	56,516	0.72
5	6	56,485	56,842	56,636	0.72
4	6	56,346	56,725	56,506	0.72
4	5	56,705	56,974	56,818	0.70

Note. AIC = Akaike information criterion; BIC = information criterion; aBIC = adjusted Bayesian information criterion; Entropy = degree of profile separation. Lower information criteria indicate better model-data fit, for LTA particularly the BIC. Finally selected model marked in bold.

to BIC = 56,715. The exploratory inclusion of gender did not show any substantial relations. Finally, we added students' score on the transfer questions as another indicator variable, to estimate mean values of students per profile on this score. The inclusions of students' CVS scores and the transfer scores did not alter the profile structures, which can sometimes happen after direct inclusion of additional variables (Hickendorff, Edelsbrunner, Schneider, Trezise, & McMullen, in press).

Our first results concern students' content knowledge profiles at pretest and posttest. We classified the profiles according to the level and prominence of the estimated indicator mean values (Marsh, Lüdtke, Trautwein, & Morin, 2009). In total, there were seven different profiles (Fig. 3 presents the estimated profiles). There were three profiles with a prominent number of misconceptions on an overall low level (low misconceptions profile), on a moderate level (moderate misconceptions profile), and on a high level (high misconceptions profile). There was one profile with an above-average number of all three types of conceptions (fragmented profile, cf. Schneider & Hardy, 2013). There was one profile with a high number of intermediate conceptions (intermediate profile), one with high numbers of intermediate conceptions and scientific concepts (prescientific profile, cf. Schneider & Hardy, 2013), and one with a moderate but prominent number of scientific concepts (scientific profile, cf. Schneider & Hardy, 2013).

At pretest, only four profiles were present: The three misconceptions-profiles, and the fragmented profile. At posttest, six profiles were present; the moderate misconceptions profile, in which most students (35%) were at pretest, was not present anymore. In addition, the three most proficient profiles, that is, the intermediate, prescientific, and

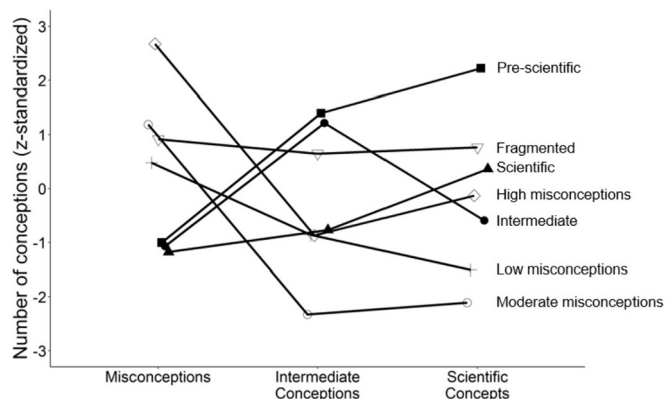


Fig. 3. The seven estimated knowledge profiles based on students' numbers of misconceptions, intermediate conceptions, and scientific concepts about floating and sinking.

Table 3
Estimated percentage of students in each profile at pretest and posttest, and mean transfer score estimates including 95% confidence intervals.

Profile	% pretest	% posttest	Transfer score [CI95]
High misconceptions	29%	6%	1.86 [1.60, 2.12]
Moderate misconceptions	35%	na	na
Low misconceptions	30%	28%	1.06 [0.84, 1.28]
Fragmented	6%	23%	2.41 [2.07, 2.75]
Intermediate	na	11%	2.82 [2.44, 3.18]
Scientific	na	21%	3.98 [3.50, 4.46]
Pre-scientific	na	11%	4.40 [3.98, 4.84]

Note. na indicates that the knowledge profile was not present at the relevant time point.

scientific profiles were present only at posttest. Thus, at posttest, the students had become more homogeneous in terms of less proficient content knowledge profiles, and they had developed three new, proficient content knowledge profiles. Descriptively, on the transfer test score the low misconceptions profile had the lowest estimate, followed by the high misconceptions profile, then the fragmented and intermediate profiles, and the prescientific and scientific profiles had the highest estimates (see Table 3).

Fig. 4 depicts the interrelations between students' CVS score and the four content knowledge profiles at pretest. By including students' CVS score as a predictor variable in the latent transition analysis, we estimated how many of the students show either of the four knowledge patterns depending on their CVS score. This correlational analysis between the CVS score and the profile frequencies at pretest shows a clear pattern: The higher students' CVS score was, the less likely they were to begin instruction in the high or moderate misconceptions-profiles. Instead, with increasing CVS scores, the frequency of students entering instruction in the low misconceptions-profile or in the fragmented profile increased. This result pattern indicates a positive relation between students' understanding of the CVS and their content knowledge profiles at pretest.

Next, we estimated students' profile transitions from pretest to posttest. All profile transitions detected for at least 5% of students are depicted in Fig. 5. These were just six transitions, which together represented the knowledge development of more than 72% of students. None of these six transitions concerned the fragmented profile. The most frequent transition represented almost a fourth of the students, who changed from the moderate misconceptions profile into the low misconceptions profile. Thus, these students could successfully restructure their knowledge by getting rid of misconceptions, and acquiring some intermediate conceptions, although their scientific concepts-score stayed almost at the same level. A further 5% changed from the moderate misconceptions into the fragmented profile, similarly discarding misconceptions while at the same time acquiring more intermediate conceptions and scientific concepts. The second most frequent transition was from the high misconceptions into the fragmented profile, with students discarding misconceptions, and acquiring both intermediate conceptions, and scientific concepts. A similar number of students changed from the low misconceptions into the scientific profile, discarding misconceptions and acquiring scientific concepts. From the low misconceptions profile, there were two further transitions: Students either transitioned into the intermediate profile, discarding misconceptions and acquiring intermediate conceptions, or into the pre-scientific profile by acquiring in addition a large number of scientific concepts. All of these most frequent transitions indicate positive knowledge development in the sense of science education; that is, these students could discard misconceptions, and acquire intermediate conceptions and/or scientific concepts. In addition, only 4.2% of student stayed in the same profile from pretest to posttest.

The results regarding our main question, whether students' understanding of the CVS predicts profile transitions from pretest to posttest, are depicted in Fig. 6. In the figure, all transition paths are depicted the

