Asterhan, C. S. C. & Dotan, A. (in press). Feedback that corrects and contrasts students' erroneous solutions with expert ones improves expository instruction for conceptual change. *Instructional Science.* DOI: 10.1007/s11251-017-9441-1

December 7, 2017

Author Note

Correspondence should be addressed to Christa Asterhan, School of Education, Hebrew University of Jerusalem, Mount Scopus, Jerusalem, Israel, 91905. E-mail: asterhan@huji.ac.il

Abstract

In the present study, we examined the effects of feedback that corrects and contrasts a student's own erroneous solutions with the canonical, correct one (CEC&C feedback) on learning in a conceptual change task. Sixty undergraduate students received expository instruction about natural selection, which presented the canonical, scientifically accepted account in detail. Two-third of these received CEC&C feedback on their self-generated solutions to open-ended test items. Students either received this feedback on their pretest solutions (prior to instruction), or on their immediate posttest solutions (following instruction). Students in the control condition only received the correct canonical answers to the immediate post-test items and compared these with their own solutions autonomously. Conceptual understanding on transfer items was assessed after one week. Results showed that students in the CEC&C feedback conditions outperformed control students. Timing of feedback did not affect learning, however. These findings add to accumulating evidence from different lines of research on the importance of instructional support that explicitly compares and contrasts between erroneous student models and canonical models in conceptual change tasks.

Introduction

A vast body of research has demonstrated effects of feedback on student learning in a range of settings (see Hattie & Timperley, 2007; Kluger & DeNisi, 1996, for meta-analyses and overviews). Surprisingly, and perhaps due to its strong roots in constructivism and student-directed learning, the role of externally provided feedback has received little empirical attention in the conceptual change literature, however. In the present study, we address this gap in the literature by testing the effects of feedback on learning that requires conceptual change. We consider both the content as well as the timing of feedback (prior to or after instruction). In order to be effective, feedback in conceptual change tasks should support the cognitive processes involved in conceptual change and design decisions about content and timing of feedback should be closely related to current theoretical accounts of this type of learning. We first provide a short overview of current thinking on conceptual change and existing approaches to instruction for conceptual change.

Conceptual change

For more than four decades, scholars from science education, developmental psychology and cognitive science have documented how children's and adults' naïve theories about natural phenomena do not align with the scientifically accepted, but often counter-intuitive concepts that they are exposed to in science instruction. A group of misconceptions that are known to be particularly resistant to change concern emergent processes, which students often misinterpret for sequential processes (Chi, 2009; Chi, Roscoe, Slotta, Roy, & Chase, 2012). A sequential process is, among others, characterized by the fact that it has a clear beginning and end, a sequence of distinct actions that are contingent and causal, and an identifiable, explicit goal. Emergent processes, such as for example diffusion and evolution, on the other hand are uniform, simultaneous and ongoing, and have no clear goal (Chi, 2008; Chi et al, 2012). The observable outcomes of these processes often resemble sequential processes, however. Coming to understand and being able to correctly use these canonical, scientific explanations is not a matter of "gap-filling", in that learners lack the necessary knowledge, but rather involves a substantive re-organization of existing, misconceived, intuitive knowledge, an outcome which is usually referred to as "conceptual change" (e.g., Chi, 2008; Thagard, 1992; Vosniadou & Brewer, 1994).

Traditionally, conceptual change has often been described as a correction or replacement of misconceived conceptions that reside in the mind (e.g., reviews in Özdemir & Clark, 2007; Vosniadou, 2009). Current cognitive accounts of conceptual change describe it in terms of a response competition at a deeper cognitive level. Accordingly, it constitutes an increase in the probability with which more advanced schema configurations are activated and used to construct temporary mental representations in working memory, when an individual is required to apply that knowledge to solve a problem (e.g., Ohlsson, 2002; Potvin, Sauriol, & Riopel, 2015; Ramsburg & Ohlsson, 2016; Schnotz & Preuss, 1999). This response competition account is further supported by recent empirical evidence showing that conceptual change involves both an improved capability to construct the correct scientific explanation, as well as more efficient inhibition of automatically activated, but irrelevant schemas and propositions (e.g., Babai, Sekal & Stavy, 2010; Dunbar, Fugelsang & Stein, 2007; Masson, Potvin, Riopel, & Foisy, 2014; Shtulman & Valcarcel, 2012; Potvin, Masson, Lafortune & Cyr, 2015).

Instructional approaches to conceptual change

In order to be effective, instructional approaches for conceptual change should then preferably support both these cognitive processes: to provide students with opportunities to become aware of and understand the errors in (their) naïve theories, as well to fully comprehend the scientifically accepted theory that is often counterintuitive to everyday experiences (see also Chan, Burtis & Bereiter, 1997). Yet, in traditional tell-and-practice instruction, learners are presented with the correct scientific explanations and then practice this newly acquired knowledge with further exercises. Perhaps not surprisingly, then, tell-and-practice instruction has not been found to be very effective for learning that requires conceptual change, especially in the case of robust misconceptions (Chi, 2008; Vosniadou & Mason, 2013).

In their search for alternative instructional approaches, scholars of conceptual change have been heavily influenced by Piagetian ideas. Two characteristics stand out in this research tradition: (1) A strong emphasis on task designs that are meant to induce learner awareness of and dissatisfaction with their own intuitive understanding; and (2) the expectation that students will understand or even arrive at the correct scientific explanation by themselves. For example, students are asked to solve a set of problems according to their own naïve understanding and then presented with contradictory information by running an experiment (e.g., Howe, Tolmie, Duchak-Tanner, & Rattay, 2000; Schwarz, Neuman & Biezuner, 2000), or they are paired with another student who has a different understanding (e.g., Ames & Murray, 1982; Asterhan & Schwarz, 2007, 2009; Doise & Mugny, 1978). In these approaches, learners rarely receive detailed feedback about their errors and often have to deduce the correct explanations themselves. However, students experience substantive difficulty in doing so (e.g., Chinn & Brewer, 1993; Fugelsang & Dunbar, 2005). Thus, even though these instructional methods have produced some positive effects, these have not been consistent across studies and are overall not very strong (Límon, 2001; Ramsburg & Ohlsson, 2016).

In recent years, there has been increasing interest in instructional approaches that contrast common, erroneous student solutions with the correct explanation or procedure. This type of instruction is a particular case of the more general *contrasting cases approach*, which includes a range of techniques in which students are given two contrasting cases, solutions or examples which are then compared (e.g., Bransford & Schwartz, 1998). The more specific technique of *contrasting erroneous* solutions with the canonical, *correct* solution on the critical features (CEC) has recently been tested in several studies and found effective in different formats, such as classroom teaching (Loibl & Rummel, 2013; Kapur & Bielaczyc, 2012), refutation text reading (e.g., Diakidoy, Kendeou, & Ioannides, 2003; Sinatra & Broughton, 2011) and teacher-provided worksheets with common erroneous student solutions (Durkin & Rittle-Johnson, 2012; Gadgil, Nokes & Chi, 2012).

