**Learning science concepts by teaching peers in a cooperative environment: a longitudinal study of pre-service teachers**

**Abstract**

Misconceptions of science principles are frequent among university students. This is particularly problematic in teacher education since these misconceptions can be passed on by teachers to pupils at school. In this study, 40 pre-service teachers studied 8 common misconceptions by three different approaches - Learning-by-Teaching (*LbT*) and Learning-from-Peers (*LfP*) (each organized in cooperative groups and using self-developed activities) and conventional lecturing. Participants’ understanding was evaluated prior to, 1-month after, and 2-years after instruction, together with their views about the effectiveness of these methods to achieve learning. The three approaches improved scores in all post-tests, but *LbT* achieved the best results. This coincided with perceptions of the participants who gave diverse reasons for the success of *LbT*. The longitudinal perspective showed for lecturingthat gains in student learning at 1-month had partially declined by 2-years, even though there was still a medium effect size compared to the pre-test. However, for *LbT* and *LfP* gains remained statistically stable throughout. A combination of these methods might not only contribute efficiently to conceptual change for a good number of misconceptions, but also connect teacher education to an almost-real school classroom, giving opportunities to translate student learning into teaching practice.

**Keywords**

*Learning-by-teaching, learning-from-peers, lecturing, misconceptions, teacher education.*

It is well known that new teachers often have misconceptions in their content knowledge which can be transmitted to pupils via their teaching (Butler, Mooney Simmie & O’Grady, 2014; Kind, 2014). These misconceptions can be very difficult to correct (Thorn, Bissinger, Thorn & Bogner, 2016). For this reason, it is particularly important in teacher training courses to raise students’ awareness of the existence of misconceptions, give them opportunities to explore and correct their own, and provide strategies to address them in their future classrooms.

This study investigated the effectiveness of three teaching methods for addressing the misconceptions of new science teachers: *Lecturing*, *Learning-from-Peers* (*LfP*) and *Learning-by-Teaching* (*LbT*). *Lecturing* is the most common mode of instruction in many countries (Biggs, 2005) and is recommended by some learning scientists (eg Kirschner, Sweller & Clark, 2006). *LbT* is often believed to be superior to other methods. It is highly engaging, and it is often said that someone does not really learn something until they have to teach it to somebody else. In the ICAP framework (see Chi & Wylie, 2014) constructive activity is required in order to develop explanations. *LfP* - by which students learn from other students - is now used more frequently, due perhaps to the increase in numbers of students and limited teaching resources (Burgess, McGregor & Mellis, 2014). *LfP* may be as passive as lecturing, though even that can lead to better results. Although the learning effects of the three methods are well known, few studies have examined them longitudinally.

Therefore, in this study we examined their effects in the short term and 2-years after the activities concluded, at which time we also asked the teachers to self-assess which teaching method had been most effective for them. This has given new insight into the impacts of each method and led to recommendations for teacher education.

**science teachers and conceptual change**

Although student and teacher misconceptions have been a focus in science education research since the 1970s (Duit & Treagust, 2003), there is no single universally accepted and tested theory of conceptual change (diSessa, 2014). Conceptual change entails the acquisition of new ideas and the revision of prior understanding (Vosniadou, 2013), but with emphasis on “change” rather than the simple accumulation of knowledge and building new concepts in the context of previous ones (diSessa, 2014).

Duit and Treagust (2003) and Duit, Treagust and Widodo (2013) traced the development of conceptual change theories in cognitive and constructivist traditions over the years; these mainly involved an interplay of Piagetian assimilation and accommodation, where new ideas always emerge from old ones (diSessa, 2014).

In the 1990s, models began to change from those using rational behaviours (“cold” models) to others taking into account motivational aspects (“hot” models). In this spirit, Pintrich, Marx and Boyle (1993) proposed linking cognition to motivational and affective factors: to address any misconception, teaching has to be organized keeping in mind the role of emotions, interest, motivational and affective frameworks (Broughton, Sinatra & Nussbaum, 2013; Duit et al., 2013; Duit & Treagust, 2003), and providing ideas, evidence, time and opportunities to practise and modify original constructs (Boesdosfer et al., 2012; Butler et al., 2014).

Gadgil, Nokes-Malach and Chi (2011) proposed “holistic confrontation” (p. 47) consisting of comparing and contrasting a student’s mental model, reviewing the interrelations between features of the prior knowledge and the scientific one. They proposed that, in making this comparison, students focussed on revising their knowledge of the false beliefs leading to robust knowledge.

Many studies addressing science misconceptions of pre-service teachers have been reported. For example, Trumper (2006) compared lecture instruction on moon phases and eclipses with a more constructivist approach showing that both strategies led to learning, but the constructivist approach was more effective. Similarly, Baser (2006) obtained better results using cognitive conflict in a traditional lecture. In a more complete consideration, Trundle, Atwood and Christopher (2007) pursued a longitudinal study of 12 students finding that only three had reverted to their original ideas 6-13 months after inquiry-based instruction: reversion to alternative conceptions was explained in terms of students’ low metacognitive awareness during instruction.

Bulunuz and Jarret (2009) evaluated learning using “hands-on” approaches, readings and conceptual mapping. Participants scored higher following the hands-on approaches. Boesdorfer et al. (2011) used vicarious learning through a video that showed a discussion among students. They evaluated understanding immediately afterwards and 10 weeks later, and found that half of the initial misconceptions of the pre-service teachers had been changed to accepted scientific models. Södervik, Mikkilä-Erdmann, & Vilppu (2014) contrasted refutational vs. non-refutational text, reporting better results using the first, and argued that refutational text awakens the readers’ meta-conceptual awareness when they compare text concepts with their own.

In summary, different approaches can lead to conceptual change in pre-service teachers, and methods used become more effective as they become more active and provoke higher metacognitive awareness. However, no study has assessed conceptual change through *LbT* or *LfP*, and only a few have considered learning impact over time, the aim of this work.

**learning from lectures, Learning from Peers, and Learning by Teaching**

This section reviews the conceptual background of the three approaches to conceptual change learning that we investigated.

**Learning from Lectures**

Lecturing remains the conventional method of instruction in higher education (Biggs, 2005, Freeman, 2012). Nevertheless it has been frequently criticized since effective teaching cannot be achieved by telling students what to think and why but by implementing meaningful activities that challenge everyday understanding (Loughran, Berry & Mulhall, 2012). Most lectures leave few opportunities to reflect on learning, do not strengthen critical skills and do not ensure the application of learning in practice (Azer, 2009). In such cases, lectures are considered a passive mode of engagement (Chi et al., 2014). One perception is that conventional lectures require less teacher expertise (Biggs, 2005).

Probably one of the most extensive meta-analyses contrasting active instruction methods with the conventional lecture was that of Freeman et al. (2012) who analysed 642 studies related to science, technology and mathematics learning. Results showed how the failure rate of students decreased from 34% for lecturing to 22% using active methods, while students’ scores increased notably. The study also highlighted that achievement increased with the amount of active learning and this was independent of the instructor. In this vein, studies dealing with misconceptions among pre-service teachers (eg. Trumper, 2001; Baser, 2006) showed that traditional lectures are less likely to rectify misconceptions than teaching approaches involving more intellectual and practical reasoning.

If the cognitive load (the total amount of mental effort used consciously by the working memory (WM) when the individual is performing a particular task) is not appropriate for the WM, learning will be ineffective (Chu, 2014). For unguided instruction methods like discovery, inquiry or problem-based learning (PBL), students use WM trying to discover the information or looking for the way of solving the problem, giving less room to assimilate new material (Kirschner et al., 2006). This is particularly important in novice students (Sweller, 2016). These authors proposed direct instruction, where all that students needed to learn was supplied in full, often through lectures, and often complemented by different type of activities.

Many studies have supported this theory for children (Klahr & Nigam, 2004; Matlen & Klahr, 2013; Chase & Klahr, 2017) but also for university students: Seery and Donnelly (2012) showed how pre-lectures about key concepts for chemistry students with little prior knowledge helped them during subsequent lectures by reducing the cognitive load. Similarly, Kennedy, Hirsch, Dillon, Rabideux and Alves (2016) using podcasts as a way of increasing student knowledge found better results and less perceived cognitive load. Wijnia, Loyens, van Gog, Derous and Schmidt (2014) examined the performance of guided and unguided PBL, finding greater learning in the guided group.

**Learning from Peers**

“Learning-from-peers” (*LfP*) consists of students learning from other students. To foster metacognition, a deep interaction between both parties is necessary (Roscoe, 2014; Roscoe & Chi, 2007). Spontaneous questions from tutees can motivate dialogue, scaffolding their own deep learning (Chi, Siler, Jeong, Yamuauchi, & Hausmann, 2001), and stimulating knowledge-building in both (Roscoe, 2014; Roscoe & Chi, 2007).

These effects are facilitated by tutors, who are concerned about the quality of the learning environment (Moust & Schmidt, 1995), thus creating a relaxed, comfortable, friendly, non-intimidating and interactive learning space (Hudson & Tonkin, 2008) compared with lecture situations. Tutors behave in a less authoritarian way, are more receptive to tutees’ problems, pay more attention to the preparation for final exams (Moust & Schmidt, 1995), are more accessible (Hudson & Tonkin, 2008) and give more time and freedom to discussion (Moust & Schmidt, 1994).

All of this offers greater equality and mutuality between both parties (Moust & Schmidt, 1994) giving the chance to admit uncertainty, speak up, question or speculate about any problem (Hudson & Tonkin, 2008; Moust & Schmidt, 1994), maximizing opportunities to verbalize learning, admit to mistakes and learn from them (Hattie, 2009), address doubts, and create interactive scenarios of dialogue that scaffold deep learning (Roscoe, 2014), in every case more so than in ordinary lectures. *LfP* might also enhance self-esteem and self-confidence (Topping, Campbell, Douglas, & Smith, 2003; Zentz et al., 2014), reduce anxiety (Zentz et al., 2014) and give greater satisfaction (Mills et al., 2014).