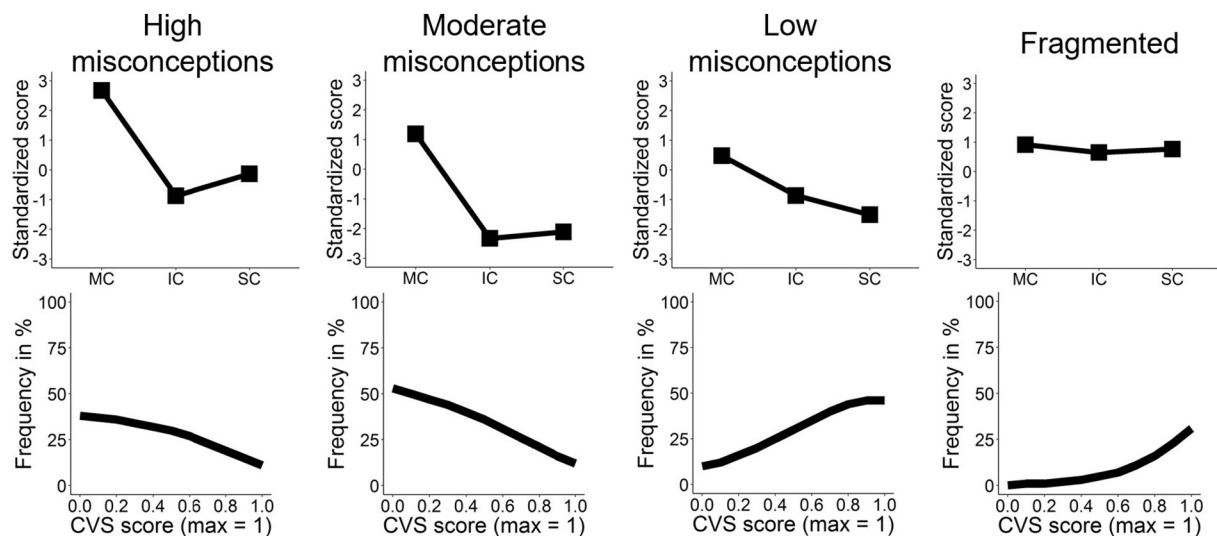


Fig. 4. Cross-sectional relation of students' CVS score to knowledge profile frequencies at pretest. Upper panel: the four knowledge profiles at pretest, represented by students' numbers of misconceptions (MC), intermediate conceptions (IC), and scientific concepts (SC). Lower panel: covariation of estimated profile frequencies (percent of students in the respective profile) at pretest with students' scores on the CVS assessment, controlling for school grade.

probability of which varied at least 10% between students with high (+1.5SD) or low (−1.5SD) CVS scores. The estimates of all transition paths for students with high and low CVS scores are provided in the Appendix Table A1.

For students starting in the high misconceptions profile (Fig. 6A), with increasing CVS the probability decreased that they would stay in this profile. Rather, they were more likely to transition into the intermediate profile. This indicates a positive predictive value of CVS. For students starting in the moderate misconceptions profile (Fig. 6B), with increasing CVS the probability decreased that they would transition into the low misconceptions-profile. Instead, the probability increased that they would transition into the scientific profile, indicating again a positive predictive value of the understanding of the CVS. For students starting in the low misconceptions-profile (Fig. 6C), the probability decreased that they would transition into the scientific profile, while increasing in the probability to transition into the intermediate profile.

Whether this indicates a positive predictive value will be discussed. Finally, for students starting in the fragmented profile (Fig. 6D), the probability decreased to transition into the prescientific profile, and the probability increased to transition into the scientific profile. As we will discuss, most of these predictive patterns indicate a positive predictive value of students' understanding of the CVS for their content knowledge development. The additional covariate of school grade did not have substantial additional predictive value beyond CVS. Between lower and higher grades, the predictive value of students' CVS score for profile frequencies at pretest, and for the transition probabilities was similar.

4. Discussion

We examined whether and how elementary school students' understanding of the control-of-variables strategy (CVS) interplays with their prior content-specific knowledge and predicts their content knowledge development in inquiry-based science instruction. These findings reinstate earlier lab-based findings (Schauble, 1990, 1996) and extend them to guided inquiry-based instruction in the classroom. Specifically, our analysis of a large-scale sample from six school grades indicates that CVS in fact matters for concept learning in the classroom: Students' understanding of the CVS is a positive predictor of their prior content knowledge (i.e., more proficient knowledge structures), and of their content knowledge development (i.e., transitions to more proficient knowledge structures). Even under teacher guidance, students' understanding of the CVS predicts their knowledge development on the

topic floating and sinking.

4.1. The association between CVS and students' content knowledge profiles

The relation of students' understanding of the CVS to their prior content knowledge was positive: Students' understanding of the CVS was a positive predictor for having one of the two more proficient knowledge profiles already at pretest. It is possible that this relation can be explained by students' broader reasoning abilities because these are associated with CVS (Bryant et al., 2015), or by their socioeconomic background. For example, parents' with higher socioeconomic background typically provide more support in students' everyday inquiry activities and critical thinking, which might contribute positively to their understanding of the CVS and also to their content knowledge development (Gleason & Schauble, 1999). Thus, students' socioeconomic background and general reasoning abilities both might positively influence their understanding of the CVS, and also their science content knowledge, explaining the statistical association. Nevertheless, skills based on the CVS have unique predictive value beyond general reasoning (Wüstenberg et al., 2012). Reasoning also has been found to be moderately related to the CVS mostly when it is based on verbal tasks, but the relation to general non-verbal reasoning seems to be rather weak (Mayer, Sodian, Koerber, & Schwippert, 2014; Wagenfeld et al., 2015). We therefore argue that our results indeed point towards a direct interrelation between the understanding of the CVS and science content knowledge. Crucial for education, the present finding implies that students entering instruction with less proficient prior content knowledge tend to also have a less advanced understanding of the CVS. At the same time, our longitudinal findings indicate that students with more advanced understanding of the CVS have a higher probability to transition and change from less proficient into more proficient content knowledge profiles during the instruction. Thus, students with a good understanding of the CVS, but less advanced prior knowledge still benefit from instruction.

A close look at the content knowledge profiles offers insight into conceptual knowledge structures. No knowledge profile comprised only one type of conceptions. For example, in the profile with the highest number of misconceptions, students concurrently held moderate numbers of intermediate conceptions and scientific concepts. On the transfer tasks, students in this profile outperformed those in a profile with a lower number of misconceptions. How can this be explained? We assume that this resulted from the co-existence of misconceptions with