In contrast to the more traditional cognitive conflict-based instructional approaches, CEC-based instruction does not rely on students to detect the critical differences between common misconceptions and the canonical, correct explanation on their own and/or to generate the correct account by themselves. Having said that, however, the effectiveness of CEC may be further increased by tailoring it to the individual learner: The aforementioned implementations of the CEC approach rely on the use of pre-prepared materials and are therefore not tailored to the specific errors made by a given individual student. Even though the most frequently encountered student misconceptions are presented, the individual learner may not share, understand or recognize the targeted misconception that is presented in pre-prepared CEC materials. Overall, personalized feedback that compares and contrasts the students' *own* personal solution to a canonical correct explanation may then be a more productive approach.

Feedback and conceptual change

To our best knowledge, the issue of whether and how feedback may improve learning on conceptual change tasks has not been subjugated to extensive empirical research. It is likely that in some studies, especially those that were conducted in in-vivo classroom settings, students received some form of feedback as part of an overall learning sequence, either through the task itself (e.g., Light & Glachan, 1985; Schwarz & Linchevski, 2007; Schwarz et al., 2000) or through a human partner (e.g., Asterhan & Schwarz, 2007; Asterhan, Schwarz & Cohen-Eliyahu, 2014). However, feedback as an instructional method has rarely been a focus for empirical investigations in the conceptual change literature.

This paucity of empirical attention is surprising, since research on the role of feedback in other types of learning (procedural knowledge, factual knowledge, fluency) is abundant. In fact, feedback effects are among the most robust and longstanding findings in the empirical literature. Perhaps it is best explained against the historical rift between constructivist and behaviorist research traditions: Whereas conceptual change research originated from Piagetian theory, research on feedback has strong behaviorist, Skinnerian roots. Given the modest effects of other instructional methods for conceptual change and given our extensive knowledge about feedback effects, however, empirical investigation into the effects of feedback in conceptual change tasks are timely.

Findings from a recent research by Asterhan et al. (2014) shows that the role of feedback in conceptual change tasks deserves further attention: They tested the effects of outcome feedback (whether the response is correct or not) on 9th graders' proportional reasoning in a range of different dyadic learning set-ups. Results showed that providing outcome feedback improved learning gains only when learners had also access to the correct explanation. Thus, simply knowing that one had erred (outcome feedback) did not cause students to deduce the correct explanation on their own. Vice versa, only having access to the correct explanation did not improve learning either. The *combination* of outcome feedback with access to the correct explanation was key to conceptual change in this task. These findings align well with the aforementioned current views on conceptual change processes: irrelevant knowledge structures should be recognized and actively inhibited and access to relevant ones should be facilitated.

Whereas outcome feedback goes some way, however, the literature on feedback shows that effects are strongest for corrective, elaborated feedback (Hattie & Timperley, 2007). Corrective, elaborated feedback not only states whether the response is wrong, but also indicates where the mistakes are made, why they are mistakes and corrects them. Based on the findings from the CEC research and current theoretical models of conceptual change, effective feedback for conceptual change should then in addition include explicit comparisons with an expert solution. We term this type of feedback CEC&C (Contrast between Erroneous and Canonical solution and Correct). In comparison, only receiving an expert solution and letting students deduce the differences by themselves is likely to be insufficient (Loibl & Rummel, 2013). Based on the aforementioned rationale, it is therefore expected that students who receive instructor-provided CEC&C feedback on their own solutions will gain better conceptual understanding than students who are given an expert solution and compare it with their own solution, without further support (H1).

Timing of feedback

Another issue to consider when providing feedback in conceptual change tasks is its timing. Existing research on feedback effects has compared immediate and delayed feedback on student-produced responses to practice items that follow a period of instruction (e.g., Butler, Karpicke, & Roediger, 2007; Kulik & Kulik, 1988). In conceptual change tasks, however, students already have, by definition, some existing (albeit misconceived) knowledge about the topic. Being made aware of this *prior* to instruction may increase their attention to pivotal aspects of the scientific explanation during the instruction phase (Richland, Kornell, & Kao, 2009). Moreover, when the differences between the misconceived and the correct explanations are highlighted upfront, the information provided in the instruction phase is less likely to be assimilated into existing erroneous knowledge structures. In both cases, students who receive CEC&C feedback prior to instruction are expected to benefit more from this expository instruction phase, compared to students who do not (H2).

However, it is less clear whether receiving CEC&C feedback prior to instruction will also produce better learning when compared to receiving it *after* the instruction phase. On the one hand, making students aware of their erroneous, intuitive understanding of a phenomenon before presenting them the canonical scientific explanation may indeed lead to a better accessible, more stable understanding of the correct account. When receiving CEC&C feedback only after the instruction and the subsequent practice questions, learners may have already further consolidated the incorrect explanation structures. Moreover, since feedback is given post factum, students do not have further opportunities to practice and consolidate the correct explanation structures.

On the other hand, however, these expectations are somewhat tempered by recent findings by Loibl and Rummel (2013). In their study, teachers compared and contrasted between erroneous student solutions and canonical explanations in 10th grade whole classroom instruction on a statistical concept. These teacher-led activities either followed or preceded student-led practice activities. Whereas contrasting-and-comparing-based instruction improved conceptual learning outcomes compared to expository instruction, timing did not have an additional effect.

Due to the limited research available, our expectations regarding timing effects (prior compared to after instruction) are then not very strong, but nevertheless lean toward a positive effect. In the present study, we test the hypothesis that giving CEC&C feedback prior to instruction improves learning gains on conceptual change tasks, compared to a condition in which CEC&C feedback is given after the instructional phase (H3).

The present study

These hypotheses will be tested in a controlled experimental study on university students' understanding of natural selection. Even without formal education on the topic, most people have intuitive theories about natural selection, which are usually based on explanatory schemata that are incommensurate with the scientifically accepted account. Previous research has shown that instruction that only presents the correct account is insufficient to induce a lasting change in students' understanding of natural selection (e.g., Asterhan & Schwarz, 2007; Jensen & Finley, 1996; Jimenez-Aleixandre, 1992). Students are given CEC&C feedback on their self-generated solutions to test items on natural selection. This feedback is either given prior to or after the expository instruction phase presenting the correct scientifically accepted account of natural selection. Their conceptual understanding is tested on a delayed posttest and compared to students who compared their own solution with an expert one, but did not receive CCE&C feedback at any stage in the experiment. Natural selection is a complex concept whose correct understanding requires the integration of a number of different principles. Therefore, we assess student understanding on two different levels: the schemata students use to explain evolutionary phenomena and the number of correct principles they produce in their explanations.