In addition, peer-tutors, with their recent experiences of learning the content, are better able to recognize particular challenges with the subject matter than lecturers and can focus on them to facilitate learning. They are more able to give suggestions because tutors are able “to speak the language of the students” (Moust & Schmidt, 1995, p. 298) making them more comprehensible, fitting Vygotsky’s zone of proximal development. Thus, tutors are more ‘cognitively congruent’ (Moust & Schmidt, 1994, p. 480). In contrast, teachers are likely to use their more extensive subject knowledge, which might make learning difficult. For science, Topping Peter, Stephen, & Whale (2004) also noted that students find the topic more interesting and enjoyable when it is explained by peers.

 Other reported benefits of *LfP* are that students achieve a better integration of theory and practice, stimulation to think from other perspectives, more effective critical thinking (Loke & Chow, 2007), skills in systematizing their learning and planning their studies (Arco-Tirado et al., 2011; Loke & Chow, 2007), better use of study material, improved cognitive and metacognitive strategies (Arco-Tirado et al., 2011) and, for pupils with learning difficulties, better results and attitudes towards school than in ordinary lectures (Topping, 2005). Finally, Chi, Kang and Yaghmourian (2016) exposed learners to videos showing tutor-tutee dialogues and contrasted these with monologue videos. They found greater learning gain after the dialogues, explaining that observers paid greater attention to tutees’ comments and built on them constructively.

However, some studies found negative effects among tutees when compared with conventional lectures: Brannagan et al. (2013) found that tutees learning laboratory skills felt more anxiety with peer-tutors than with lecturers. Indeed, with lecturers they received feedback that was moderately more helpful, had greater gains in critical thinking and problem solving, and increased their sense of responsibility. Learners might also have spent more time on self-study to compensate for their tutors’ lack of expertise (Moust & Smith, 1994). Discrepancies have also been found in tutees’ levels of performance. In most studies, tutees achieved similar results to control groups whether they were guided by peers or lecturers (Arco-Tirado et al., 2011; Hudson & Tonkin, 2008; Topping et al., 2004). In other studies, tutees performed better, whether the tutees were children (Topping et al., 2004) or undergraduate students (Roscoe, 2014).

**Learning by Teaching**

Teaching has always been central to human development. For Okita and Schwartz (2013, p. 376) teaching “involves a more knowledgeable person purposefully communicating accumulated experience by word and deed”. Teaching can also be used as a way of learning, in what is called “Learning-by-teaching”. This consists of selecting and organizing the most relevant content to be included in the teaching, developing the material, incorporating existing ideas, doing it in ways that make sense to the learners (Fiorella & Mayer, 2015) and giving feedback (Okita & Schwartz, 2013). *LbT* is equivalent to peer-tutoring or peer-learning (Herberg, Levin & Saylor, 2012), and Topping (2005) and Hattie (2009) emphasized that cooperative learning is a form of peer-learning.

For *LbT* to work properly, students should use a “knowledge-building strategy” (Roscoe and Chi, 2007; Roscoe, 2014), a process that helps learners to “metacognitively reflect upon their own expertise and comprehension, and constructively build upon their prior knowledge by generating inferences, integrating ideas across topics and domains, and repairing errors” (Roscoe & Chi, 2007, p. 541).

The literature shows many reasons why *LbT* achieves deeper learning: students must organize their own understanding into meaningful representations, filling gaps or clearing confusions (Burgess et al., 2014; Fiorella & Mayer, 2013; Roscoe, 2014) in order to deliver complete, suitable and correct explanations (De Backer, Van Keer, & Valcke et al., 2012). In this process students become curious (Roscoe, 2014), fostering logical reflection and further elaboration. They must take greater responsibility for their own learning (Brannagan et al., 2013) because they are accountable for others’ learning, creating a sense of duty and a determination to learn (Johnson, Johnson & Holubec, 1999). *LbT* facilitates greater involvement and increases motivation, a factor closely related to deep learning (Biggs, 2005; Loke & Chow, 2007).

An important feature of *LbT* is that tutors are challenged by learners’ questions when they see contradictions, lack knowledge or are simply curious (De Backer et al., 2012; Roscoe & Chi, 2008). This questioning deepens the tutor’s own learning (Fiorella & Mayer, 2015; Roscoe, 2014), fostering the re-examination of their own understanding in order to remedy inconsistencies (Hudson & Tonkin, 2008; Roscoe & Chi, 2008). Students’ dialogue leads to co-inferring processes (Chi & Wylie, 2014) and scaffolding interactions (Roscoe & Chi, 2007). Tutors also need to be alert to discover and correct learners’ misconceptions (Roscoe & Chi, 2007).

 In addition, research on *LbT* shows an increase in self-confidence and self-esteem (Loke & Chow, 2007; Hudson & Tonkin, 2008; Zentz et al., 2014) and a reduction in anxiety (Zentz et al., 2014). It enhances important skills like communication (Loke & Chow, 2007; Mills et al., 2014), time management and responsibility (Loke & Chow, 2007), critical and reflective thinking (Loke & Chow, 2007; Zentz et al., 2014) and lifelong-learning (Zentz et al., 2014). *LbT* fosters social skills (Arco-Tirado et al., 2011; Topping et al., 2003), personal growth (Loke & Chow, 2007) and professional development (Brannagan et al., 2013; Zentz et al., 2014). Teaching is an important skill for many professionals which is often not developed in curricula, so *LbT* can be a means of filling this gap (Burgess et al., 2014). Grossman et al. (2009) found that pre-service teachers had fewer opportunities to engage in approximations of practice than other professionals, so *LbT* seems particularly important for teacher training programs.

Nevertheless, there also are studies with less positive findings:Loke and Chow (2007) reported frustration in dealing with mismatched learning styles between tutors and tutees, and a high time commitment. There are also discrepancies in academic achievements: some studies reported better results (Fiorella & Mayer, 2014; Hudson & Tonkin, 2008; Loke & Chow, 2007) including for students with learning difficulties (Topping, 2005); others have not found an improvement (Brannagan et al., 2013; Burgess et al., 2014; Topping et al., 2004), and occasionally, students who teach have achieved less (Topping et al., 2004). Thus, Stigmar (2016) concluded in his review that this aspect is still unclear and requires further study.

**Research Questions**

Although many studies mention the potential of these three instructional methods to detect and address misconceptions, none deals with any misconception specifically for pre-service teachers. Therefore, the research questions for the study were:

1. How do the three learning methods compare in their impacts on the conceptual change of pre-service teachers over a short (1-month) and longer (2-years) term?
2. Why, from the perspectives of the participants 2-years after, do the three different learning methods differ in their effectiveness?

**Method**

**Participants**

The study used an intentional non-random sample approach with a single group of pre-service teachers with a within-subjects design: the same group served in all tested treatments. Although this had methodological challenges, it was considered the most ecologically suitable given the small number of participants and the curricular requirement to provide all participants access to all methods and concepts studied.

Students were in the final year of a 3-year degree at a public university in Spain. As part of their training to be primary school teachers, they studied science education for 2 hours per week for 15 weeks. This was the only science education component of their degree. For many students, this was the first time they had studied a science-related subject since middle school.

From the initial 58 students (22 female, mean age: 22.6, *SD*= 3.6), 40 (69%; 15 female, mean age: 22.8, *SD*= 4.2) completed the study protocols 1-month and 2-years after instruction. Between the initial and final samples there was no significant difference in gender (38 vs 37% female, *χ*2= 0.002, *p*>.05), in age (*U*: 1141, *p*>.05), nor in final exam marks (*t*= -1.254, p>.05), so the final sample can be representative of the initial. The lecturer explained the rationale for the research and requested participants to give informed consent after emphasising the voluntary nature of the research. Those who agreed were asked for e-mail addresses to enable future contact.

The intervention occurred in a conventional university classroom setting, being part of the degree program’s planned coursework and being one of three assignments with a similar design that was therefore familiar to the students (Author et al. 2011; 2015). Since it was a situated intervention, not all variables could be strictly controlled (Brown, 1992).

**Pre-test**

All pre-service teachers completed a pre-test at the beginning of the course in Fall 2009. They were asked to respond to eight statements dealing with misconceptions, selected from the authors’ experience and also widely discussed in the science education literature (Table 1). They were asked to answer “yes “or “no” to each sentence and requested to explain their answers. Items were presented as both statements and questions (Grasso, 2006) and they were mainly formulated in a positive way because some studies show that those in the negative can introduce confusion and errors in answers (Colosi, 2005)..

**Instructional Methods**

**Research and preparation for teaching**

Immediately after the pre-test, the participants were assigned to a group task lasting three weeks. Groups were required to become experts on a specific misconception. Working in cooperative learning groups ensured that there was sufficient expertise among the group members to be able to complete the task effectively since the teaching would be the product of all the students working together; this addressed a problem of reciprocal teaching that could have been caused by possible variation in teaching style and competence among group members (Seymour & Osana, 2003). This cooperative work could also foster an interactive mode of engagement that promotes a higher level of learning (Chi & Wylie, 2014).

Each group was required to: (1) select one theme (or more) from the eight listed in the pre-test, (2) investigate the theme to increase their subject knowledge, but focusing also on how to teach the concept, and (3) develop actively engaging activities (as recommended by the National Research Council, 2012) suitable for use by school pupils in a cooperating group, designed to help the pupils identify their own misconceptions and consider different explanations. This way, students learned the essential scientific principles but also emphasized the practical application of these principles.

These types of activities help to foster knowledge-building and thus they are a good approach to produce generative learning (Roscoe & Chi, 2007) which “involves actively making sense of to-be-learned information by mentally reorganizing and integrating it with one’s prior knowledge, thereby enabling learners to apply what they have learned to new situations” (Fiorella & Mayer, 2015, p. 1). Activities had to be able to engage school pupils in communicative actions, developing meaning-making and argumentation (Butler et al., 2014), and providing opportunities for pupils to connect with their previous knowledge and for group discussion and decision making (National Research Council, 2012). To do this, students had to formulate questions to promote discussion and to find strategies to enable pupils to contrast their mental models with the scientific one (Gadgil et al., 2011), but also to encourage discussion to generate further knowledge in an interactive mode of engagement (Chi & Wylie, 2014).