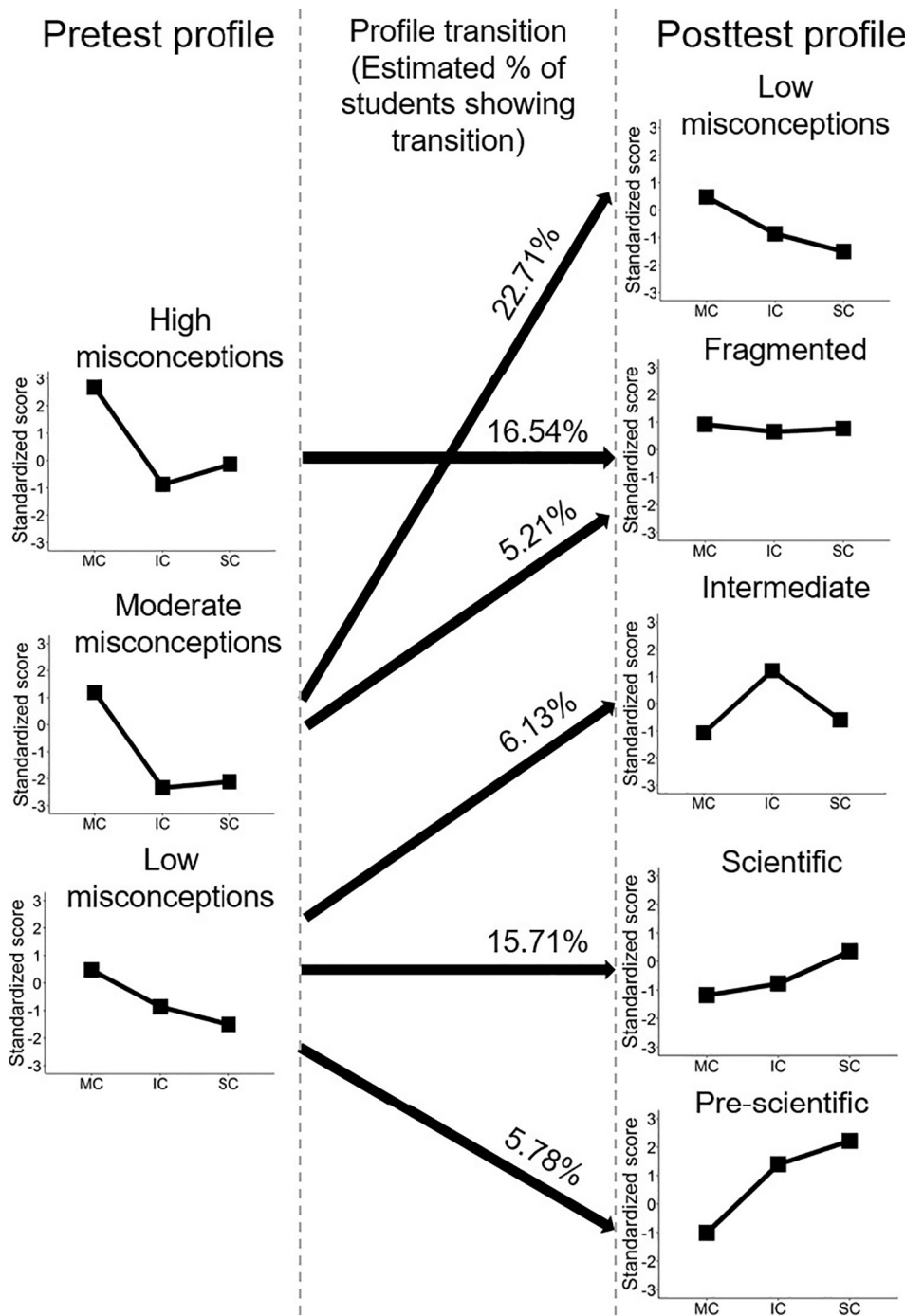


Fig. 5. The six most frequent content knowledge profile transitions from pretest to posttest. Note that only three of the four pretest profiles are shown because none of the most frequent transitions started from the *fragmented profile*.

more advanced conceptions: Misconceptions are not necessarily detrimental. Prior research indicates that having some misconceptions is not a major problem, as long as these misconceptions do not directly contradict related scientific concepts. The coexistence of a large number of conceptions is advantageous if conceptions are taken up in instruction to make students aware of their usefulness and limitations in different contexts (Ohlsson, 2009). Thus, if misconceptions might interfere with students' initial learning, in the long run it can be beneficial to discuss and reflect as many initial conceptions as possible (Kapur & Rummel, 2012). This can trigger powerful knowledge changes even if students' initial knowledge comprises primarily misconceptions (Ramsburg & Ohlsson, 2016).

4.2. The predictive value of CVS for students' content knowledge development

Further results illuminate the relation of students' understanding of the CVS and their content knowledge development from pretest to posttest. Most of the predictive estimates point towards a positive predictive value of the understanding of the CVS for content knowledge development: For students starting in the two content knowledge profiles with the highest number of misconceptions (i.e., the high and moderate misconceptions profiles), the probabilities to leave their initial profile and transition into a more proficient profile increased with increased understanding of the CVS (Fig. 6A & B). A likely

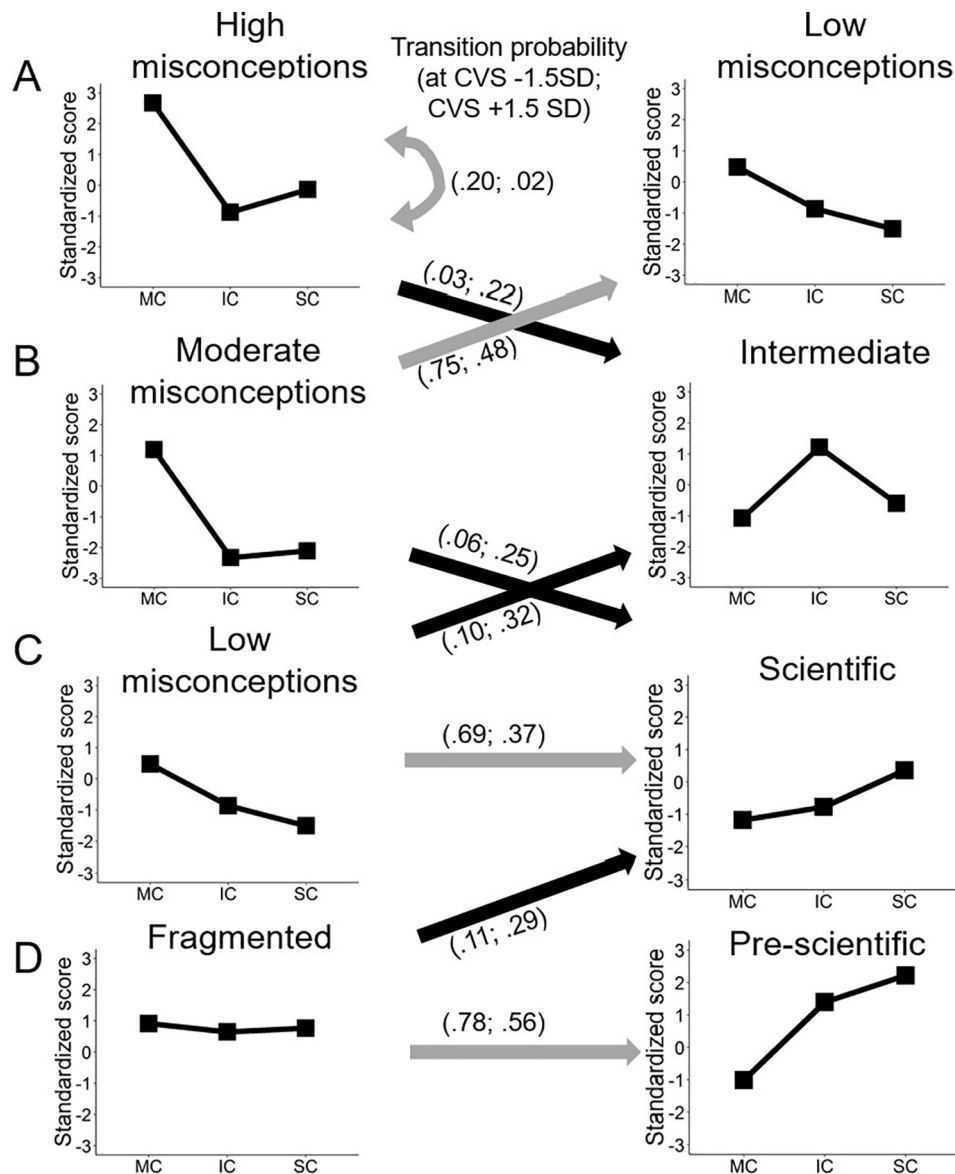


Fig. 6. Relations of students' CVS score to knowledge profile transitions from pretest to posttest. Left panel: THE four knowledge profiles present at pretest. Right panel: The four profiles at posttest for which transitions were predicted by students' understanding of the CVS. Gray arrows indicate transitions with decreasing probability for students with higher CVS scores, black arrows transitions with increasing probability. In accordance with arrow color, numbers between panels indicate transition probabilities for students with low CVS scores ($-1.5SD$, first number) and with high CVS scores ($+1.5SD$, second number). Curved arrow at high misconceptions profile indicates students staying in this profile.

interpretation of these results is that the understanding of the CVS helped these students to restructure their initial knowledge, and get rid of most of their misconceptions. For students starting in either of the two more proficient profiles, we found an opposed pattern. Students starting in a profile with a low number of misconceptions (Fig. 6C) were more likely to pass into a profile dominated by intermediate conceptions when they had a good understanding of the CVS. Students starting in the fragmented profile showed the contrary pattern (Fig. 6D): that is, they were more likely to end up in a profile dominated by scientific conceptions instead of both scientific and intermediate conceptions if they had a good understanding of the CVS. Their good understanding of the CVS thus might have helped them to either build up intermediate conceptions, or, in case they already had more intermediate conceptions initially, to get rid of these and instead develop scientific concepts. In a nutshell, this interpretation of the results can be summarized as follows: the understanding of the CVS helps students with lower-level prior knowledge pass into an intermediate profile. It helps those already in an intermediate profile to pass into a scientific profile. Thus,

independent of the content knowledge profile students have in the beginning, a good understanding of the CVS supports their knowledge restructuring and transitioning into a more scientifically valid understanding of floating and sinking.