Method

Participants

Sixty undergraduates from the Social Science and Humanities departments at the Hebrew University of Jerusalem, Israel, participated in this study (mean age= 24.86 *yrs*). One participant failed to complete the delayed posttest and the relevant data was therefore omitted from the relevant analyses. Participants were recruited through the university's computerized system of experiment registration and through publications on student social network sites. Hebrew proficiency was a requirement for participation, as was a lack of formal education in the Life Sciences Departments. Each participant was offered the choice between course credit (25%) or a financial reward of approximately $8 (75%) for their participation.

Design

A 1X3 experimental design with random assignment to condition was used. Individual conceptual understanding of natural selection was assessed on three separate test occasions (pretest, immediate post-test and delayed post-test). The experimental conditions differed in whether they received CEC&C feedback on their individual solutions, or only received the correct answers only. The CEC&C feedback condition was furthermore subdivided in receiving feedback prior to or following the expository instruction phase (see Figure 1).

Materials

Demographic information questionnaire. Participants completed a questionnaire regarding the following demographic details: gender, age, major, academic year, previous knowledge in the field of biology and evolution, and religious affiliation.

Instructional movie. A 20 *min* excerpt of an educational movie in the "New Frontiers" series, which presents Darwin's theory of the evolution of species, with Hebrew subtitles (see Procedure). The movie excerpt presented the evolutionary development of different species on the Galapagos Islands (iguanas, turtles, sea-lions), without an account of how that change had occurred. In addition, changes in a population of Galapagos finches (also known as Darwin's finches) were discussed in detail and explained in terms of natural selection. This explanation also included a step-by-step graphical presentation. Natural selection was presented as the scientifically proven explanation for the evolution of species.

Test items. Based on previous work (Asterhan & Schwarz, 2007; 2009), eight open test items were compiled to assess conceptual understanding of natural selection. All items had structural similarity but different surface features. They included information about ways in which a certain species' ancestors used to be different on a given trait. Subjects were then requested to explain how evolutionary theory would account for how this process of change occurred. Following is an example of a test item:

> "*Ducks have webbed feet. Thousands of years ago, the ancestors of the current ducks lived mostly in dry lands and their feet were similar to those of current pigeons or chickens. It is also known that as a result of global warming and consequent sharp increases in the amounts of rain, the living areas of these proto-ducks became mostly flooded. Given this information, please explain how evolutionary theory would account for how the change that occurred in the duck's feet (to their current* shape of webbed feet)?"

Four types of trait changes were included (2 items for each type): (1) A change in an observable physical trait (i.e., longer necks in giraffes, webbed feet in ducks); (2) an improvement of an existing ability (i.e., swimming in sea iguana, running speed of cheetahs); (3) the loss of an ability despite the physical existence of the "relevant" organ (i.e., ability to fly among emus and the loss of sight among cave salamanders); and (4) a change in color (i.e., polar bear fur and the coloring of the pepper moth). The pretest included one item of each type (4 altogether). The immediate posttest included two items (one from type 1 and one from type 2), as did the delayed posttest (i.e., one from type 3 and one from type 4). Each participant answered each of the eight items only once throughout the experiment. CEC&C feedback was always given on the responses to the duck and cheetah items.

Procedure

All participants received a short explanation about the study and signed a consent form. They were then led to a separate room where they completed a demographic survey and the pretest. Upon completion, students in the pre-instruction CEC&C feedback condition received corrective feedback on their solutions of items 1 and 2 of the pretest (i.e., the duck and the cheetah items, see Figure 1). They received a maximum of 10 *min* to review the feedback. Students in the other two conditions did not receive corrective feedback on their pretest solutions. Pretest items 1 and 2 in the feedback-after-instruction condition referred to different phenomena of the same item category (i.e., the iguana and the giraffe items, see Figure 1). For participants in the control condition, this order was counterbalanced.

Following, each participant watched a 20 *min* excerpt of an educational movie on natural selection (see Materials section) and then completed the immediate post-test questionnaire which consisted of two transfer items: The iguana and giraffe items in the feedback-prior-instruction condition, and the duck and cheetah items in the feedback-after-instruction condition. Upon completion, participants in the feedback-prior condition were excused. Participants in the feedback-after condition received CEC&C feedback on their solutions of the duck and the cheetah items, and were given time to study the feedback. Students in the control condition were given an equal amount of time to study the correct answer sheet and compare it with their own solutions, but did not receive any feedback.

All participants completed the delayed posttest at least one week following the first part of the experiment (ranging from 7 to 11 days, M = 7.76, SD = 1.21). The delayed posttest was composed of two test items (one from type 3 and one from type 4) and referred to two novel phenomena participants had not encountered yet: Half of the participants received questions on the wings of emus and the color of the peppered moth, whereas the other half received questions on the eyesight of cave salamanders and the color of polar bear fur. This was counterbalanced with the content of items 3 and 4 of the pretest (see Figure 1).

Correction procedure. Regardless of condition, the correction of participants' solutions was always performed on the duck and cheetah items. After the subject completed the questionnaire, the experimenter took the answer sheets for review, while the participant waited in the room. CEC&C feedback consisted of the following:

(1) Corrective notes on the participant's answer sheet. Incorrect propositions were highlighted with colored markers and a highlighted section was connected to a written note that appeared under the subject's solution. These notes always included an explanation of the error, as well as a correction. For example, a sentence such as "*The duck developed webs*" would be highlighted and the connecting feedback note would read: "*The individual duck is unable to change genetically determined traits on its own*

("to develop" webs) and to pass these changes on to its offspring. Natural selection is based on initial, existing variance between individual members of a population and on differential selection and reproduction of those members that by chance already possess the advantageous trait". When statements were unelaborated, incomplete or vague (but not necessarily wrong) they were marked by an asterisk and were clarified with a corrective note as well.

(2) A pre-prepared sheet with the complete, correct solutions for both test items was given to each participant. The experimenter highlighted sentences that were specifically relevant to the errors made by a specific participant. In the control condition, participants were given the correct answer sheet only, without any corrective feedback or highlights.

Coding

Conceptual understanding of evolutionary change processes was assessed according to a coding procedure developed and validated by Asterhan & Schwarz (2007), which comprises two separate but complementary scoring schemes assessing (1) the type of conceptual models student use, and (2) the number of correct principles they incorporate in their explanations. Interrater reliability on these schemes was calculated on a separate data set with identical test items and was good, Cohen's κ = .72 on the conceptual model score and r = .95 on the Darwinian principles score.

(1) The first assessed the correctness of the explanatory schema (Ohlsson, 2002) a respondent uses to explain changes in a species' traits. In a slight adaptation of the original coding scheme (Asterhan & Schwarz, 2007), we distinguished between three explanatory schemas: typological, natural selection and hybrid models. Solutions that are based on natural selection schemas include references to differential selection and existing intra-species variation was considered (grade: 1). Following is an example

from the data set: "*Following particular changes to the giraffes' ancestors' living conditions, there was a shortage of food in the lower branches of trees and shrubs, and therefore the giraffes with longer necks (among all the different giraffes that lived in a particular time) survived better because they managed to eat more and therefore mated more than giraffes with shorter necks and thus passed their genetic traits to the next generations*" (subject 1223).