In a decomposition process (Grossman et al., 2009), students also had to plan to evaluate activities at all stages in order to inform their decisions on the most appropriate activities and learning strategies (Gomez-Zwiep, 2008) and to support pupils as they organize their knowledge, giving proper guidance (Kirshner et al., 2006). For this, students needed to anticipate any difficulties and have alternative activities available to suit different circumstances and different pupil abilities and starting points. Students were required to tailor multi-perspective activities that merged motivation, emotions and affective frameworks, as recommended when addressing pupils’ misconceptions (Duit et al., 2013; Duit & Treagust, 2003; Vosniadou, 2013).

**Instructor-led group tutorial and feedback on teaching plans.**

When students explore new themes by themselves, their understanding can remain superficial (Moust & Schmidt, 1994), disorganized, and can even cause new misconceptions to occur (Kirschner et al., 2006). To minimize this risk, the lecturer led a 15 to 30-minute tutorial for each group in which students explained the theory of the selected misconception, and how they planned to address it. Here, the lecturer ensured that enough knowledge and all the main elements of the concepts were considered and understood. Students also made use of other opportunities to consult the lecturer about any uncertainties. After three weeks, students submitted to the lecturer a written report containing an explanation of the misconception and a lesson plan. This was then reviewed by the lecturer to check that all essential elements were correct.

All students had previously received a workshop on cooperative learning and many other subjects required group-work activities. Nevertheless, to avoid a passive role during the whole process, each group had to appoint a coordinator to ensure that every group member played a full part. In addition, students were informed that the lecturer was to observe the peer teaching process, which encourages all group members to participate actively (Johnson et al., 1999).

**Learning-by-teaching.**

Next, each group taught another group in a rehearsal format (Grossman et al., 2009). The group that had studied the theme acted as “tutors” to, normally, three separate groups of their peers who had not studied that theme (“tutees”). Tutors were encouraged to intervene only to give necessary guidance (Kirschner et al., 2006) and promote reasoning by asking planned questions and addressing tutees’ own questions. Also, they were required to give the tutees opportunities to reflect on the misconception, on the activities and on the interactive discussion (Chi & Wylie, 2014) – and not simply explain the misconception to them.

**Learning-from-peers.**

Groups then switched roles of tutor and tutee, starting a new activity and theme, in a procedure similar to Aronson’s jigsaw in which all students are tutors and tutees at some point in the process. The tutees, in a cooperative learning environment, carried out the activities prepared by the tutors, asked questions and proposed possible explanations, replicating the process that would take place among pupils in a primary classroom. This was to promote knowledge-building strategies (Roscoe & Chi, 2007; Roscoe, 2014), forcing tutees to reflect on their own knowledge, integrating and generating new ideas through inference and reasoning in groups, leading to an interactive engagement (Chi & Wylie, 2014). Each block of peer teaching lasted 20-30 minutes. By the end, each student had received peer tutoring in normally three themes.

**Learning from lectures.**

On the day the *LbT* and *LfP* concluded, the lecturer gave formal explanations of all the misconceptions in a conventional 35-minute lecture which included some practical demonstrations, and in which students were encouraged to take notes (active engagement, in the Chi and Wylie [2014] ICAP framework). The purpose was to give all students the opportunity to consider briefly the key elements of all eight misconceptions. Explanations took 3-4 minutes per misconception and did not provide new information to the tutors nor to the tutees who had learned the topic from their peers. It was only new for those who were neither tutors nor tutees for each specific misconception.

**Summary of intervention.**

By the end, all the participants a) had learned about one or more misconception within their group, and taught it to several groups of peers (*LbT*); b) learned about some other misconceptions from their peers (*LfP*); and c) learned about the remaining misconceptions from the *Lecturer explanation*.

**Post-test and delayed post-test**

The same assessment of understanding used in the pre-test was included in a final exam for all students a month after the peer tutoring. This assessment had been used for different students in previous years. In the examination rubric, students were informed that questions answered as “yes” or “no” but without explanation would be treated as “no answer” and would be awarded no marks. In this paper, we refer to the data from this exam as the “1-monthafter” sample.

In early 2012, more than two years after the experience, consenting participants were invited to collaborate by completing the same assessment of science understanding. Again, the necessity of giving explanations for all answers - and not a simple “yes” or “no” - was highlighted. They were asked to do this assessment without reviewing any resource to help them answer the questions. Respondents were also invited to identify which of the three teaching methods used in the science education course was most effective in helping them learn deeper, retain longer, and explain why. This was the “2-years after” sample.

These opinions given two years after the end of the university course could be valuable because, by then, participants will have been less influenced by positives and negatives of daily university life. Equally, during the two years they might have had other learning and teaching experiences to contrast and provide a basis for reflection. However, because of the structure of the Spanish public education system, the probability that these students would have yet become classroom teachers was very low, so experiences in this time would likely have been informal.

 [TABLE 1 NEAR HERE]

**Data Analysis**

Students’ answers were scored as follows:

* “0”: Not answered, or answered with a simple “Yes” or “No”, because it was not possible to determine if the question was answered randomly.
* “1”: Incorrect answer (normally including a misconception). For example, for Misconception 5 (density and viscosity): “Yes, something that is viscous is because it is dense”; Misconception 7 (speed an object falls): “Sure, heavier objects are attracted stronger, so fall faster”
* “2”: Partially correct or an answer without full explanation. For example, Misconception 1 (photosynthesis and respiration): “No, plants need to do both as they are living things, but only respire during the night”; Misconception 2 (plant nutrition): “Most materials come from the soil, but also they use the sunlight”.
* “3”: Satisfactory response. For example, Misconception 3 (spontaneous generation): “Not at all, all living things are originated by other living things”; Misconception 6 (floor temperature): “No, because nothing is heating or cooling any floor but the environment. It depends on their thermic conduction capacity. Marble conducts very well, but not wood”.

Following Boesdorfer et al. (2011), 10 students were scored in full by the first author and independently by another science education lecturer in a process of reiterative cross-checking to secure judgements. Inter-rater reliability was very high (Cohen’s *Kappa* =0.855, *p*<.001). Answers of the remaining students were reviewed by the first author. For data analyses, allowances were made for the following characteristics of the sample data: a) comparison within a method: for any one method of instruction, answers coming from the same student were treated as related samples; b) comparison among methods: when comparing different methods of instruction, answers coming from different students were treated as independent samples. Although all the students experienced the three instruction methods, different misconceptions of each student were addressed. They were not the same ones for every student.

Since the values used in the analysis were ordinal, non-parametric statistics were applied. Friedman’s ANOVA was computed to check if, for each instruction method, results were different over time (pre-test, 1-month and 2-years after). For example, tests for *LbT* were: a) Pre-test, *LbT* vs. 1-month after, *LbT*; b) 1-month after, *LbT* vs. 2-years after, *LbT*; c) Pre-test, *LbT* vs. 2-years after, *LbT*. Similar tests were also applied to *Lecturer explanation* and to *LfP*. Wilcoxon signed-rank tests were computed, with the Bonferroni correction for multiple test, to discover effects over time for the three instruction methods.

Secondly, Kruskal-Wallis test for several independent groups was computed to help clarify any differences among the three instruction methods at each evaluation point in time. For example: a) 1-month after, *Lecturer explanation* vs. 1-month after, *LbT*; b) 1-month after, *Lecturer explanation* vs. 1-month after, *LfP*; c) 1-month after, *LbT* vs. 1-month after, *LfP*. Similar tests were applied for 2-years after. *Post hoc* Mann-Whitney test was used to compare these pairs of methods (also using Bonferroni’s correction).

The effect size test was computed to indicate the magnitude of the effect between contrasted methods. Cohen’s (1988) and Rosenthal’s (1996) criteria were applied as a benchmark: *r*<.30: small; .30≤*r*<.50: medium; .50≤*r*<0.70: large; *r*≥.70: very large effect. SPSS 19 was used to compute these tests.

Students’ written responses to open questions were analysed from different perspectives. To assess the most effective method, categorisation was kept flexible to allow for additional and unexpected themes beyond the three methods used. Equally, within each category, sub-categories were not defined in advance; instead they were derived from students’ answers using an inductive focus (Bryman, 2016). For this, following Johnson and Christensen (2008), data were firstly segmented, secondly coded and thirdly allocated to categories and sub-categories. For example, María’s answer (Table 5) was divided into three segments: The first, “I have to make an effort to understand the information perfectly”, was coded as “*LbT/GEL*” (Category: *LbT*; subcategory: Greater Effort to Learn). The second segment, “I can answer questions from my tutees”, was coded as “*LbT/GIn*” (*LbT*; Greater Involvement). The third, “I have reached a deeper understanding about the misconception”, was coded as “*LbT/ME*” (*LbT*; More Effective). Later, codes *GEL* and *ME* were merged in the same sub-category “Greater effort to understand, so teaching can be more effective”.

Additionally, from each segment, emotions were highlighted according to their relevance for learning and for conceptual change. They were divided (Broughton et al., 2013) into positive (enjoyment, pride, hope) and negative (anxiety, anger, shame, boredom, hopelessness). For example, Mati (Table 5) said, “…but we got it!” showing pride, and “I like it!” showing enjoyment. Similarly, Milan said, “…my activity, it wasn’t engaging” showing boredom.

Descriptive statistics were obtained for each category and sub-category. HyperResearch 2.8.3 was used to summarize and analyse qualitative data.

**Results**

Results are presented in two sections. Quantitative data show the changes in student knowledge over time for each learning method, then contrasts between methods, then the trends for each misconception. Qualitative data focus on participants’ views two years after the experience, in particular on the effectiveness of the different learning methods and emotional responses.

**Quantitative Test Results**

Descriptive statistics are shown for the pre-test, 1-month, and 2-years tests (Table 2).

[INSERT TABLE 2 NEAR HERE]

In the pre-test, most participants showed misconceptions for the majority of the concepts (Figure 1): 5.8% of the answers were completely correct, 24.0% partially correct, 59.2% incorrect, and 10.5% unanswered. In contrast, 1-month after the different interventions, 55.3% of the responses were completely correct, 26.6% partially, 10.3% incorrect and 7.8% unanswered. Two-years after, 49.1% were completely correct, 27.2% partially, 7.5% incorrect and 16.3% unanswered.