Our analytic approach - latent transition analysis - offers valuable information on how to target specific groups of students. The knowledge profile with the highest number of misconceptions (i.e. about three standard deviations above mean) after instruction remained only for students with very low CVS understanding. Consequently, teachers should particularly support students lacking proper understanding of this experimentation skill in inquiry-based instruction. Such support might help them to get rid of their misconceptions and to restructure their relevant concept knowledge towards a more proficient knowledge profile.

A result demanding further discussion is that students with better CVS understanding were more likely to develop a high number of intermediate conceptions, instead of scientific concepts. A potential explanation is that intermediate conceptions, such as material kind and

object hollowness, describe directly visible object features. Such obvious object features can easily be tested in experiments by applying the CVS to test for their effects. Explanations based on scientific concepts, however, do not directly relate to individual, directly visible and testable object features. For example, the scientific concept of object density demands understanding of the relation between two object features, mass and volume. Additional skills and abilities, such as proportional reasoning ability, might interact with CVS in contributing to such understanding, while understanding of CVS on its own directly benefits the development of intermediate conceptions.

For students' content knowledge development in inquiry-based instruction, our results indicate that their understanding of the CVS might be a relevant constraint. If students do not understand principles of experimentation to a sufficient degree, they might not be able to grasp the importance and implications of well-controlled experiments (Schwichow et al., 2016). Children often see the meaning of experimentation in producing optimal outcomes, which has been termed the *engineering-approach* to experimentation (Schauble, Klopfer, & Raghavan, 1991). The controlled comparison between different conditions might be meaningful for students only if they understand its value for gaining conceptual insights about causal relations, beyond the aim of using an engineering-approach for producing optimal outcomes. For instruction, this would imply that preparing students with the necessary understanding of the CVS beforehand might support their conceptual learning from experimental designs. Students' understanding of the CVS can be trained efficiently with relatively low time investment, especially in teacher-directed training settings (Chen & Klahr, 1999; Lorch Jr et al., 2010; Strand-Cary & Klahr, 2008). We therefore suggest examining in future studies whether and under which circumstances training students' understanding of the CVS indeed raises their knowledge gains from subsequent inquiry-based classroom instruction.

4.3. Potential of the mixture modeling approach

The latent transition analysis yielded informative results in comparison to a traditional regression model. Given that interindividual differences in students' content knowledge structure exist, results from traditional regression models might be less informative than results from mixture modeling, because regression models are based on the assumption of a homogeneous multivariate normal distribution population across all students. Thus, in accordance with conceptual change theory, interindividual differences in students' knowledge and its development can be more adequately represented in mixture models. This comes at the price of setting up, estimating, and interpreting complex models with many parameters. In our study, this was possible due to the large sample size. With smaller sample sizes, such complex models cannot be estimated because they would not yield reliable information. However, in other studies on conceptual change, moderate sample sizes of a few hundred students were sufficient for estimating and yielding reliable information from similar models (McMullen et al., 2015; Schneider & Hardy, 2013; Straatemeier et al., 2008). Thus, in future quantitative studies on conceptual change, we suggest researchers to consider mixture modeling and to plan sample sizes accordingly. In case of moderate sample sizes, Bayesian estimation offers high model flexibility and ease of parameter estimation by incorporating information from prior studies into the models (see e.g., Chung, Lanza, & Loken, 2008; Etz et al., 2017). Most importantly, mixture modeling offers a way of yielding the information that conceptual change researchers really want from quantitative data.

4.4. Comparison to Schneider and Hardy (2013)

The present study represents a partial replication, and extension of Schneider and Hardy (2013). Schneider and Hardy (2013) used the same teaching materials and also a latent transition analysis approach to model third graders' knowledge profiles about floating and sinking.

Our results indicate that similar profiles are present in a new sample from a different country and across all grades of elementary school. Similar to Schneider and Hardy (2013), we found, for example, profiles with high numbers of misconceptions, and a profile with high numbers of intermediate conceptions and scientific concepts. We also found a fragmented profile, in which students showed higher-than-average numbers of all three types of concepts. We did however also find a profile that was not present in the data of Schneider and Hardy, the profile with a high number of intermediate conceptions. This profile was found only after instruction and revealed informative transition patterns.

The six most frequent transitions, comprising more than 70% of students, all described knowledge development into more advanced profiles. That is, these students all reduced their misconceptions, and acquired new intermediate conceptions or scientific concepts. This general pattern of results is in accordance with the findings of Schneider and Hardy (2013), indicating that most students manage to restructure their knowledge, and that although there is heterogeneity in students' knowledge, both their initial knowledge and its development can be described by a limited number of profiles and transition patterns. In addition, only about 4% of students stayed in their initial profile after instruction, similar to the findings by Schneider and Hardy (2013).

However, although the instruction and assessment materials were the same in our study and in Schneider and Hardy (2013), not all profiles we estimated concur with those found in their study. Differences might stem from various disparities in study characteristics. First, in our study first- to sixth graders instead of only third graders were examined; second, students stemmed from a different country (Switzerland instead of Germany); third, trained teachers instead of the study authors provided the instruction; and fourth, we analyzed raw scores instead of scaled scores, an analytic choice which might influence the findings of a latent transition analysis.

Overall, however, the major conclusions from both studies converge. This convergence is an important finding as to the best of our knowledge, such close conceptual replications using a mixture modeling approach have not yet been conducted in educational research. The similarity in findings between the two studies indicates that mixture modeling allows a rather robust description of learning patterns that can generalize across samples.

4.5. Limitations and outlook

In our study, all teachers received the same training but this does not imply that they all delivered the same instruction. It was not our intention to strictly control the instructional characteristics because we aimed at examining learning in realistic classroom situations. Still, in future studies, it might be informative to assess characteristics of teachers and their instruction to explain further variance in students' knowledge development. For example, the amount of teacher guidance provided during experimentation might influence to which degree students' own understanding of the CVS matters. From a statistical view, mixture models can be estimated as multilevel models, making it possible to estimate models similar to ours with predictors on the student level and also on the teacher level (Fagginger Auer et al., 2016; Vermunt, 2003).

Researchers can use the intraclass correlation estimates from Table A2 to inform whether multilevel modeling might be appropriate and informative in their future research.