Typological explanations, on the other hand, state that evolutionary change is due to (small) changes that occur to all members in a given generation of the population (Shtulman, 2006) in reaction to a change in the environment which "requires" the change. Each member in the next generation becomes better adapted to the environment, until the full development of the necessary trait. Explanations based on a typological schema can contain different references to *how* this change occurs exactly (e.g., acquired trait changes as an outcome of effort, spontaneous mutations in genetic make-up in reaction to a specific need, awakening of "dormant genes" (Asterhan & Schwarz, 2007). These all have in common that they do not contain references to some form of selection mechanism, nor to intra-species variance (grade: 0). Following is an example from the data set: "*When the cheetah did not need to hunt prey or when other predators were slower, their maximum speed was lower (32 km). With the passing of time, when the other predators began to run faster or when the cheetah's prey started to develop [improve] their running ability, the cheetah also made sure its running abilities became faster, and the fastest in the hunting territories*" (subject 4859). A zero grade was also given to responses that simply did not answer the question, for example, when students stated that they did not know the answer or just repeated the data provided, without providing additional information.

Half a point was given to solutions that reflect a hybrid or mixed model of evolution. Hybrid models include explanations with references to either selection or intra-species variance, in combination with some characteristics of typological models (Asterhan & Schwarz, 2007). For example, some explanations claiming that only those individuals that somehow managed to change themselves (i.e., a directed, intentional change) in their lifetime survived, and then passed these newly acquired changes to their offspring. Half a point was also given to responses in which students presented two alternative solutions to an item (one typological and one Darwinian).

Based on these explanatory schema scores, a dichotomous variable for conceptual change was compiled, separately for gains from pretest (items 1 and 2 only) to immediate posttest and for gains from pretest (items 3 and 4 and 4b only) to delayed posttest. Evidence for a substantive change in the types of conceptual models that were used in students' explanations (e.g., conceptual change) was defined as: (1) an improvement on the mean explanatory schema score from an imperfect (<. 67) to a perfect or near perfect average score (.83-1.00); or (2) a mean increase of at least .5 points on the mean explanatory schema score. A mean increase of .5 indicates that either the student has improved from *consistently* using typological to hybrid, or from hybrid to Darwinian explanatory models. It may also indicate, however, that the student has moved from a typological to a Darwinian account on at least one of the two test items. Each of these possibilities was included in the dichotomous variable of substantive conceptual improvement. Students with (near) perfect pretest scores were not regarded, because, by definition, such individuals could not show substantive improvement on this measure.

(2) The complementary coding scheme assesses the explicit use of the following six principles of natural selection in a test item response: Intra-species variability,

source of intra-species variability (i.e., random changes in genetic material), differential survival rates, differential reproduction rates, accumulation of changes (i.e., the process is repeated in each generation), changes within the population (i.e., the proportion of individuals carrying the advantageous trait(s) will increase within the population). The appearance of each principle in each test item was assessed according to the following grading key: The principle is mentioned explicitly and applied correctly (2 points), only a part of the principle is mentioned and used correctly (1 point), or the principle is not mentioned or applied incorrectly (0 points). Further details about this coding scheme can be found in Asterhan and Schwarz (2007). The total grade for each item ranges between 0 and 12, and between 0-24 on a test occasion (based on 2 items in each).

Whereas the explanatory schema score provides an indication of the extent to which a participant bases his or her explanations on the correct schema structures, the principles score provides an indication of the fullness of that explanation, i.e., whether the individual elaborates and explicitly refers to the different principles of natural selection. Thus, for example, an explanation may be correct but include explicit references to two of the principles (intra-species variability and differential survival patterns), as is exemplified in the aforementioned example. Vice versa, a typological explanation may include partial references to certain principles (e.g., the need to survive without reference to differential survival patterns or a generic reference to accumulation over many generations, each only receive 1 out of two points in the principle score coding scheme).

Results

For each hypothesis, two separate analyses were conducted, one for each of the two conceptual understanding scores: Analyses for the natural selection principles scores were conducted with analyses of variance with repeated measures for each of the three hypotheses. Mean principles scores are presented in Table 1. Checks for variance as well as covariance homogeneity were conducted with Levene's and Box's tests respectively, and were non-significant in all cases.

Analyses on the explanatory schema score were conducted with non-parametric tests (Chi square). The number of students who achieved substantive improvement on the explanatory schema score per condition are presented in Table 2. This table also shows the types of substantive change that were observed: From mainly typological explanations (mean grade 0 to .17) to hybrid or mixed explanations (mean grade between .33 to.67), from mainly typological explanations to mainly natural selection-based explanations (mean grade .83 to 1.00), or from hybrid or mixed explanations to natural selection-based explanations.

No differences between conditions were found on the combined pretest principles scores, F (2, 58) = 1.17, p = .319, nor on the demographic variables. An independent sample t-test comparing the pretest(1+2) and the pretest(3+4) principles scores revealed that students performed better on the former (M = 5.51, SD = 4.84) than on the latter (M = 3.90, SD = 4.33), t(58) = 3.27, p = .002. As the specific items were counterbalanced within each of the four item types, this means that the item types that were used for the delayed posttest (item types 3 and 4) were slightly more difficult than the ones used for the immediate posttest (item types 1 and 2). Shapiro-Wilk tests showed that gains from pretest to immediate and to delayed posttest distributed normally, p = .795 and p = .305, respectively.

Manipulation check

The corrections in both conditions were given to the same test items (the duck and the cheetah items). We counted the number of times a participant in the feedback-prior or the feedback-after condition received a correction that explicitly referred to one of the six natural selection principles. An independent sample t-test comparing the overall number of corrections given showed no differences between the feedback-prior (M = 5.53, SD = 1.90) and the feedback-after condition (M = 4.90, SD = 1.76), $t(36) = 1.06$, $p = .294$. We compared the number of corrections for each of the six principles (either zero, once or twice) with separate Chi square tests. No differences were found, $.292 < p < .497$

The effect of receiving CEC&C feedback on immediate posttest scores

To test the hypothesis that receiving CEC&C feedback leads to larger learning gains on immediate posttest scores, student gains from the pretest to the immediate posttest were compared between two conditions: At the time of the immediate posttest, only students in the pre-instruction CEC&C feedback condition had received CEC&C feedback prior to watching the instructional movie, and their performance was therefore compared to that in the other two conditions combined. Mean pretest scores (items 1+2) were not found to differ across conditions, $t < 1$.This also implies that, even though students in the feedback-prior and the feedback-after condition received different items for the first pretest part (duck + cheetah and giraffe + iguana, respectively), no differences in item difficulty were found.