**Improvements over time for each learning method.**

Friedman’s ANOVA for repeated measures (*χ*2) showed that, for all methods, participants’ scores were significantly different over time. Wilcoxon signed-rank test (*Z*) compared pairs of scores for each instruction method. In all of them (see Table 2, last rows), comparisons of pre-test scores vs. 1-month and 2-years are positive (based on negative ranks) and statistically significant, considering Bonferroni’s correction (*p*<.0167). The effect size (*r*) was large and very large in all cases (except for *Lecturer explanation*: pre-test/2-years – medium-; and 1-month/2-years –small). This is good evidence that, whatever instruction method, students scored higher 1-month and 2-years later in relation to their initial knowledge.

However, there were some interesting differences between scores for 1-month and 2-years (Figure 1). For *Lecturer explanation* the relationship is negative and significant (based on positive ranks): students scored significantly less well 2-yearsafter instruction than 1-monthafter. The median and mean values both clearly decrease. There is a similar tendency for *LfP*, although this difference is not statistically significant. For *LbT*, the trend is opposite: students scored better 2-years after instruction than 1-month after, although this difference is not statistically significant.

[INSERT FIGURE 1 NEAR HERE]

**Learning method.**

Using the Kruskal-Wallis test, there was, as expected, no statistical difference in pre-test scores among groups of students involved in each different instruction method. However, subsequent scores were significantly affected by the method used 1-month and 2-years after instruction (Table 3, and Table 2 for scores).

*Post-hoc* analysis using the Mann-Whitney test (and considering Bonferroni’s correction) showed statistical differences (*p*<.001) in all cases except the comparison of *Lecturer explanation* vs *LfP* at 1-month after instruction (Table 3).

[INSERT TABLE 3 NEAR HERE]

To consider changes in performance over time, we compared the proportion of each score (Figure 1) and considered the effect size (*r*, Table 3).

*One-month* after instruction, the highest proportion of students scoring “3” and the highest mean value are those from *LbT* (80.4%, Figure 1, Table 2). *LbT* also had the lowest proportion of answers “0” and “1” (0 and 3.6% respectively). *LbT* is statistically better with a small-medium effect size for *Lecturer explanation* and *LfP*. Although mean values reveal a slight advantage to students involved in *LfP* vs *Lecturer explanation*, this difference is not statistically significant. Overall, students show better results with *LbT* after 1-month, even though the effect size is not especially large.

*Two-years* after instruction, *LbT* again has the highest proportion of students scoring “3” (78.6%, Figure 1) and the highest mean value (Table 2). The effect size is higher when comparing *LbT* with *Lecturer explanation* (moderate) than when comparing with *LfP* (medium). Although a higher proportion of students scoring “3” are from *LfP* (50.4%) than from *Lecturer explanation* (36.7%), the proportions scoring “1” are similar (9.4% vs. 8.8%) and the effect size is low. This means that any differences between these two methods are small.

In summary, the best results were achieved using *Learning-by-Teaching*. These students - who explored misconceptions for themselves, then developed activities for children, then taught their peers - scored higher at 1-month and 2-years. The results from the other learning approaches were not significantly different from each other 1-month after instruction, but after 2-years, *LfP* had higher scores than *Lecturer explanation* (albeit with small effect size). It is also noteworthy that all who had learned by *LbT* attempted to answer questions in all the post-tests (Figure 1), while the number failing to answer after *LfP* and *Lecturer explanation* increased from the 1-month to the 2-years assessment.

Even though learning from *Lecturer explanation* was less than from the other approaches anddeclined in the longer term, lecturing still gave an interesting moderate effect size, an outcome placed in what Hattie (2009) called the zone of desired effect.Lecturing could still be a fast and effective way to deal with a high number of misconceptions for pre-service teachers. . However, it is questionable whether learning from *Lecturer explanation* alone is enough to give newly qualified teachersthe understanding they need.

*LfP* achieved comparable results after 1-month and 2-years, suggesting that the knowledge gain was stable, unlike the results from *Lecturer explanation*. *LbT* gave the best empirical results with very large effect sizes and was also very stable. The data on the stability of knowledge gain are important results of this longitudinal study.

**Analysis of Specific Misconceptions**

Table 4 summarizes descriptive statistics, but not inferential, since the number of participants for each individual misconception is too small.

[INSERT TABLE 4 NEAR HERE]

For misconception 1 (“Is photosynthesis the plant’s equivalent of respiration?”) a high score was maintained after all methods of instruction, though in all cases mean scores decreased slightly from 1-month to after. Misconception 2 (“What raw materials do plants need for their metabolism and how do they obtain them?”) is the one with the highest mean after 1-month and 2-years for *Lecturer explanation*. For *LbT*, means are higher and remain the same after 1-month and 2-years. For *LfP*, mean post-test scores are lower than for the other instruction methods. In contrast, Misconception 3 (“Does spontaneous generation exist?”) has the lowest mean after 2-years, and the highest level of reversion for *Lecturer explanation*. Indeed, scores after 2-years are worse than in the pre-test. For *LbT* the mean is much higher and increases from 1-month to 2-years. For *LfP*, the post-test scores are not quite so high and they decrease slightly by 2-years. Misconception 4 (“Hardness and fragility are opposite concepts”) follows the typical decreasing pattern from 1-month to 2-years for *Lecturer explanation* and *LbT*. However, the mean increased for *LfP*, though the value is still below that for *LbT*.

 For Misconception 5 (“Density and viscosity are equivalent concepts”) the mean score for *Lecturer explanation* decreases from 1-month to 2-years, but increases in *LbT* and, to a lesser extent, in *LfP*. The mean score for Misconception 6 (“A marble floor is colder than a wood floor”) decreases markedly from 1-month to 2-years after *Lecturer explanation*, but increases for *LbT* and keeps the same value for *LfP*. Like Misconception 2, Misconception 7 (“The speed an object falls depends on its mass”) has a high mean after 2-years for *Lecturer explanation,* higher than for *LfP*. Misconception 8 (“Weight and mass are the same concept”) has the typical decreasing trend for *Lecturer explanation* and *LfP* between 1-month and 2-years, though it is more pronounced for *Lecturer explanation*, while students in *LbT* improve their score over time (Table 4).

In summary, in the longer term, *Lecturer explanation* only obtained moderately good results (mean values) for misconceptions 2 and 7, and was particularly ineffective for misconception 3, for which the majority of participants did not even answer the question 2-years after. *LfP* plays an effective role in learning with the exception of misconceptions 3 and 5. Apart from misconception 2 (and 6 in the shorter term), *LfP* always had higher means than *Lecturer explanation*. *LbT* was effective for all the concepts analysed, always gaining the highest outcomes, and indeed in two of them (misconceptions 3 and 6) all participants achieved completely correct answers. Interestingly, for misconceptions 3, 4 and 5 in *LfP* and for misconceptions 3, 5, 6 and 8 in *LbT*, participants showed a better understanding with time.

**Participant Views about Learning Methods**

During the 2-years post-test, students were asked to identify which of the three learning methods was most effective in helping them learn more deeply and retain their understanding. Their responses fell into four categories, each with several sub-categories (Table 5). However, due to the richness of some students’ testimonies, different parts of the same comment were assigned to different categories-subcategories.

First, *LbT* produced the deepest and most long-lasting learning: at the same time *LbT* was recognized as most effective learning method by 80% of the participants. 98 relevant statements were identified and grouped into nine sub-categories (Table 5):

*(a) Testing everything with experiences* (17 comments): Students appreciated the benefit of seeking, preparing and testing activities with their group peers, and implementing practical activities for their tutees during the *LbT* process. Ricardo saw these activities as the way to truly understand misconceptions and scientific concepts and he also highlighted the value of making mistakes as part of the learning process.

*(b) Careful research leads to metacognition* (16 comments):Here metacognition is the awareness by students of their own understanding of a concept (Karpicke, 2009). For example, Jonathan showed metacognitive thinking when he was becoming aware of what he did and did not know, and his own inconsistencies related to a concept. *LbT* requires higher order thinking as learners contrast and select important ideas and make conclusions. Also, Jonathan emphasised the need to test everything by providing activities for peers.

*(c) Greater degree of involvement with the learning process* (14 comments): Juan demonstrated this in several different ways: i) the large number of tasks entailed in the process, which made him “think and re-think about the concept” (also related to metacognition); ii) the influence of his peers’ opinions (the power of interaction with others); iii) the metacognitive process contrasting his own and his peers’ opinions; iv) the use of some higher order thinking skills (for example, “contrast information” or selecting “interesting activities”); v) the application of knowledge to new scenarios through different activities; vi) the power of cooperative learning.

*(d) Greater effort to understand, so teaching can be more effective* (14 comments): María and Pedro emphasized this value because they were determined to give their tutees an accurate understanding. They also noted that they were worried about peers’ questions and anticipating what peers might ask. Pedro demonstrated a willingness to learn (“I wanted to understand why”), an important prerequisite for good learning.

 *(e) Taking responsibility for teaching* (13 comments): *LbT* seemed to awaken students’ sense of responsibility. Esther’s comments also revealed some motivational and emotional components (“you are responsible for teaching something so important …”) that engendered a willingness to work harder. Her last point, on the importance of making concepts easy to understand, describes an essential feature of any teacher.

*(f) Greater reflection* (10 comments): Llanos’ comments fit Ausubel’s theory of meaningful learning and the Piagetian disequilibrium: she thought that *LbT* elicited her prior knowledge and challenged it with new evidence, leading to the breakdown of her old schemes. Llanos also highlighted the importance of tutees’ questions, forcing her to “think hard” to answer. In her opinion, this reflective process was what led her to remember and better understand the concept.

*(g) More time invested* (eight comments): Ana’s argument was based on her need to devote time to work with the misconception to understand it properly. She acknowledged that *LbT* involved a sequence of tasks: seeking and understanding information, report writing, developing activities with peers and preparing teaching materials, all of which required her to address the same misconception. This provided “enough time to think properly” about it, something she believed was not provided by *LfP* or *Lecturer explanation*.