Finally, our results show that students' understanding of the CVS is an informative predictor of their content knowledge development. But we did not randomly manipulate students' understanding of the CVS in a training study. Therefore, we do not know whether the present effects are direct and to which extent they are caused by other student characteristics such as intelligence and other cognitive abilities (Mayer et al., 2014, e.g., executive functions), or further endogeneity factors like feedback effects and common method variance from questionnaires

(Antonakis, Bendahan, Jacquart, & Lalive, 2014). In the present study, students' school grade was included in the analyses, which could not explain knowledge development beyond students' CVS scores. However, school grade is a limited proxy of intelligence. We doubt however that intelligence can fully account for the present findings because the predictive power of intelligence for learning is limited when students enter learning situations with relevant prior knowledge (Murayama, Pekrun, Lichtenfeld, & Vom Hofe, 2013). Its effect is often soaked up in the prior knowledge (Rütsche & Schalk, submitted; Schneider, Körkkel, & Weinert, 1989), although intelligence can sometimes explain variation in learning beyond prior knowledge (e.g., Primi, Ferrão, & Almeida, 2010). In addition, CVS has been found to predict science achievement beyond intelligence (Bryant et al., 2015). Furthermore, intelligence can be conceptualized as a network of interacting cognitive abilities (Conway & Kovacs, 2015; Kovacs & Conway, 2016; Van Der Maas et al., 2006). For future studies examining the causal status of the CVS based on longitudinal but non-experimental data, we suggest to examine the role of the CVS from a network perspective to yield insights into its dynamic relations with other constructs in the course of knowledge development. To further scrutinize whether the CVS has a direct causal effect on students' learning, we also re-emphasize our proposal to examine whether training students in CVS is indeed beneficial for their content knowledge development.

5. Conclusion

Our results indicate that students' understanding of the control-of-variables is relevant in teacher-guided inquiry-based instruction in

Appendix A. Additional information on the Bayesian change score model

In the latent change score model, no knowledge profiles are estimated, instead this is a traditional regression-based model in which variance in students' knowledge on the three scores at posttest was predicted by their knowledge at pretest, and additionally by their CVS score (Fig. A2).

The latent change score model was estimated using Bayesian estimation with Markov Chain Monte Carlo (MCMC) sampling. This estimation method can be more effective than maximum likelihood estimation and overcomes interpretational issues of other estimation methods and of p-value hypothesis testing (Edelsbrunner, 2014; Etz et al., 2017; Wagenmakers, 2007; Wagenmakers et al., 2016). We used default priors in Mplus and MCMC with Gibbs sampling with four chains, each with 10,000 draws, 5000 samples burn-in, and a thinning factor of 5.

Default priors were chosen because we did not have substantial prior information concerning the predictive strength of CVS measures for conceptual change in regression models, which is reflected in the default priors.

There were no convergence problems, potential scale reduction factors were equal or smaller than 1.020 for all estimated parameters (Brooks & Gelman, 1998). Posterior trace plots, density plots, and autocorrelation plots of the main parameters, the regression of the three knowledge indicator change scores on the CVS score, are depicted in Fig. A1. Posterior trace plots depict the parameter estimates yielded in each MCMC draw by the four chains (one in each color) and the figures indicate that the chains converged very early because the variances in the four chains seem to be mostly deviating randomly from each other, and stable around the same estimate. The posterior density plots indicate the distributions of these estimates from each fifth draw (i.e., after thinning to minimize the influence of potential autocorrelation between draws). These posteriors seem almost perfectly Gaussian, and the medians, which provide the point estimates for the respective parameter, are indicated by blue lines and corresponding numbers.

The autocorrelation plots, finally, indicate that consecutive MCMC draws were only very marginally dependent on each other, which increases trust in precise posterior estimates (for further guidance on interpreting Bayesian plots, see e.g. McElreath, 2015). All the actual estimated standardized model parameters are presented in Fig. A2.

elementary school. This finding opens new prospects for improving students' learning. Inquiry and other types of instruction have been researched for decades in science education. With the present study, we connected two research traditions by relating the CVS to inquiry-based learning in the classroom: Understanding of CVS is a main determinant for students' knowledge and their knowledge development in inquiry-based education. The latent transition analysis indicates that students' conceptual change depends on their knowledge structures. These findings are significant for theories of conceptual change; even the most advanced students do hold some misconceptions. We have shown that latent transition analysis is a tool that adequately and beneficially connects conceptual development theories and empirical observations. Future research might benefit from similar approaches to study whether and to what extent students' cognitive preconditions shape their learning of basic and advanced science concepts.

Acknowledgments

The Swiss MINT Study was funded by the Jacobs Foundation from 2010 to 2015 under the title “Boosting Hidden Potential in Science Education”. We are thankful to Mikko Kainulainen and Ansgar Opitz for helpful feedback on a draft of this manuscript and suggestions for figure design.

Ethical approval

Ethical approval was granted from the first author's institution under request number 2011-N-43.

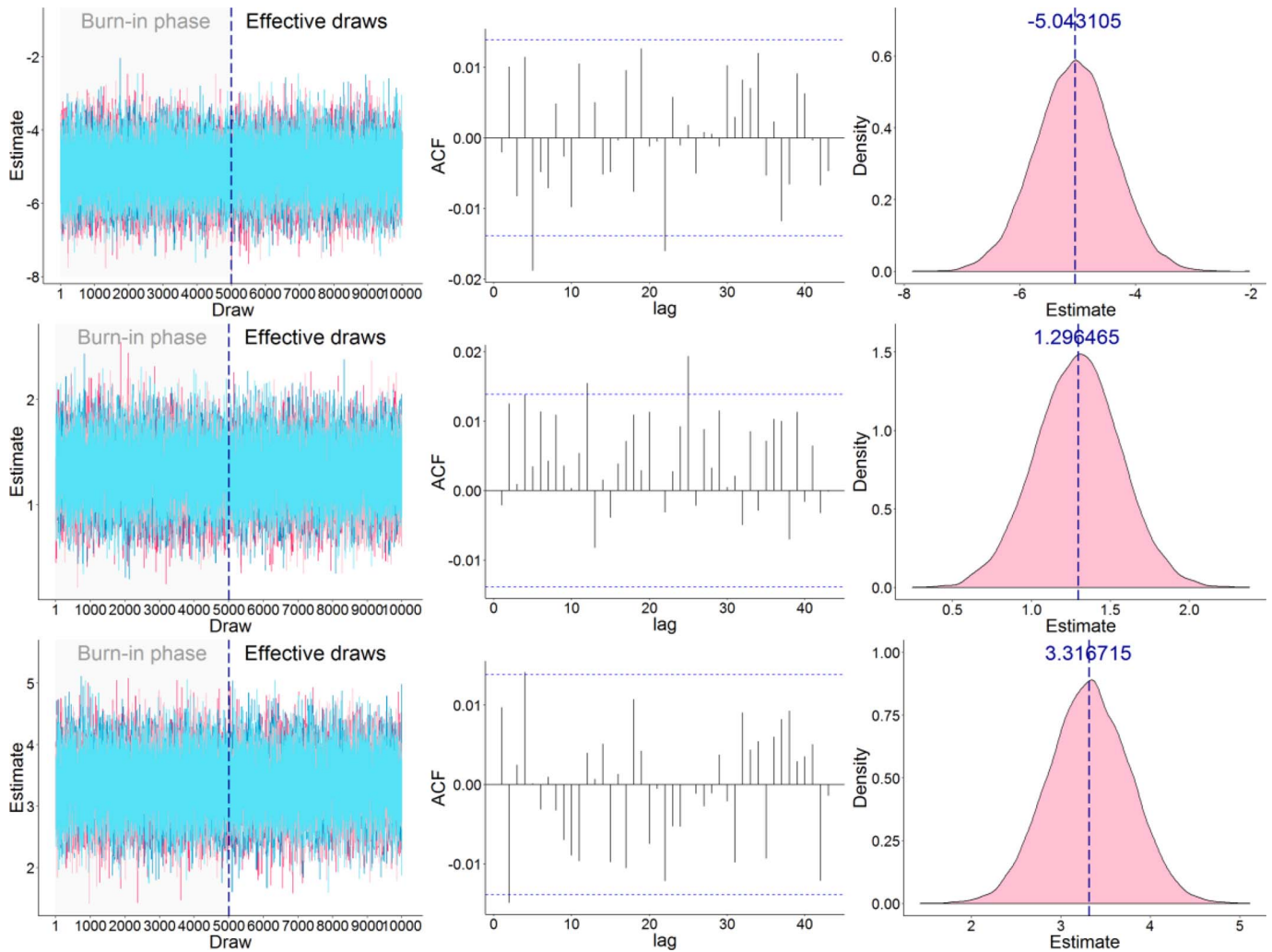


Fig. A1. Trace plots (left column panel), autocorrelation plots (middle column panel), and posterior density plots (right column panel, median indicated) from the latent change score-model to predict students' prior knowledge and knowledge change from their understanding of the control of variables-strategy. The plots stem from the MCMC draws of the three unstandardized parameters for the regression of students' change on the misconceptions score (upper panel), intermediate conceptions score (middle panel), and scientific concepts score (lower panel) on their CVS score.