A 2 (time) X 2 (condition 1 *vs* 2 + 3 combined) analysis of variance with repeated measures was conducted on the sum of the natural selection principles score of the pretest (items 1 and 2) and the immediate posttest score. An interaction effect for time with condition was found, $F(1,57) = 16.45$, $p < .001$, with a large effect size, $\eta_p^2 = .22$: Students who had received CEC&C feedback used a larger number of natural selection principles in their solutions at the immediate posttest (gains from *M* = 4.63 to *M* = 11.58), compared to students who had not (yet) received CEC&C feedback (increase from *M* = 5.63 to *M* = 7.68).

A Chi square test was conducted to examine whether students who received CEC&C feedback achieved conceptual change more often than those who had not received CEC&C feedback by the time of the immediate posttest. Nine individual attained a full or near full (>.83 out of 1) explanatory schema score on the pre-test (3 in the feedback-prior, 4 in the feedback-after and 2 in the control condition), and were therefore excluded from the analysis. Fifteen out of 17 participants (88%) in the CEC&C feedback condition showed substantive improvement from the pretest to the immediate posttest, compared to 10 out of 34 students (29%) in the no feedback condition, $\chi^2_{(1,N=51)} = 15.69, p < .001$), with a large effect size, Cramer's V = .55

The effect of receiving CEC&C feedback on delayed posttest scores

To test the hypothesis that receiving CEC&C feedback, in and by itself and irrespective of its timing, leads to larger learning gains on delayed posttest scores, student gains from the pretest (items 3 and 4) to the delayed posttest scores were compared between two conditions: Those who had received CEC&C feedback (either prior or after instruction) and those who had received the correct answer sheet only. Mean pretest scores (item types 3+4, counterbalanced) were not found to differ between conditions, *t*(57) = 1.08, *p* = .285.

A 2 (time) X 2 (conditions 1+2 *vs* 3) analysis of variance with repeated measures was conducted on the sum of the natural selection principles scores of the pretest (items 3 and 4) and the delayed posttest. Students who had received CEC&C feedback increased their natural selection principles score from *M* = 4.33 to *M* = 11.28, compared to students who had not received CEC&C feedback who only increased their score from $M = 3.05$ to $M = 7.86$. The interaction effect of time with condition was only marginally significant, however, $F(1,57) = 3.48$, $p = .067$, with a medium effect size, $\eta_p^2 = .06$.

A Chi square test was conducted to examine whether more students who had received CEC&C feedback at any time during the experiment substantively changed their explanations of natural selection (i.e., conceptual change) compared to those who only had received the correct answer sheet. Again, students with a perfect or near perfect score on pretest were excluded from this analysis. In the CEC&C feedback condition, 22 out of 32 participants (69%) showed substantive improvement on the mean explanatory schemas score, compared to only 5 out of 18 students (28%) in the no feedback condition, $\chi^2_{(1,N=50)} = 7.79, p = .005$, with a medium effect size, Cramer's V = .38.

The effect of timing of CEC&C feedback on learning

In order to test the hypothesis that receiving CEC&C feedback prior to the instruction phase leads to larger learning gains, compared to receiving it after instruction, student performance on two pretest items 3 and 4 was compared with their delayed posttest scores. Pretest (3+4) principle scores were not found to differ between conditions, $t < 1$.

A 2 (time: pretest to delayed posttest) X 2 (condition: CEC&C feedback prior to or after instruction) analysis of variance with repeated measures was conducted. A main effect of time was found, $F(1, 37) = 121.41$, $p < .001$, $\eta_p^2 = .77$. Differences in timing of CEC&C feedback did not affect learning, however: Receiving CEC&C feedback prior to instruction resulted in roughly equal gains (from $M = 3.68$ to $M = 11.16$) compared to receiving feedback after instruction (from $M = 4.95$ to $M = 11.40$), $F(1,37) < 1$.

A Chi square test was conducted to examine whether timing of CEC&C feedback had an effect on the conceptual change measure (i.e., substantive gains on the conceptual model level), including only those students who, based on their pretest scores, could substantively improve. When feedback was given prior to instruction, 75% (N = 12) of students showed substantive improvement in the types of explanatory models they used to explain evolutionary change, compared to 63% (N = 10) when this feedback was given after the instruction. This difference was not significant, $\chi^2_{(1,N=32)} = .58, p = .45$.

Discussion

In the present work, we examined whether feedback that corrects and contrasts a student's own erroneous solutions with the canonical, correct one (CEC&C feedback) yields improved learning gains in an expository instruction-based conceptual change task. This was compared to two situations: (1) receiving no feedback and (2) autonomous comparison of own and canonical solutions without further feedback or support. It was also tested whether the timing of CEC&C feedback impacts learning gains. We discuss our main findings, the study's limitations and venues for future research in relation to three different literatures: Conceptual change, contrast-and-comparison methods of instruction, and feedback effects.

CEC&C feedback and conceptual change

Our findings show, first and foremost, that standard tell-and-practice instruction for conceptual change can be substantively improved by giving students detailed CEC&C feedback on their self-generated solutions. When students were given the canonical solution and asked to read and compare it on their own, only a quarter of them achieved conceptual change, but this increased to two-third when they had received CEC&C feedback, either before or after the instruction phase.

Given the fact that standard tell-and-practice instruction has recurrently been shown ineffective for conceptual change of robust misconceptions (e.g., Newtonian mechanics, evolution and diffusion), research has intensively focused on alternative, resource-intensive methods of instruction. These involve, among others, hypothesis testing, peer group work, and project-based activities. Our findings show, however, that standard tell-and-practice methods can be upgraded significantly with a less resource-intensive method that has often been overlooked in the conceptual change literature: detailed and precise feedback that contrasts and compares between erroneous and correct explanations. Even though CEC&C feedback did not produce perfect scores, the effect size and the number of participants who did reach conceptual change compare very favorably to other studies who rely on more resource-intensive learning activities on similar concepts (e.g., Asterhan & Schwarz, 2007; 2009; 2014; Chi et al, 2012).

Taken together, findings from the present and the abovementioned recent studies highlight the importance of explicit comparison and contrasting of student-generated, erroneous solutions and canonical, scientific accounts for conceptual change learning. Students often do not notice, gloss over, refute or dismiss differences (Chinn & Brewer, 1998). The expectation that students will detect and understand the difference in the deep structure of the conceptual models underlying two different solutions is not realistic, not even for most university students. The benefits of CEC&C feedback are expected to be even larger for school-aged students, who are likely to meet with more difficulty when required to compare and contrast erroneous and correct explanations on their own. Future research should further test the generalizability of the current findings to other content domains, and different misconceptions.