*(h) Cooperative learning* (three comments): For some, discussing concepts within the group and helping each other to understand was the reason *LbT* was better. Mati emphasized the impact of listening to each other’s ideas to obtain the best results, a key feature of cooperative learning. Likewise, she expressed positive emotions (“… to celebrate the success”). Beatriz highlighted the important role of others in her group in developing her understanding of the concept, with positive emotions (“…my peers were happy and very patient…”) that denote social and affective interactions among the participants.

*(i) Motivation to teach* (three comments): Although all the students in this study were training to be teachers, only three emphasized this. Manuel’s comments emphasised situated learning. He explained the advantages of learning about something that he would use in his profession, and not simply deepening his theoretical understanding but complementing this with examples and activities useful for his future pupils. Manuel saw *LbT* as real training, a way of connecting university learning and the profession of teaching

Second, *LbT* and *LfP* were equally effective for five (out of 40) participants (12.5%), but in two different ways (Table 5):

*(j) Peers as tutors work just as hard* (three comments): Francisco, although he recognized that researching for himself was the best way for him to learn [we assume he was referring to his role in *LbT*], acknowledged that his tutors worked as hard and as effectively as he did, so he trusted that he could learn from them equally well. He valued receiving explanations from the tutors and interactions among tutors and tutees through questions, making the process active, constructive and interactive.

*(k) Quality of activities* (two comments): Esther P.’s opinion is relevant because she highlighted not only the importance of the peers’ activities to successfully achieve proper (and long-lasting) conceptual change but also mentioned the characteristics these activities should have: relevance (explicitly demonstrating the misconception), originality, attractiveness and encouraging active learning.

Third, *LfP* was better for two students (5%):

*(l)* Rocio’s comment justified the success of *LfP* from an empathetic perspective: she thought that when tutors prepared their teaching, they themselves found some difficulties in understanding the concept. Being aware of these difficulties for themselves helped tutors to imagine the perspectives of their tutees and focus specifically on these difficulties when teaching. This helped Rocio to understand “the trickiest things” of the misconception.

*(m)* Jorge considered *LfP* superior because it blended explanations of the misconception by tutors with the chance of doing simple activities addressing the concept. Because they were simple activities designed for children, they made it easier for him to understand the misconception. Jorge also noted that *LfP* modelled activities he himself could use for teaching with his future pupils.

Fourth, only one student felt that *Lecturer* *explanation* was superior. She argued that the lecturer gave the best illustrations and she remembered them. She also felt disenchanted with her own group’s activities and others groups’ behaviour, showing negative emotions. Perhaps she had an individualistic perspective. She criticized the process and asked for more supervision from the lecturer.

Finally, participants with views supporting *LbT* showed a general increase in self-confidence about their knowledge of the misconceptions. Ricardo, María, Ana, Mati and Beatriz expressed this clearly. Moreover, participants (María, Ana and Pedro, for example) felt ready (and confident) to address the doubts of their tutees. Independently of the learning method, none of the students made any comment related to stress.

**Discussion and concLusionS**

This study investigated longitudinal impact of *Learning-by-Teaching*, *Learning-from-Peers* and *Lecturer explanation* on pre-service teachers, at the same time exploring participants’ views on these methods. Taken as a whole, while pre-service teachers’ understanding of science misconceptions improved in all cases, there were significant differences in effectiveness: *LbT* > *LfP* > *Lecturer explanation*. Furthermore, the longitudinal perspective shows that knowledge acquired through *LbT* and *LfP* does not decline throughout the time, while *Lecturing* does

These findings matched participants’ views. They believe that testing any misconception with activities gives opportunities to make themselves aware of differences between their own and the scientific conception, which is the first step for change (Gadgil et al., 2001; Karpudewan, Roth & Chandrakesan, 2014). At the same time, new scenarios foster the creation of more differentiated and complex knowledge (Gadgil et al., 2001), and reflexion on similarities and contradictions between models is essential if learners are to revisit their understanding of how features of prior knowledge interrelate and focus on correcting misconceptions (Allen, 2010; Gadgil et al., 2001). The responsibility that students feel for teaching their peers well makes them feel more motivated, which leads to a greater effort to understand. In fact, Biggs (2005) asserts that the first element of learning is simply the willingness to learn. Finally, time is an essential requirement to achieve conceptual change, since change is often not dramatic but gradual (Vosniadou, 2013).

Furthermore, the deep social support and interactions that arise during group meetings, when students share their understanding and doubts, seek, select and test teaching activities, and prepare to teach, provide them with different viewpoints that are a starting point for new questions and answers (Karpudewan et al., 2014). This iterative dialogue fosters an interactive engagement (in the ICAP framework) which causes the highest level of learning (Chi & Wylie, 2014). All of this seems to raise students’ self-confidence, as shown in comments and quantitative data (all who learned by *LbT* attempted to answer in all post-tests, but this was not the case for *LfP* or *Lecturer explanation*), giving evidence of this benefit over the longer term.

Perhaps it is surprising that students show such significant learning given that not all group members necessarily participate fully in discussions; some might act more as spectators. Nevertheless, as Chi et al. (2016) demonstrated, dialog among peers can serve as a model of learning, and conflict among participants in the dialog can motivate and encourage onlookers to learn more through empathy or through a fear of failure.

One of the reported concerns about *LbT* is knowledge-tellingbias (Roscoe, 2014; Roscoe & Chi, 2007) that appears when tutors show little reflection or reasoning. Participants’ comments suggest the opposite, a knowledge-building orientation because of their frequent hints of metacognitive and reflective approaches. In addition, there are also recurrent remarks about peers’ or tutees’ challenging questions, with the recognition that these promoted learning and understanding. This process of searching for alternative explanations helps students deepen their own knowledge, addressing inconsistencies that helps to deepen conceptual change and make it long-lasting (Hudson & Tonkin, 2008; Roscoe & Chi, 2008).

Regarding *Lecturer explanation* and *LfP*, the longitudinal perspective gives interesting differences, with an important decline over time only for *Lecturer explanation.*. This is consistent with the ICAP perspective (Chi & Wylie, 2014): in lectures students are active (because they are taking notes), but in *LfP* are *interactive* (because there is mutual exchange of thoughts, resulting in new ideas for all participants) as emphasised in participants’ comments.

Interestingly, the fact that *LfP* and *Lecturer explanation* gave similar results at 1-month does not match the ICAP perspective. Our hypothesis is that, since students received the short lecture right at the end of the process after already addressing other misconceptions, they are more motivated to engage constructively (in the sense of ICAP) with the lecture than perhaps they would normally be in other lectures, although this seems to serve only in the shorter term. Further study is needed to confirm this.

Students also feel that peer tutors have a good understanding of which aspects of the science their peers would find most challenging, so they are in Vygotsky’s zone of proximal development, confirming a principle frequently used to explain the better outcomes of peer-teaching (Roscoe, 2014). Moreover, this reinforces the observation that teachers’ learning is enhanced when learning from other teachers (Fishman et al., 2014).

Finally, *LfP* blends explanations, questions and guided activities fitting Kirschner et al.’s (2006) defence of direct instruction based on Swellers’ cognitive load theory. As these authors note, careful explanations and guidance avoid overloading the working memory, increasing the amount and quality of learning and avoiding misconceptions. Also, VanLehn et al. (2007) show how students cannot follow a line of reasoning without a tutors’ help through frequent feedback. *LfP* shows evidence of full, active engagement by tutees and careful guidance from tutors, which could lead to better results than *Lecturer explanation*.

It is noteworthy that participants’ comments suggest that the blending *LbT-LfP* develops emotions positively, which seems directly related to deep learning and conceptual change (Duit et al., 2013; Duit & Treagust, 2003). Interactions among tutors and between tutors and tutees can create motivational and affective frameworks which enable some social, emotional and affective opportunities that help learning (Fiorella & Mayer, 2014, 2015; Roscoe & Chi, 2007).

Another contribution of this study is the interesting pattern shown by the “no answer” category which has not been included in other longitudinal studies (Trumper, 2006; Trundle et al., 2007). Most striking are the data for *Lecturer explanation*, where the number giving no answer in the longer term almost replicates the pre-test (Figure 1). We suggest this is because students were conscious that their original idea was incorrect, were not sure of the correct one and so preferred not to answer rather than get it wrong - or perhaps they did not remember the correct answer.

Regarding individual misconceptions, while *LbT* consistently obtained the best results, there are interesting differences. *Lecturer explanation* seemed especially poor at dealing with spontaneous generation. Samarapungavan and Wiers (1997) explained that formal instruction cannot easily provide an appropriate framework to structure or restructure an understanding of evolution. Nevertheless, *LfP* has one of the poorest results on this theme too. Perhaps students reverted to this misconception because they continued to make the same everyday observations as they had always made: worms or fungi simply “appear” (Vijapurkar & Konde, 2014). *LbT* shows excellent results: 100% of students responded correctly 2-years after, even better than at 1-month.

Misconception 5, on density and viscosity as equivalent concepts, obtained the poorest results in *LfP* and *LbT.* In everyday Spanish, both words are used indifferently, so perhaps students have not realised that science makes a distinction. However other studies also find this confusion to be highly prevalent among in-service teachers (Van Duzor, 2011).

**Implications**

This study shows that the combination of *LbT* and *LfP* through cooperative learning groups, and in the rehearsal format, has considerable potential for use in teacher education programs, since students learn difficult concepts with high self-confidence, select and implement effective and engaging learning activities, train to teach (situated learning) and see others translate ideas into practice (approaching communities of practice), all of which are elements highly demanded by teachers (Fishman et al., 2014). While Ball and Forzani (2009) consider that this type of authentic teaching practice is difficult to provide in teacher education programs, the format we used is a feasible way to give teacher education a classroom focus. Furthermore, focussing critical feedback and dialogue on their own small-group teaching (and not the whole classroom, where interactions might be diluted) enables them to gain further expertise.