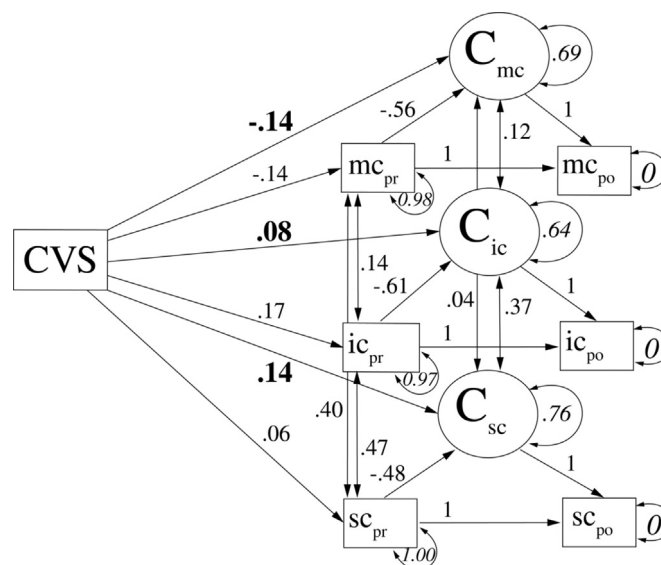


Fig. A2. The latent change score-model to predict students' knowledge change from their understanding of the CVS. CVS = CVS questionnaire mean score; C_{mc} = change on misconceptions score from pretest mc_{pr} to posttest mc_{po} ; C_{ic} = change on intermediate conceptions score from pretest (ic_{pr}) to posttest (ic_{po}); C_{sc} = change on scientific concepts score from pretest (sc_{pr}) to posttest (sc_{po}). Double-headed arrows indicate variance or correlation parameters, single-headed arrows regression parameters. Main parameter estimates are the predictive paths in bold from CVS to C_{mc} , C_{ic} , and C_{sc} . Change score variables represent intercepts and residual variances at posttest, controlling for pretest values, like in a generic

panel regression model. 0 s and 1 s are fixed parameters, the other values represent standardized Bayesian parameter estimates.

Table A1
Transition probabilities between all profiles for students high (upper panel) and low (lower panel) on CVS.

CVS – 1.5SD		Post					
		LMC (0.41)	HMC (0.08)	FRA (0.27)	PRE (0.01)	INT (0.04)	SCI (0.14)
Pre	MMC (0.47)	0.75	0.02	0.14	0.00	0.03	0.06
	LMC (0.16)	0.00	0.00	0.02	0.19	0.10	0.69
	HMC (0.36)	0.17	0.20	0.56	0.04	0.03	0.00
	FRA (0.01)	0.00	0.00	0.00	0.78	0.12	0.11
CVS + 1.5SD		Post					
		LMC (0.12)	HMC (0.01)	FRA (0.16)	PRE (0.23)	INT (0.22)	SCI (0.26)
Pre	MMC (0.21)	0.48	0.04	0.18	0.00	0.05	0.25
	LMC (0.44)	0.00	0.00	0.05	0.26	0.32	0.37
	HMC (0.19)	0.10	0.02	0.54	0.13	0.22	0.00
	FRA (0.16)	0.00	0.00	0.00	0.56	0.15	0.29

Note. MMC = moderate misconceptions profile; LMC = low misconceptions profile; HMC = high misconceptions profile; FRA = fragmented profile; PRE = prescientific profile; INT = intermediate profile; SCI = scientific profile. Numbers in brackets next to profiles indicate profile size at pretest (Pre) and posttest (Post). All other numbers indicate transition probabilities between respective profiles. No column for MMC shown because this profile was not present at posttest.

Table A2
Descriptive statistics of study variables: Means, SDs, estimates of intraclass correlation coefficients, (co-)variances and correlations.

	Grade	Gender	CVS score	MC_pr	MC_po	IC_pr	IC_po	SC_pr	SC_po
Mean	3.58	0.51	0.47	17.93	11.60	6.75	8.83	6.79	10.12
SD	1.23	0.50	0.20	5.62	6.24	2.51	2.39	3.34	4.16
ICC	0.94	0.00	0.12	0.19	0.22	0.11	0.18	0.10	0.24
Grade	1.51	0.01	0.15	-0.23	-0.22	0.11	-0.01	-0.05	0.18
Gender	0.01	0.25	-0.01	-0.08	-0.04	0.09	0.04	0.03	0.02
CVS score	0.04	0.00	0.04	-0.14	-0.22	0.18	0.16	0.07	0.18
MC_pr	-1.62	-0.24	-0.16	31.53	0.40	-0.01	0.06	0.40	0.02
MC_po	-1.70	-0.13	-0.28	14.02	38.92	-0.15	-0.07	0.11	-0.10
IC_pr	0.34	0.11	0.09	-0.17	-2.34	6.31	0.32	0.35	0.28
IC_po	-0.03	0.05	0.08	0.77	-1.10	1.93	5.73	0.19	0.33
SC_pr	-0.19	0.05	0.05	7.49	2.31	2.93	1.52	11.15	0.32
SC_po	0.93	0.04	0.15	0.35	-2.46	2.94	3.28	4.46	17.28

Note. MC_pr = misconceptions score at pretest; MC_po = misconceptions score at posttest; IC_pr = intermediate conceptions score at pretest; IC_po = intermediate conceptions score at posttest; SC_pr = scientific concepts score at pretest; SC_po = scientific concepts score at posttest; ICC = intraclass correlation coefficient. Variable intercorrelations are presented above the diagonal, covariances below the diagonal. Covariances of variables with themselves on diagonal are variances. Covariances and correlations are full information maximum likelihood estimates.

Estimates of ICCs stem from two level model in which mean parameters are modeled varying randomly across teachers.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*, 1–18. <http://dx.doi.org/10.1037/a0021017>.

American Association for the Advancement of Science (1993). *Benchmarks for science literacy*. USA: Oxford University Press.

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education, 13*, 1–12. <http://dx.doi.org/10.1023/A:1015171124982>.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. *The Oxford Handbook of Leadership and Organizations, 93–117*.

Bennett, J., Lubben, F., & Hogarth, S. (2007). Bringing science to life: A synthesis of the research evidence on the effects of context-based and STS approaches to science teaching. *Science Education, 91*, 347–370. <http://dx.doi.org/10.1002/sce.20186>.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*, 434–455. <http://dx.doi.org/10.1080/10618600.1998.10474787>.