Unfortunately, the current study did not include process data collection, which could have provided some insight into the cognitive and motivational processes during the feedback and the instruction phase. This focus on outcome measures and paucity of process measures is common in the literature on instruction for conceptual change. Models of conceptual change processes, on the other hand, are often based on cross-population comparisons, such as experts *vs.* laypersons (e.g., Babai et al., 2010; Dunbar et al., 2007; Masson et al., 2014; Shtulman & Valcarcel, 2012) or children of different age groups (e.g., Vosniadou & Brewer, 1994). Only few have attempted to describe the processes of conceptual change *while individuals receive instruction* (e.g., Opfer & Siegler, 2004; Chinn & Brewer, 1998). Future research on the effects of feedback for conceptual change should preferably include both outcome measures, as well as process data.

Contrasting and comparing erroneous and correct solutions

The findings presented here also add to a growing evidence base on the effects of instructional methods that include some form of activities in which erroneous and correct solutions are contrasted and compared (Asterhan & Schwarz, 2016; Diakidoy et al., 2003; Durkin & Rittle-Johnson, 2012; Gadgil et al., 2012; Loibl & Rummel, 2014; Kapur & Bielaczyc, 2011; Sinatra & Broughton, 2011). In these studies, the contrasts and comparisons were either explicitly given to students (e.g., as in refutation texts and in classroom teaching), or students were asked to compare and contrast pre-selected erroneous and correct solutions side-by-side while receiving additional support and guidance (e.g., Durkin & Rittle-Johnson, 2012; Gadgil et al., 2012; Grosse & Renkl, 2007). Both formats produced better learning gains than receiving expository instruction only and/or than comparing the correct and erroneous solutions without any support.

The present study adds to this line of research by extending the CEC approach to an additional instructional method, namely written feedback. Moreover, the contrasting and comparing is done with the student's own erroneous solutions, instead of with a general, yet common misconception. The abovementioned studies, as well as the current one, use different assessment methods, age groups and content domains. Future research needs to include direct comparisons of CEC&C feedback with other methods, to determine which of the different CEC activities is more effective. For example, refutation texts provide an explicit statement of commonly held misconceptions, directly refutes them, and then introduces scientific explanations as alternatives (Sinatra & Broughton, 2011; Tippett, 2010; Vosniadou & Mason, 2012). Since CEC&C feedback is personally tailored to individual errors, one may expect it to be more effective than refutation texts. Moreover, the strongest effects for refutation text are typically reported on less complex topics, such as whether ostriches bury their heads in the sand or not (Tippett, 2010), with very few studies on complex, robust misconceptions (e.g., Diakidoy, Kendeou, & Ioannides, 2003). Finally, recent research suggests that refutation texts primarily aid in the recognition of errors, but are less successful in improving understanding of the correct account (Diakidoy, Mouskounti, Fella, & Ioannides, 2016; Van Loon et al., 2015).

Providing CEC&C feedback that is tailored to every individual student can prove to be time-consuming for teachers, however. Another practical drawback of CEC&C feedback is that it heavily relies on teacher expertise to assess and detect student misconceptions. Research has revealed that there are considerable differences in teachers' individual ability to identify and correct common student misconceptions in a content domain (Hill, Rowan & Ball, 2005; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). Future research should then compare the effectiveness of CEC&C

feedback with pre-prepared material that contrast common misconceptions with correct ones, such as refutation text, both in controlled as well as classroom settings. If differences prove to be small, then practical and logistic considerations in classroom settings may tip the scales towards the latter. CEC&C feedback may also be considered for implementation in intelligent tutoring systems that target conceptual change types of learning, provided that the linguistic components analyzing student explanations will be sophisticated and sensitive enough to pick up on the misconceptions. For example, keyword use may be misleading as students often use phrases such as “develop”, “adaptive” and “mutations” but in an erroneous way. Finally, research should also explore how different instructional methods can be combined to maximize learning gains in conceptual change tasks. For example, combining CEC&C feedback with refutation texts, instead of expository texts, may further improve success rates.

The finding that timing of CEC&C feedback did not have an effect on learning outcomes in this study is, on the one hand, surprising. Nevertheless, it is in accordance with findings from a recent study by Loibl and Rummel (2014), in which teachers contrasted key features of common student erroneous solutions with the canonical one improved student learning of statistical concepts. Teacher-led CEC&C classroom instruction was either followed or preceded by student group work on open-ended tasks. Both the present and the Loibl and Rummel study found that not timing, but the actual presence of CEC&C activities improved learning.

Given the scarcity of research on timing effects of CEC&C activities and the differences in learning activities and settings between the two available studies, we believe it to be too early to draw definite conclusions concerning timing effects and argue for more empirical research. It is possible that the advantages of feedback timing effects will become evident in a larger sample, in other content domains, or with different populations (in particular, with younger students). Moreover, collecting process data during the presentation of feedback and the instruction phase in future research could provide further insight as to the reasons behind the current results, as well as solutions to improve the effectives of one timing approach over the other.

Feedback effects and conceptual change

Finally, the findings presented here also contribute to the feedback literature by focusing on conceptual change, a learning goal and learning outcome that has not been considered in the feedback literature. The design of the feedback in this study was based on current theories of conceptual change processes (e.g., Ohlsson, 2002; Potvin et al., 2015; Schnotz & Preuss, 1999) and, as such, aimed to facilitate both inhibition of irrelevant schema structures and improved retrieval of the relevant ones. CEC&C feedback compared favorably to providing the correct account only. However, the present study did not include a comparison condition in which student were only made aware of their mistakes (i.e., outcome feedback). Based on this model as well as on previous findings (Asterhan et al., 2014), we expect it unlikely that outcome feedback alone will be sufficient for learning that requires conceptual change.

Finally, the present study was conducted with college students and in laboratory settings. It is possible that due to impression management, learners invested more cognitive effort in studying the feedback, compared to what they would have in everyday classroom settings. Since this argument also holds for reading the correct solutions in the control condition, however, we have no reason to believe that it has affected the pattern of difference that was found between conditions.

In conclusion.

Whereas research has repeatedly shown the ineffectiveness of standard tell-and-practice instruction for conceptual change types of learning, the present study reveals that it can be improved substantively by giving students detailed corrective feedback that explicitly contrasts the correct, full explanation with the erroneous, student-generated one. Whereas these first results are promising, future research should seek both to replicate these results with a larger sample size, as well as to extend it to additional topic domains that are traditionally considered in the conceptual change literature (e.g., diffusion, forces, genetics). Given the well-documented success of elaborate feedback in other fields of learning and given the complexity and resource-intensity of many existing instructional approaches for conceptual change, we recommend to broaden and deepen research on the role of feedback for conceptual change.

Acknowledgements

This research was funded by Israeli Science Foundation award 1044/13. We thank Maya Resnick, Roni Segal, Noa Ettinger and Morag Pitaro for their assistance in data collection and coding efforts.