Moreover, Ball, Thames, & Phelps (2008) see pedagogical content knowledge as a way of relating the academic world of disciplinary knowledge to the practical world of teaching, “placing the emphasis on the use of knowledge in and for teaching” (p. 394). This is what *LbT* achieves in this study: students recognise specific misconceptions of their own (and probably of pupils), explore and understand them in depth, develop strategies to address them with children, and practise these approaches by teaching their peers.

All of this also has implications for the participants’ practicum experiences, when new teachers have the opportunity to test, through their classroom practice, all they have learned from *LbT* and *LfP* and situate it within their broader philosophy of teaching, expressed in their testimonies.

However, even though *LbT-LfP* gives such good results, it is unrealistic to learn each difficult science concept this way because of the time it takes. Yet when complemented by brief lecture inputs on a wider number of less complicated misconceptions towards the end of the teaching period, the approach is likely to demonstrate a significant improvement across all of the concepts, with relatively little time invested.

Another important implication comes from the results from *Lecturer explanation* between 1-month and 2-years. Many studies comparing lectures with alternative methods use evaluation data collected immediately after the experience. Some of these studies do not find differences and some suggest that lectures give better results. However, these conclusions must be treated with caution because, as we show, impacts can be significantly different in the longer term.

**Limitations and Future Work**

We recognize that time-on-task was very different for the three methods because of their very nature. *LbT* involved hours of effort, *LfP* 20-30 minutes, and lecturer explanation only a few minutes, so the expected learning was always likely to be very different. However, we chose deliberately to explore natural interactions in authentic classroom settings to give high ecological validity (Raes, Schellens, & De Wever, 2014). Such authentic settings are central in the learning sciences, which look to study learning in realistic contexts rather than using artificial experiments (Bransford, Brown, & Cocking, 2000, in Reimann, 2011). In the spirit of Brown (1992), we wanted “improved cognitive productivity under the control of the learners, eventually with minimal expense, and with a theoretical rationale for why things work.” (p.167) so we used a familiar university environment and familiar teaching approaches. As a result, not all variables - like time - could be controlled. This gives reality to this study and an opportunity for conclusions to be transferred to other teacher education contexts.

A useful extension of this work would be to capture differences in the kinds of discussions during *LbT* and *LfP*, to evidence use and development of higher order and social skills, and to correlate these with individual results 1-month and 2-years after instruction. It would also be intriguing to explore whether outcomes from *LbT* and *LfP* differ if participants were to work as individuals rather than in cooperative groups.

**References**

Allen, M. (2010). *Misconceptions in primary science*. Berkshire, UK: McGraw-Hill.

Arco-Tirado, J. L., Fernández-Martín, F. D., & Fernández-Balboa, J. M. (2011). The impact of a peer-tutoring program on quality standards in higher education. *Higher Education*, *62,* 773-788.

Azer, S. A. (2009). What makes a great lecture? Use of lectures in a hybrid PBL curriculum. *The Kaohsiung journal of medical sciences*, *25*, 109-115.

Ball, D., Thames, M., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *29*, 389-407.

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and challenge of teacher education. *Journal of Teacher Education*, *60*, 497 – 511.

Baser, M. (2006). Fostering conceptual change by cognitive conflict based instruction on students’ understanding of heat and temperature concepts. *Eurasia Journal of Mathematics, Science and Technology Education*, *2*, 96-114.

Biggs, J. (2005). *Calidad del aprendizaje universitario [Teaching for quality learning at University*]. Madrid: Narcea.

Boesdorfer, S., Lorsbach, A., & Morey, M. (2011). Using a vicarious learning event to create a conceptual change in pre-service teachers’ understandings of the seasons. *Electronic Journal of Science Education*, *15*(1).

Brannagan, K. B., Dellinger, A., Thomas, J., Mitchell, D., Lewis-Trabeaux, S., & Dupre, S. (2013). Impact of peer teaching on nursing students: Perceptions of learning environment, self-efficacy, and knowledge. *Nurse Education Today*, *33*, 1440-1447.

Bransford, J. D., Brown, A. L., & Cocking, R. C. (Eds.). (2000). *How People learn: Brain, mind, experience and school*. Washington DC: National Academy Press.

Broughton, S. H., Sinatra, G. M., & Nussbaum, E. M. (2013). “Pluto has been a planet my whole life!” Emotions, attitudes, and conceptual change in elementary students’ learning about Pluto’s reclassification. *Research in Science Education*, *43*, 529-550.

Bryman, A. (2016). *Social research methods*. Oxford: Oxford University Press.

Bulunuz, N., & Jarrett, O. S. (2009). Understanding of Earth and Space Science Concepts: Strategies for Concept‐Building in Elementary Teacher Preparation. *School Science and Mathematics*, *109*, 276-289.

Burgess, A., McGregor, D., & Mellis, C. (2014). Medical students as peer tutors: a systematic review. *BMC Medical Education*, *14*, 115.

Butler, J., Mooney Simmie, G., & O’Grady, A. (2014). An investigation into the prevalence of ecological misconceptions in upper secondary students and implications for pre-service teacher education. *European Journal of Teacher Education, 38,* 300-319

Chi, M. T., Kang, S., & Yaghmourian, D. L. (2016). Why Students Learn More From Dialogue-Than Monologue-Videos: Analyses of Peer Interactions. *Journal of the Learning Sciences*, *26*, 10-50.

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, *25*, 471-533.

Chi, M. T., Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*, 219–243.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

Chu, H. C. (2014). Potential Negative Effects of Mobile Learning on Students' Learning Achievement and Cognitive Load-A Format Assessment Perspective. *Educational Technology & Society, 17*, 332-344.

Colosi, R. (2005). Negatively worded questions cause respondent confusion. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 2896-2903.

De Backer, L., Van Keer, H., & Valcke, M. (2012). Fostering university students’ metacognitive regulation through peer tutoring. *Procedia-Social and Behavioral Sciences*, *69*, 1594-1600.

diSessa, A. A. (2014). A history of conceptual change research: Threads and fault lines. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd Edition) (pp. 88-108). Cambridge, UK: Cambridge University Press.

Duit, R. & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*, 671-688*.*

Duit, R., Treagust, D., & Widodo, A. (2013). Teaching Science for Conceptual Change: Theory and Practice, in S. Vosniadou (Ed), *International Handbook of Research on Conceptual Change* (487-503). Oxon, UK: Routledge.

Fishman, B. J., Davis, E. A., & Chan, C. K. (2014). A learning sciences perspective on teacher learning research. In R. Swayer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 707-725). New York: Cambridge University Press

Fiorella, L., & Mayer, R. E. (2014). Role of expectations and explanations in learning by teaching. *Contemporary Educational Psychology*, *39*, 75-85.

Fiorella, L., & Mayer, R. E. (2015). Eight Ways to Promote Generative Learning. *Educational Psychology Review*, *28*, 717-741.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*, 8410-8415.

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

Galvin, E., Mooney Simmie, G., & O'Grady, A. (2015). Identification of misconceptions in the teaching of biology: a pedagogical cycle of recognition, reduction and removal. *Higher Education of Social Science, 8*, 1-8.

Gomez-Zwiep, S. (2008). Elementary teachers’ understanding of students’ science misconceptions: Implications for practice and teacher education. *Journal of Science Teacher Education*, *19*, 437-454.

Gönen, S. (2008). A study on student teachers’ misconceptions and scientifically acceptable conceptions about mass and gravity. *Journal of Science Education and Technology*, *17*, 70-81.

Grasso, L. (2006). *Encuestas. Elementos para su diseño y análisis*. Córdoba (Argentina): Encuentro.

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*, 2055-2100.

Hattie, J. (2009). *Visible Learning*. Abingdon: Routledge.

Herberg, J. S., Levin, D. T., & Saylor, M. M. (2012). Social audiences can disrupt learning by teaching. *Journal of Experimental Social Psychology*, *48*, 213-219.

Hudson, J.N., & Tonkin, A. L. (2008). Clinical skills education: outcomes of relationships between junior medical students, senior peers and simulated patients. *Medical education*, *42*(9), 901-908.

Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches*. London: Sage.

Johnson, D. W., Johnson, R. T., & Holubec, E. J. (1999). *El aprendizaje cooperativo en el aula*. Buenos Aires: Paidós.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*, 469-486.

Karpudewan, M., Roth, W. M., & Chandrakesan, K. (2014). Remediating misconception on climate change among secondary school students in Malaysia. *Environmental Education Research*, *21*, 631-648.

Kind, V. (2014). Science teachers’ content knowledge. In Venkat, H., Rollnick, M., Loughran, J. & Askew, M. (Eds.), *Exploring Mathematics and Science Teachers’ knowledge* (pp. 15-29). Abingdon, UK: Routledge.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*, 75-86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological science, 15*, 661-667.

Lehavi, Y., & Galili, I. (2009). The status of Galileo’s law of free-fall and its implications for physics education. *American Journal of Physics*, *77*, 417-423.

Loke, A. J., & Chow, F. L. (2007). Learning partnership—the experience of peer tutoring among nursing students: A qualitative study. *International journal of nursing studies*, *44*, 237-244.

Loughran, J., Berry, A., & Mulhall, P. (2012). *Understanding and Developing Science Teachers’ Pedagogical Content Knowledge* (2nd Edition). Rotterdam: Sense Publishers.

Matlen, B. J., & Klahr, D. (2013). Sequential effects of high and low instructional guidance on children’s acquisition of experimentation skills: Is it all in the timing? *Instructional Science, 41*, 621-634.

Monteiro, A., Nóbrega, C., Abrantes, I., & Gomes, C. (2012). Diagnosing Portuguese Students' Misconceptions about the Mineral Concept. *International Journal of Science Education*, *34*, 2705-2726.

Moust, J. C., & Schmidt, H. G. (1994). Effects of staff and student tutors on student achievement. *Higher Education*, 28, 471-482.

Moust, J. H., & Schmidt, H. G. (1995). Facilitating small-group learning: a comparison of student and staff tutors' behavior. *Instructional Science*, *22*, 287-301.

National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Okita, S. Y., & Schwartz, D. L. (2013). Learning by teaching human pupils and teachable agents: The importance of recursive feedback. *Journal of the Learning Sciences*, *22*, 375-412.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, *63*, 167-199.