Bryant, P., Nunes, T., Hillier, J., Gilroy, C., & Barros, R. (2015). The importance of being able to deal with variables in learning science. *International Journal of Science and Mathematics Education, 13*, 145–163. <http://dx.doi.org/10.1007/s10763-013-9469-x>.

Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. *Individual development from 3 to 12: Findings from the Munich longitudinal study* (pp. 38–54). Cambridge University Press.

Carey, S. (1985). *Conceptual change in childhood*. MA: Cambridge: MIT Press.

Carey, S. (1992). *The origin and evolution of everyday concepts*. Minneapolis: University of Minnesota Press.

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](http://dx.doi.org/10.1016/S0193-3973(99)00046-5).

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70*, 1098–1120. <http://dx.doi.org/10.1111/1467-8624.00081>.

Chi, M. T. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 61–82). New York: Routledge.

Chi, M. T., & Ohlsson, S. (2005). In K. J. Holyoak, & R. G. Morrison (Eds.), *Complex declarative learning*. Cambridge University Press.

Christou, K. P., & Vosniadou, S. (2012, January). What kinds of numbers do students assign to literal symbols? Aspects of the transition from arithmetic to algebra. *Mathematical Thinking and Learning, 14*, 1–27. <http://dx.doi.org/10.1080/10986065.2012.625074>.

Chung, H., Lanza, S. T., & Loken, E. (2008, May 20). Latent transition analysis: Inference

- and estimation. *Statistics in Medicine*, 27, 1834–1854. <http://dx.doi.org/10.1002/sim.3130>.
- Conway, A. R., & Kovacs, K. (2015). New and emerging models of human intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6, 419–426. <http://dx.doi.org/10.1002/wcs.1356>.
- Croker, S., & Buchanan, H. (2011, September). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies: children's hypothesis testing. *British Journal of Developmental Psychology*, 29, 409–424. <http://dx.doi.org/10.1348/026151010X496906>.
- diSessa, A. A. (2008). A theory bite on the meaning of scientific inquiry. A companion to Kuhn and Pease. *Cognition and Instruction*, 26, 560–566. <http://dx.doi.org/10.1080/07370000802391760>.
- Edelsbrunner, P. A. (2014, November). Bayesian statistics: What is it and why do we need it? [Blog post]. Retrieved from <http://blog.efpsa.org/2014/11/17/bayesian-statistics-what-is-it-and-why-do-we-need-it-2/>.
- Etz, A., Gronau, Q. F., Dablander, F., Edelsbrunner, P. A., & Barbault, B. (2017, June). How to become a Bayesian in eight easy steps: An annotated reading list. *Psychonomic Bulletin & Review*. <http://dx.doi.org/10.3758/s13423-017-1317-5> (bibtex: Etz2017 bibtex[day=28]).
- Fagginger Auer, M. F., Hickendorff, M., Van Putten, C. M., Béguin, A. A., & Heiser, W. J. (2016, April 2). Multilevel latent class analysis for large-scale educational assessment data: Exploring the relation between the curriculum and students' mathematical strategies. *Applied Measurement in Education*, 29, 144–159. <http://dx.doi.org/10.1080/08957347.2016.1138959>.
- Finch, W. H., & Bronk, K. C. (2011, January 13). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 18, 132–151. <http://dx.doi.org/10.1080/10705511.2011.532732>.
- Flick, L. B. (1995). Complex instruction in complex classrooms: A synthesis of research on inquiry teaching methods and explicit teaching strategies. *Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco (ED 383 563)*.
- Frède, V., Nobes, G., Frappart, S., Panagiotaki, G., Troadec, B., & Martin, A. (2011). The acquisition of scientific knowledge: The influence of methods of questioning and analysis on the interpretation of children's conceptions of the earth. *Infant and Child Development*, 20, 432–448. <http://dx.doi.org/10.1002/icd.730>.
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012, August 8). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82, 300–329. <http://dx.doi.org/10.3102/0034654312457206>.
- Gleason, M. E., & Schauble, L. (1999, December). Parents' assistance of their children's scientific reasoning. *Cognition and Instruction*, 17, 343–378. http://dx.doi.org/10.1207/S1532690XCI1704_1.
- Hardy, I., Jonen, A., Moeller, K., & Stern, E. (2006). Effects of instructional support within constructivist learning environments for elementary school students' understanding of "floating and sinking". *Journal of Educational Psychology*, 98, 307–326. <http://dx.doi.org/10.1037/0022-0663.98.2.307>.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London; New York: Routledge.
- Hickendorff, M., Edelsbrunner, P. A., Schneider, M., Trezise, K., & McMullen, J. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analyses. *Learning and Individual Differences*. <http://dx.doi.org/10.1016/j.lindif.2017.11.001> (in press).
- Kapur, M., & Rummel, N. (2012). Productive failure in learning from generation and invention activities. *Instructional Science*, 40, 645–650. <http://dx.doi.org/10.1007/s11251-012-9235-4>.
- Klahr, D., & Chen, Z. (2003). Overcoming the positive-capture strategy in young children: Learning about indeterminacy. *Child Development*, 74, 1275–1296.
- Kleickmann, T., Hardy, I., Pollmeier, J., & Möller, K. (2011). Zur struktur naturwissenschaftlichen wissens von grundschulkindern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 200–212. <http://dx.doi.org/10.1026/0049-8637/a000053>.
- Kovacs, K., & Conway, A. R. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, 27, 151–177. <http://dx.doi.org/10.1080/1047840X.2016.1153946>.
- Kuhn, D. (2002). In U. Goswami (Ed.), *What is scientific thinking and how does it develop?* (pp. 497). Oxford, England: Blackwell.
- Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction*, 18, 495–523. http://dx.doi.org/10.1207/S1532690XCI1804_3.
- Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (2008). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*, 23, 435–451.
- Kuhn, D., Ramsey, S., & Arvidsson, T. S. (2015, July). Developing multivariable thinkers. *Cognitive Development*, 35, 92–110. <http://dx.doi.org/10.1016/j.cogdev.2014.11.003>.
- Leuchter, M., Saalbach, H., & Hardy, I. (2014, July 3). Designing science learning in the first years of schooling. An intervention study with sequenced learning material on the topic of 'floating and sinking'. *International Journal of Science Education*, 36, 1751–1771. <http://dx.doi.org/10.1080/09500693.2013.878482>.
- Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, 36, 837–858.
- Linn, M. C., Eylon, B.-S., & Davis, E. A. (2004). The knowledge integration perspective on learning. *Internet environments for science education* (pp. 29–46).
- Liu, X. (2004, May). Using concept mapping for assessing and promoting relational conceptual change in science. *Science Education*, 88, 373–396. <http://dx.doi.org/10.1002/see.10127>.
- Lorch, R. F., Jr., Lorch, E. P., Calderhead, W. J., Dunlap, E. E., Hodell, E. C., & Freer, B. D. (2010). Learning the control of variables strategy in higher and lower achieving classrooms: Contributions of explicit instruction and experimentation. *Journal of Educational Psychology*, 102, 90–101. <http://dx.doi.org/10.1037/a0017972>.
- Maclin, D., Grosslight, L., & Davis, H. (1997, September). Teaching for understanding: A study of students' preinstruction theories of matter and a comparison of the effectiveness of two approaches to teaching about matter and density. *Cognition and Instruction*, 15, 317–393. http://dx.doi.org/10.1207/s1532690xci1503_2.
- Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, 16, 191–225. <http://dx.doi.org/10.1080/10705510902751010>.
- Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014, February). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, 29, 43–55. <http://dx.doi.org/10.1016/j.learninstruc.2013.07.005>.
- McArdle, J. J. (2009, January). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605. <http://dx.doi.org/10.1146/annurev.psych.60.110707.163612>.
- McElreath, R. (2015). *Statistical rethinking: A bayesian course with examples in r and stan*. Boca Raton: Chapman and Hall-CRC.
- McMullen, J., Laakkonen, E., Hannula-Sormunen, M., & Lehtinen, E. (2015, June). Modeling the developmental trajectories of rational number concept(s). *Learning and Instruction*, 37, 14–20. <http://dx.doi.org/10.1016/j.learninstruc.2013.12.004>.
- Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching*, 47, 474–496. <http://dx.doi.org/10.1002/tea.20347>.
- Möller, K., Hardy, I., Jonen, A., Kleickmann, T., & Blumberg, E. (2006). Naturwissenschaften in der primarstufe: zur förderung konzeptuellen verständnisses durch unterricht und zur wirksamkeit von lehrerfortbildungen. *Untersuchungen zur bildungsqualität von schule: Abschlussbericht des dfg-schwerpunktprogramms* (pp. 161–193). Waxmann.
- Murayama, K., Pekrun, R., Lichtenfeld, S., & Vom Hofe, R. (2013). Predicting long-term growth in students' mathematics achievement: The unique contributions of motivation and cognitive strategies. *Child Development*, 84, 1475–1490. <http://dx.doi.org/10.1111/cdev.12036>.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.
- Nussbaum, J., & Novak, J. D. (1976). An assessment of children's concepts of the earth utilizing structured interviews. *Science Education*, 60, 535–550. <http://dx.doi.org/10.1002/see.3730600414>.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14, 535–569.
- Ohlsson, S. (2009, January 23). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist*, 44, 20–40. <http://dx.doi.org/10.1080/00461520802616267>.
- Osterhaus, C., Körber, S., & Sodian, B. (2016). Experimentation skills in primary school: An inventory of children's understanding of experimental design. *Frontline Learning Research*, 3, 56–94.
- Penner, D. E., & Klahr, D. (1996). The interaction of domain-specific knowledge and domain-general discovery strategies: A study with sinking objects. *Child Development*, 67, 2709–2727.
- Piekny, J., & Maehler, C. (2013, June). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology*, 31, 153–179. <http://dx.doi.org/10.1111/j.2044-835X.2012.02082.x>.
- Primi, R., Ferrão, M. E., & Almeida, L. S. (2010, October). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences*, 20, 446–451. <http://dx.doi.org/10.1016/j.lindif.2010.05.001>.
- Ramsburg, J. T., & Ohlsson, S. (2016). Category change in the absence of cognitive conflict. *Journal of Educational Psychology*, 108, 98. <http://dx.doi.org/10.1037/edu0000050>.
- Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, 17, 265–279. <http://dx.doi.org/10.1080/10705511003659417>.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47, 667–696. <http://dx.doi.org/10.1080/00273171.2012.715555>.
- Rütsche, B., & Schalk, L. (2018). *Constructing and refining relational categories: Disentangling the interplay between prior knowledge, reasoning, working memory capacity, and strategy*. Manuscript submitted for publication. (submitted).
- Schalk, L., Edelsbrunner, P. A., Schumacher, R., & Stern, E. (2018). *Early inquiry-based science education matters beyond content learning*. (submitted).
- Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology*, 49, 31–57. [http://dx.doi.org/10.1016/0022-0965\(90\)90048-D](http://dx.doi.org/10.1016/0022-0965(90)90048-D).
- Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology*, 32, 102. <http://dx.doi.org/10.1037/0012-1649.32.1.102>.
- Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, 28, 859–882. <http://dx.doi.org/10.1002/tea.3660280910>.
- Schneider, M., & Hardy, I. (2013). Profiles of inconsistent knowledge in children's pathways of conceptual change. *Developmental Psychology*, 49, 1639. <http://dx.doi.org/10.1037/a0030976>.
- Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and