References

Ames, G. J., & Murray, F. B. (1982). When two wrongs make a right: Promoting cognitive development through cognitive conflict. *Developmental Psychology, 18*(6), 894-897. doi:10.1037/0012-1649.18.6.894.

Asterhan, C. S. C. & Schwarz, B. B. (2016). Argumentation for learning: Well-trodden paths and unexplored territories. *Educational Psychologist, 51*(2), 164-187.

Asterhan, C. S. C., & Schwarz, B. B. (2009). The role of argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialogue. *Cognitive Science, 33*, 373-399.

Asterhan, C. S. C., & Schwarz, B. B. (2007). The effects of monological and dialogical argumentation on concept learning in evolutionary theory. *Journal of Educational Psychology, 99,* 626-639.

Asterhan, C. S. C., Schwarz, B. B., & Cohen-Eliyahu, N. (2014). Outcome feedback during collaborative learning: Contingencies between feedback and dyad composition. *Learning and Instruction, 34* (4), 1-10.

Babai, R., Sekal, R., & Stavy, R. (2010). Persistence of the intuitive conception of living things in adolescence. *Journal of Science Education and Technology*, *19*(1), 20-26. doi:10.1007/s10956-009-9174-2

Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273-281. doi:10.1037/e527352012-769

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*, *15*(1), 1-40. doi:10.1207/s1532690xci1501_1

Chi, M. T. H. (2008). Three types of conceptual change: Belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61-82). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. doi:10.1007/978-1-4615-0605-8_20

Chi, M. T. H., Roscoe, R., Slotta, J., Roy, M., & Chase, M. (2012). Misconceived causal explanations for "emergent" processes. *Cognitive Science*, *36*, 1-61. doi:10.1111/j.1551-6709.2011.01207.x

Chinn, C. A., & Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*(6), 623-654.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*(1), 1-49. doi:10.2307/1170558

Diakidoy, I. A. N., Mouskounti, T., Fella, A., & Ioannides, C. (2016). Comprehension processes and outcomes with refutation and expository texts and their contribution to learning. *Learning and Instruction*, *41*, 60-69. doi: 10.1016/j.learninstruc.2015.10.002

Diakidoy, I.A.N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: the effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology*, *28*, 335–356. doi: /10.1016/S0361-476X(02)00039-5

Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centration in cognitive development. *European Journal of Social Psychology*, *9*(1), 245–247. doi:10.1002/ejsp.2420090110.

Dunbar, K., Fugelsang, J., & Stein, C. (2007). Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. *Thinking with data*, 193-206. doi: 10.1108/00251740810865085

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction*, *22*(3), 206-214. doi: 10.1016/j.learninstruc.2011.11.001

Gadgil, S., Nokes-Malach, T. J., & Chi, M. T. (2012). Effectiveness of holistic mental model confrontation in driving conceptual change. *Learning and Instruction*, *22*(1), 47-61. doi: 10.1016/j.learninstruc.2011.06.002

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: can this foster learning outcomes? *Learning and Instruction*, *17*, 612-634. doi:10.1016/j.learninstruc.2007.09.008

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112. doi:10.3102/003465430298487

Hill, H., Rowan, B., and Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Education Research Journal*, *42*(2), 371-406. doi:10.3102/00028312042002371

Howe, C., Tolmie, A., Duchak-Tanner, V., & Rattay, C. (2000). Hypothesis-testing in science: Group consensus and the acquisition of conceptual and procedural knowledge. *Learning and Instruction*, *10*(4), 361-391. doi:10.1016/S0959-4752(00)00004-9.

Jensen, M. S., & Finley, F. N. (1996). Changes in students' understanding of evolution resulting from different curricular and instructional strategies. *Journal of Research in Science Teaching*, *33*(8), 879-900. doi:10.1002/(sici)1098-2736(199610)33:8<879::aid-tea4>3.0.co;2-t

Jiménez-Aleixandre, M. P. (1992). Thinking about theories or thinking with theories?: a classroom study with natural selection. *International Journal of Science Education*, *14*(1), 51-61. doi:10.1080/0950069920140106

Kapur, M., & Bielaczyc, K. (2011). Classroom-based experiments in productive failure. In L. Carlson, C. Holscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 2812–2817). Austin, TX: Cognitive Science Society. doi: 10.1109/iecon.2007.4459876

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254- 284. doi:10.1037/0033-2909.119.2.254

Kulik, J. A., & Kulik, C. L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, *58*(1), 79-97. doi:10.2307/1170349

Light, P., & Glachan, M. (1985). Facilitation of individual problem solving through peer interaction. *Educational Psychology, 5*, 217-225. doi: 10.1080/01443418500503.

Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, *11*, 357–380. doi: 10.1016/s0959-4752(00)00037-2

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction* 34, 74-85. doi: 10.1016/j.learninstruc.2014.08.004

Masson, S., Potvin, P., Riopel, M., & Foisy, L. M. B. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education*, *8*(1), 44-55. doi: 10.1111/mbe.12043

Ohlsson, S. (2002). Generating and understanding qualitative explanations. In J. Otero, J. A. Leon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 91–128). Mahwah, NJ: Erlbaum.

Opfer, J. E., & Siegler, R. S. (2004). Revisiting preschoolers' living things concept: A microgenetic analysis of conceptual change in basic biology. *Cognitive Psychology, 49*, 301-332. doi: 10.1016/j.cogpsych.2004.01.002

Özdemir, G., & Clark, D. (2007). An overview of conceptual change theories. *Eurasia Journal of Mathematics, Science & Technology Education, 3*, 351-361.

Potvin, P., Masson, S., Lafortune, S., & Cyr, G. (2015). Persistence of the intuitive conception that heavier objects sink more: A reaction time study with different levels of interference. *International Journal of Science and Mathematics Education*, *13*(1), 21-43. doi: 10.1007/s10763-014-9520-6

Potvin, P., Sauriol, É., & Riopel, M. (2015). Experimental evidence of the superiority of the prevalence model of conceptual change over the classical models and traditional teaching. *Journal of Research in Science Teaching*, *52*(8), 1082-1108. doi: 10.1002/tea.21235

Ramsburg, J. T., & Ohlsson, S. (2016). Category change in the absence of cognitive conflict. *Journal of Educational Psychology*, *108*(1), 98. doi: 10.1037/edu0000050

Richland, L. E., Kornell, N., & Kao, L. S. (2009). The pretesting effect: Do unsuccessful retrieval attempts enhance learning? *Journal of Experimental Psychology: Applied*, *15*(3), 243. doi: 10.1037/a0016496

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, *50*(5), 1020-1049. doi: 10.3102/0002831213477680

Schnotz, W., & Preuss, A. (1999). Task-dependent construction of mental models as a basis for conceptual change. In W. Schnotz, S. Vosniadou, & M. Carretero (Eds.), *New perspectives on conceptual change* (pp. 193–222). Amsterdam: Pergamon Press.