Raes, A., Schellens, T., & De Wever, B. (2014). Web-based collaborative inquiry to bridge gaps in secondary science education. *Journal of the Learning Sciences*, *23*, 316-347

Reimann, P. (2011). Design-Based Research, in L. Markauskaite, P. Freebody, & J. Irwin (Eds.), Methodological Choice and Design (pp. 37-50), London, Springer,

Roscoe, R. D. (2014). Self-monitoring and knowledge-building in learning by teaching. *Instructional Science*, *42*, 327-351.

Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors’ explanations and questions. *Review of Educational Research*, *77*, 534-574.

Roscoe, R. D., & Chi, M. T. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science*, *36*, 321-350.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research, 21*, 37-59

Samarapungavan, A., & Wiers, R. W. (1997). Children's thoughts on the origin of species: A study of explanatory coherence. *Cognitive Science*, *21*, 147-177.

Seery, M. K., & Donnelly, R. (2011). The implementation of pre-lecture resources to reduce in-class cognitive load: A case study for higher education chemistry. *British Journal of Educational Technology*, *43*, 667-677

Seymour, J. R., & Osana, H. P. (2003). Reciprocal teaching procedures and principles: two teachers’ developing understanding. *Teaching and Teacher Education*, *19*, 325-344.

Södervik, I., Mikkilä-Erdmann, M., & Vilppu, H. (2014). Promoting the understanding of photosynthesis among elementary school student teachers through text design. *Journal of Science Teacher Education*, *25*, 581-600.

Stamenkovski, S., & Zajkov, O. (2014). Seventh Grade Students’ Qualitative Understanding of the Concept of Mass Influenced By Real Experiments and Virtual Experiments. *European Journal of Physics Education*, *5*, 20-30.

Stigmar, M. (2016). Peer-to-peer Teaching in Higher Education: A Critical Literature Review. *Mentoring & Tutoring: Partnership in Learning*, *24*, 124-136.

Sweller, J. (2016). Working Memory, Long-term Memory, and Instructional Design. *Journal of Applied Research in Memory and Cognition, 5*, 360-367.

Thorn, C. J., Bissinger K., Thorn, S., & Bogner, F. X. (2016). Trees Live on Soil and Sunshine!- Coexistence of Scientific and Alternative Conception of Tree Assimilation. *PLoS ONE,* *11*, e0147802.

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology*, *25*, 631-645

Topping, K. J., Campbell, J., Douglas, W., & Smith, A. (2003). Cross-age peer tutoring in mathematics with seven-and 11-year-olds: Influence on mathematical vocabulary, strategic dialogue and self-concept. *Educational Research*, *45*, 287-308.

Topping, K. J., Peter, C., Stephen, P., & Whale, M. (2004). Cross-age peer tutoring of science in the primary school: Influence on scientific language and thinking. *Educational Psychology*, *24*, 57-75.

Trumper, R. (2001). A cross-college age study of science and nonscience students' conceptions of basic astronomy concepts in preservice training for high-school teachers. *Journal of Science Education and Technology*, *10*, 189-195.

Trumper, R. (2006). Teaching future teachers basic astronomy concepts—seasonal changes—at a time of reform in science education. *Journal of Research in Science Teaching*, *43*, 879-906.

Trundle, K. C., Atwood, R. K., & Christopher, J. E. (2007). A longitudinal study of conceptual change: Preservice elementary teachers' conceptions of moon phases. *Journal of Research in Science Teaching*, *44*, 303-326.

Van Duzor, A. G. (2011). Capitalizing on teacher expertise: Motivations for contemplating transfer from professional development to the classroom. *Journal of Science Education and Technology*, *20*, 363-374.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, *31*, 3-62.

Vijapurkar, J., & Konde, P. (2014). “Omne Vivum Ex Vivo”? A Study of Middle School Students’ Explanations of the Seemingly Sudden Appearance of Some Life Forms. *Research in Science Education*, *44*, 885-902.

Vosniadou, S. (2013). Conceptual Change in Learning and Instruction. In S. Vosniadou (Ed.) *International handbook of research on conceptual change* (pp. 11-30). Oxon, UK: Routledge.

Wijnia, L., Loyens, S. M., van Gog, T., Derous, E., & Schmidt, H. G. (2014). Is there a role for direct instruction in problem-based learning? Comparing student-constructed versus integrated model answers. *Learning and Instruction*, *34*, 22-31.

Zentz, S. E., Kurtz, C. P., & Alverson, E. M. (2014). Undergraduate peer-assisted learning in the clinical setting. *The Journal of Nursing Education*, *53*, S4-S10.

Table 1. Concepts proposed to students, indicative correct answers, and number of students who learned about the concept by different learning approaches.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Indicative correct answer | *LbT* | *LfP* | *Lecturer* |
| *1. Is photosynthesis the plant’s equivalent of respiration?*(Galvin et al., 2015) | No. Like all living things, plants take in oxygen because they need it for respiration. Photosynthesis is a different process, absent in animals.  | 4 | 7 | 29 |
| *2. What raw materials do plants need for their metabolism and how do they obtain them?* (Butler et al., 2014; Södervik et al. 2014) | CO2 and O2 enters through leaves. Water and mineral salts are taken up through roots. | 7 | 9 | 24 |
| *3. Does spontaneous generation exist? (Sudden appearance of life forms)*(Samarapungavan and Wiers, 1997; Vijapurkar & Konde, 2014) | No. Any life always comes from another living being.  | 4 | 10 | 26 |
| *4. Hardness and fragility are opposite concepts*(Monteiro et al., 2012) | No. They are concepts expressing different properties of a material: hardness is the resistance to penetration or scratching; fragility is the ability to be broken under a deformation force. | 9 | 16 | 15 |
| *5. Density and viscosity are equivalent concepts*(Van Duzor, 2011; Stamenkovski & Zajkov, 2014) | No. They are different concepts expressing different properties of materials: Density is the mass of material in a given volume. Viscosity is about how easily fluids flow. | 7 | 20 | 13 |
| *6. A marble floor is colder than a wood floor.*(Baser, 2006)  | No. Both are the same temperature (environment temperature). The difference we feel is because marble is a thermal conductor and the wood is an insulator  | 7 | 17 | 16 |
| *7. The speed an object falls depends on its mass.*(Allen, 2010; Lehavi and Galili, 2009) | No. The speed it falls is independent of the mass. It depends of the air friction, the object’s shape, and the strength of gravity.  | 10 | 19 | 11 |
| *8. Weight and mass are the same concept.*(Gönen, 2008; Stamenkovski & Zajkov, 2014)  | No. They are related, but they are different concepts. Mass is the amount of material of any object. Weight is the strength with which gravity attracts it.  | 8 | 19 | 13 |

*LbT:* Learning-by-teaching; *LfP:* Learning-from-peers; *Lecturer:* Lecturer explanation.

Table 2. Summary of descriptive and inferential statistics for the three instruction methods tested for the 40 students who engaged in the whole study.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Lecturer explanation | Learning-by-Teaching | Learning-from-Peers |
|  | Pre-test | 1-month | 2-years | Pre-test | 1-month | 2-years | Pre-test | 1-month | 2-years |
| *N* | 147 | 147 | 147 | 56 | 56 | 56 | 117 | 117 | 117 |
| *Mdn* | 1 | 3 | 2 | 1 | 3 | 3 | 1 | 2 | 3 |
| *Var* | .573 | 1.033 | 1.443 | .497 | .254 | .171 | .464 | .881 | 1.057 |
| ** | 1.224 | 2.170 | 1.762 | 1.356 | 2.768 | 2.786 | 1.128 | 2.214 | 2.171 |
| *SD* | .756 | 1.016 | 1.201 | .699 | .504 | .414 | .676 | .936 | 1.028 |
| *χ*2(2) | 78.68 a | 93.76 a | 95.46 a |
|  | Pre-test / 1-month | Pre-test / 2-years | 1-month / 2-years | Pre-test / 1-month | Pre-test / 2-years | 1-month / 2-years | Pre-test / 1-month | Pre-test / 2-years | 1-month / 2-years |
| *Z* | -8.114 **b** | -5.078 **b** | -3.513 **c** | -6.327 **b** | -6.339 **b** | -.333**ns** | -7.625 **b** | -7.157 **b** | -.429**ns** |
| *r* | -0.67 | -0.42 | -0.29 | -0.85 | -0.85 |  | -0.70 | -0.66 |  |

*N*: number of answers in each category; *Mdn*: median; *Var*: variance; **: mean; *SD*: standard deviation; *χ*2: Chi-squared value (Friedman’s ANOVA, including degrees of freedom); *Z*: Wilcoxon signed-rank test; *r*: effect size.

Values: 0: “no answer”; 1: “incorrect answer (normally including misconception)”; 2: “partially correct or answer without full explanation”; 3: “satisfactory response”.

Pre-test / 1-month: contrast of pre-test results versus those obtained by the same students 1-month later; Pre-test / 2-years: idem pre-test vs 2-years after results; 1-month / 2-years: idem 1-month vs 2-years after the experience.

a *p*<.001; **b** *p*<.0167, based on negative ranks; **c** *p*<.0167, based on positive ranks; **ns** not significant.