- memory performance: A comparison of high-and low-aptitude children. *Journal of Educational Psychology*, 81, 306. <http://dx.doi.org/10.1037/0022-0663.81.3.306>.
- Schwichow, M., Christoph, S., Boone, W. J., & Härtig, H. (2016). The impact of sub-skills and item content on students' skills with regard to the control-of-variables strategy. *International Journal of Science Education*, 38, 216–237. <http://dx.doi.org/10.1080/09500693.2015.1137651>.
- Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching*, 27, 127–144. <http://dx.doi.org/10.1002/tea.3660270205>.
- Siler, S. A., & Klahr, D. (2012). Detecting, classifying, and remediating: children's explicit and implicit misconceptions about experimental design date of original version. In R. W. Proctor, & E. J. Capaldi (Eds.). *Psychology of science: Implicit and explicit processes*. New York: Oxford University Press.
- Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2014, September). Experimental evaluations of elementary science programs: A best-evidence synthesis. *Journal of Research in Science Teaching*, 51, 870–901. <http://dx.doi.org/10.1002/tea.21139>.
- Smith, C., Carey, S., & Wisner, M. (1985). On differentiation: A case study of the development of the concepts of size, weight, and density. *Cognition*, 21, 177–237.
- Smith, C. L. (2007). Bootstrapping processes in the development of students common-sense matter theories. Using analogical mappings, thought experiments, and learning to measure to promote conceptual restructuring. *Cognition and Instruction*, 25, 337–398. <http://dx.doi.org/10.1080/07370000701632363>.
- Sodian, B., & Bullock, M. (2008, October). Scientific reasoning—Where are we now? *Cognitive Development*, 23, 431–434. <http://dx.doi.org/10.1016/j.cogdev.2008.09.003>.
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62, 753–766. <http://dx.doi.org/10.1111/j.1467-8624.1991.tb01567.x>.
- Stern, E., Schumacher, R., Edelsbrunner, P. A., Schalk, L., & Deiglmayr, A. (2016). *International Journal of Psychology*, 360.
- Straatemeier, M., van der Maas, H. L., & Jansen, B. R. (2008, August). Children's knowledge of the earth: A new methodological and statistical approach. *Journal of Experimental Child Psychology*, 100, 276–296. <http://dx.doi.org/10.1016/j.jecp.2008.03.004>.
- Strand-Cary, M., & Klahr, D. (2008). Developing elementary science skills: Instructional effectiveness and path independence. *Cognitive Development*, 23, 488–511. <http://dx.doi.org/10.1016/j.cogdev.2008.09.005>.
- Van Der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113, 842. <http://dx.doi.org/10.1037/0033-295X.113.4.842>.
- Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, 33, 213–239. <http://dx.doi.org/10.1111/j.0081-1750.2003.t01-1-00131.x>.
- Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535–585. [http://dx.doi.org/10.1016/0010-0285\(92\)90018-W](http://dx.doi.org/10.1016/0010-0285(92)90018-W).
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804. <http://dx.doi.org/10.3758/BF03194105>.
- Wagenmakers, E.-J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science*, 25, 169–176. <http://dx.doi.org/10.1177/0963721416643289>.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 632–638. <http://dx.doi.org/10.1177/1745691612463078>.
- Wagensveld, B., Segers, E., Kleemans, T., & Verhoeven, L. (2015, May). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, 43, 365–379. <http://dx.doi.org/10.1007/s11251-014-9334-5>.
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence*, 40, 1–14. <http://dx.doi.org/10.1037/a0030976>.
- 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>.
- Zimmerman, C. (2007, June). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27, 172–223. <http://dx.doi.org/10.1016/j.dr.2006.12.001>.