Schwarz, B. B., & Linchevski, L. (2007). The role of task design and argumentation in cognitive development during peer interaction: The case of proportional reasoning. *Learning and Instruction*, *17*(5), 510-531. doi: 10.1016/j.learninstruc.2007.09.009

Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right... If they argue together! *Cognition and Instruction*, *18*(4), 461-494. doi: 10.1207/s1532690xci1804_2

Shtulman, A. (2006). Qualitative differences between naïve and scientific theories of evolution, *Cognitive Psychology*, *52*, 170-194. doi: 10.1016/j.cogpsych.2005.10.001

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition*, *124*(2), 209-215. doi: 10.1016/j.cognition.2012.04.005

Sinatra, G. M., & Broughton, S. H. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, *46*(4), 374-393.

Van Loon, M. H., Dunlosky, J., Van Gog, T., Van Merriënboer, J. J., & De Bruin, A. B. (2015). Refutations in science texts lead to hypercorrection of misconceptions held with high confidence. *Contemporary Educational Psychology, 42*, 39-48. Fdoi:10.1016/j.cedpsych.2015.04.003

Vosniadou, S. (Ed.). (2009). *International handbook of research on conceptual change*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. doi: 10.4324/9780203154472.ch1

Vosniadou, S., & Mason, L. (2013). Conceptual change induced by instruction: A complex interplay of multiple factors. In S. Graham, J. Royer, & M. Zeidner (Eds.), *Individual differences and cultural and contextual factors*, Vol 2 of the APA Educational Psychology Handbook Series (pp. 221-246). APA Publications. doi: 10.1037/13274-000

Table 1.

*Mean principle scores (and SD) at pretest, immediate and delayed posttest, per condition**

	CEC&C feedback		
	Prior to instruction (N = 19)	*Following instruction* (N = 20)	*No CEC&C feedback* (N = 20)
Pretest (1+2)	4.63 (4.55)	6.60 (5.58)	4.65 (4.27)
Pretest (3+4)	3.68 (4.46)	4.95 (4.83)	3.05 (3.61)
Immediate posttest	11.58 (4.81)	8.65 (4.59)	6.70 (5.14)
Delayed posttest	11.16 (4.35)	11.40 (4.15)	7.85 (5.19)

* Maximum score per test occasion is 24

Table 2.

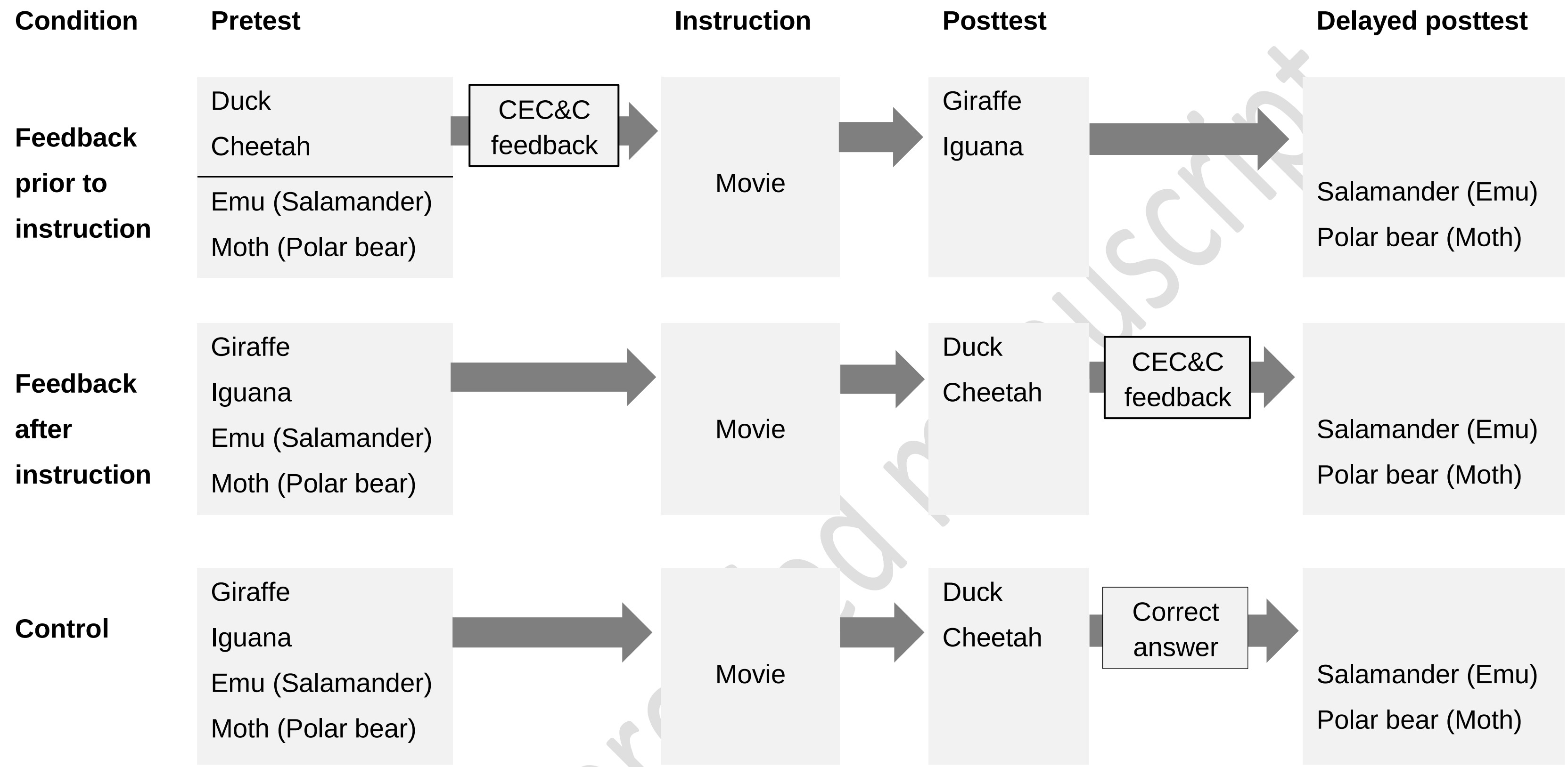
Number of students who showed substantive gains on mean conceptual model scores from pretest to delayed posttest, per condition and type of gain (N = 50)*

	CEC&C feedback		
Type of substantive conceptual model gain	*Prior to instruction*	*Following instruction*	*Control*
None	5	6	12
From level 1 (typological) → 2 (hybrid)	3	3	3
From level 2 (hybrid) → 3 (natural selection)	2	3	4
From level 1 (typological) → 3 (natural selection)	7	4	1

* Only includes students with mean pretest scores under 83%

Figure 1

Overview of the different stages of the experiment, according to condition*



* Items in parentheses refer to counterbalancing of items across test occasions