Table 3: Kruskal-Wallis test (*H*) and post-hoc Mann-Whitney (*U*) test results, comparing scores in different temporal assessments for three different instruction methods.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Pre-test | 1-month | 2-years |
|  | *H* | *H* | *H* |
| *Lec-LbT-LfP* | 4.64 **ns** | 18.67 a | 36.02 a |
|  |  | *U* | *r* | *U* | *r* |
| *Lec-LbT* |  | 2756 a | -0.29 | 2088 a | -0.41 |
| *Lec-LfP* |  | 8496 **ns** | -0.01 | 6987 **b** | -0.17 |
| *LbT- LfP* |  | 2191 a | -0.31 | 2204 a | -0.30 |

*r* = effect size; *Lec:* Lecturer explanation; *LbT:* Learning-by-teaching; *LfP:* Learning-from-peers.

a *p*<.0001; **b***p*<.01; **ns** not significant

Table 4: Summary of the results for the three instruction methods detailed by misconception. Misconceptions are numbered 1-8 as in Table 1 and the main text.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | *Lecturer* | *LbT* | *LfP* |
| Misconception | Pre-test | 1-month | 2-years | Pre-test | 1-month | 2-years | Pre-test | 1-month | 2-years |
| **1** | *Mdn* (*Vr*) | 1 (.61) | 3 (.68) | 3 (1.24) | 1 (.25) | 3 (.00) | 3 (.25) | 1 (1.29) | 3 (.14) | 3 (.24) |
| ** (*SD*) | 1.55 (.78) | 2.45 (.83) | 2.2 (1.11) | 1.25 (.5) | 3.0 (.00) | 2.75 (.0) | 1.57 (1.1) | 2.86 (.8) | 2.71 (.49) |
| **2** | *Mdn* (*Vr*)** (*SD*) | 1 (.41) | 3 (.43) | 2 (.56) | 1 (.62) | 3 (.24) | 3 (.24) | 2 (.53) | 2 (.86) | 2 (.44) |
| 1.33 (.64) | 2.50 (.66) | 2.29 (.75) | 1.57 (.8) | 2.7 (.49) | 2.7 (.49) | 1.44 (.73) | 2.11 (.93) | 2.22 (.67) |
| **3** | *Mdn* (*Vr*) | 1 (1.24) | 2 (1.64) | 0 (1.72) | 2 (.67) | 3 (.25) | 3 (.00) | 1 (.46) | 2 (.29) | 3 (2.10) |
| ** (*SD*) | 1.04 (1.11) | 1.73 (1.28) | 0.96 (1.31) | 2.00 (.82) | 2.75 (.50) | 3.00 (.00) | 1.30 (.67) | 2.10 (.49) | 1.90 (1.45) |
| **4** | *Mdn* (*Vr*) | 1 (.57) | 2 (1.35) | 2 (1.69) | 1 (.78) | 3 (.11) | 3 (.25) | 1 (.90) | 2 (.43) | 3 (.92) |
| ** (*SD*) | 1.00 (.76) | 1.93 (1.16) | 1.60 (1.30) | 1.44 (.88) | 2.89 (.33) | 2.67 (.50) | 1.31 (.95) | 2.19 (.66) | 2.38 (.96) |
| **5** | *Mdn* (*Vr*)** (*SD*) | 1 (.36) | 1 (1.27) | 2 (1.06) | 1 (.48) | 3 (.9) | 3 (.29) | 1 (.47) | 2 (1.40) | 2 (1.54) |
| 1.23 (.60) | 1.46 (1.13) | 1.31 (1.03) | 1.14 (.69) | 2.29 (.95) | 2.57 (.53) | 0.95 (.69) | 1.65 (1.18) | 1.80 (1.24) |
| **6** | *Mdn* (*Vr*) | 1 (.06) | 3 (.67) | 2 (1.32) | 1 (.10) | 3 (.10) | 3.0 (. 00) | 1 (.05) | 3 (.60) | 2 (1.25) |
| ** (*SD*) | 1.06 (.25) | 2.50 (.82) | 1.88 (1.15) | 1.29 (.49) | 2.86 (.32) | 3.00 (.00) | 1.00 (.00) | 2.29 (.85) | 2.29 (.69) |
| **7** | *Mdn* (*Vr*) | 1 (.22) | 3 (.85) | 2 (.82) | 1 (.10) | 3 (.10) | 3 (.18) | 1 (.05) | 3 (.60) | 3 (1.25) |
| ** (*SD*) | 1.27 (.47) | 2.36 (.92) | 2.27 (.90) | 0.90 (.32) | 2.90 (.32) | 2.80 (.92) | 0.95 (.23) | 2.53 (.77) | 2.16 (1.121) |
| **8** | *Mdn* (*Vr*) | 1 (.41) | 2 (.69) | 1 (1.6) | 1.5 (.55) | 3 (.21) | 3 (.13) | 1 (.50) | 3 (1.01) | 3 (1.06) |
| ** (*SD*) | 1.08 (.64) | 2.23 (.83) | 1.46 (1.27) | 1.63 (.74) | 2.75 (.46) | 2.88 (.35) | 1.05 (.71) | 2.32 (1.00) | 2.21 (1.03) |

*Mdn*: median; *Vr*: variance; **: mean; *SD*: standard deviation.

Values: 0: “no answer”; 1: “incorrect answer”; 2: “partially correct or answer without full explanation”; 3: “satisfactory response”.

Table 5: Examples of student comments about their preferences for different methods of instruction. N represents the number of statements in each category, which is the same as the number of students highlighting them.

|  |
| --- |
| **Learning by Teaching is superior (**32 of 40 students) |
| a-Testing everything with experiences (*N*=17) | Teaching my peers, because we studied the concepts and developed activities which I firstly tested with my group. That is the way to learn! Drops of water vs drops of olive and sunflower oil - which was faster? This is fluidity and I understood my misconception while I was experimenting with it. And you know: “I hear and I forget, I see and I remember, I do and I understand”. I learned a lot from the other groups and their activities too, but I think it was less because we tested a lot of things and learned from the wrong activities (Ricardo) |
| b-Careful research leads to metacognition (*N*=16) | Because I have had to investigate, search for information from several sources, contrast and select the important ideas and organize them to find a conclusion, and test it with activities - but also show it to my peers - it is very different to simply copy-paste in a traditional report. Instead I have to blow my mind thinking about what I know or understand, and what I don’t (Jonathan) |
| c-Greater degree of involvement with the learning process (*N*=14) | With this method you have to do a lot of things: search and contrast information; use your own and peers’ opinions; change your mind schema; look for interesting activities; collect the required material; test the activities - and everything in a group. This makes me think and re-think about the concept (Juan)…with all of this, I do not want to devalue the lecturer’s role and explanations but, inevitably, in a system where I’m more involved I learn more than in a typical lecture (Sonia) |
| d-Greater effort to understand, so teaching can be more effective (*N*=14) | Since I have to make an effort to understand the information perfectly so I can answer questions from my tutees, I have reached a deeper understanding about the misconception (María)I couldn’t understand how the almost empty bottle of water reached the ground at the same time as the full one. We checked it three times, and there was no doubt! So we tested with aluminium foil balls and marbles and got the same result! I searched for information and studied it carefully because I wanted to understand why, but also to know how to answer my mates after showing the examples (Pedro) |
| e-Taking responsibility for teaching (*N*=13) | The fact that you are responsible for teaching something so important to your classmates makes you put a special effort into getting proper information, assessing it and thinking of a way to make them understand it easily (Esther) |
| f-Greater reflection (*N*=10) | Using this teaching method, I can link new knowledge with previous knowledge, leading to the appearance of doubts and contradictions, and breaking mental schemes that I believed to be completely right. But also, I had to think hard when my tutees questioned me. This reflection process is what has helped me the most to remember up to now (Llanos) |
| g-More time invested (*N*=8) | I think that the main reason why I remember my task is the time I used to develop it: to search for information, find data, write the report, be sure of my perfect understanding to show it to my peers, to get all the materials to develop the activity with my peers… Learning from peers, or from you, was too fast; in a short time we saw several misconceptions without enough time to think properly about them (Ana) |
| h-Cooperative learning (*N*=3) | Undoubtedly, teaching my peers was the one, because my group and I discussed a lot about what we knew and had already learned about temperature and about the sensations we feel, up to understanding why we were wrong - but we got it! All of us worked hard searching for information and interesting activities. We even fought, by giving each other reasons why each activity was the best for children and how to do it easily! At the end we discovered that a blend of ideas from several of us was best! Many questions challenged us, but among us, we successfully got it. I still remember some. After finishing, my group and I went to have a beer to celebrate the success! How can I forget it? I did not know this way of learning before, but I will use it with my pupils in groups when I am a teacher. I like it! (Mati)…at first I didn’t even understand what “hard” meant to a scientist, but my peers were happy and very patient, explaining it to me in different ways until I finally understood it (Beatriz) |
| i-Motivation to teach (*N*=3) | I think I learned more under *LbT* because I will be a teacher and this is like an authentic training to my future. So I tried to understand at a deep level the difference between weight and mass, …, to develop useful examples and activities (eg children’s weight in space and on the Moon) to help them [pupils] modify their incorrect concept (Manuel) |
| ***LbT and* *LfP* are equally good** (5/40)Tutors work as hard and as effectively as tutees |
| j- Research by yourself is the best way to learn, but my peers [tutors] researched the same amount as me, and having the opportunity of doing their activity, questioning and receiving their explanations, I can achieve the same and I think that I learned equally. It is an original and great way to learn (Francisco)k- For long lasting, memorable experiences from your peers [tutors], activities must be original, attractive and must clearly prove the misconception..., but also you [the learner] have to participate in them (Esther P) |
| ***LfP* is better (2/40)**Peers know what is difficult and focus on those aspects |
| l- They [tutors] used a way of explaining by emphasizing those things that they themselves had difficulty in understanding. This helped me very much in learning the trickiest things (Rocio) m- It [*LfP*] has the advantage of having a teacher’s explanation (which I got from my peers [tutors]) and at the same time doing simple activities for kids to highlight the misconception... This way, it is harder to forget it. And, with these activities, not only do I learn but also I can use them to teach my future pupils (Jorge). |
| ***Lecturer explanation* is superior (1/40)** |
| Sincerely, if I try to remember what we did about misconceptions, what I remember best are your [the teacher’s] examples, even better than mine. Your explanation and examples were attractive and had the necessary key elements to understand misconceptions perfectly, and it was well selected material… About my activity, it wasn’t engaging. I think that whether it is engaging or not depends on the imagination of the people in the group… Tutors don’t take it seriously when they are explaining and showing activities to each other. We need greater supervision by the teacher when we are developing tutors’ activities… (Milan) |

Figure 1. Proportion of participants obtaining the different scores (median values: 0: no answer; 1: incorrect answer; 2: partially correct; 3: satisfactory response) in initial and subsequent assessments, grouped by learning method.

