MARIANNE KORNER and MARTIN HOPF

CROSS-AGE PEER TUTORING IN PHYSICS: TUTORS, TUTEES, AND ACHIEVEMENT IN ELECTRICITY

Received: 25 October 2013; Accepted: 25 March 2014

ABSTRACT. International comparisons reveal that lower-secondary-level students in Austria perform below the OECD mean in science. Guided by the search for remedies and improvements in science teaching, this study investigates whether cross-age peer tutoring is an appropriate method for teaching physics. A modern and concise definition of peer tutoring is available from the review of previous findings which focus on tutors as well as on tutees. This clarification leads to an experimental setup mostly in tutor–tutee dyads where a sample of $n=172$ students from grades 5 to 8 underwent a cross-age peer tutoring process dealing with different topics within the context of electricity. The overall achievement in electricity for this age group was examined in a pretest–posttest design, using test items about electricity. Additionally, analyses were carried out in order to investigate whether or not there is a correlation between the possible roles within the process (active tutors–passive tutees) and the overall achievement. The results indicate that the active role is a crucial one for the achievement. Finally, a multiple linear regression model is presented which summarizes the research results and estimates the posttest scores based on the relevant parameters: the pretest score, the active role within the tutoring process, and the first language.

KEYWORDS: Achievement, Cross-age peer tutoring, Electricity, Physics, Tutee, Tutor

Introduction

Peer tutoring, a process where students work together with students, is an interesting and promising approach to teaching. It has been tested in various academic contexts such as reading, remedial math (Cohen, Kulik & Kulik, 1982; Robinson, Schofield & Steers-Wentzell, 2005), computer literacy (Fogarty & Wang, 1982), or thinking skills (Topping & Bryce, 2004). Additionally, there are studies that reveal encouraging outcomes that peer tutoring is as well an appropriate method for science teaching (Howe, Tolmie, Greer & Mackenzie, 1995; Lumpe & Staver, 1995). In his great meta-study, Hattie (2009) certifies peer tutoring quite powerful overall effects sizes with $d=0.55$ where cross-age tutoring is reported to be even more effective.

Electronic supplementary material The online version of this article (doi:10.1007/s10763-014-9539-8) contains supplementary material, which is available to authorized users.

While some of these studies report about the positive effects on both attitudinal and academic outcomes, others refer to positive outcomes for tutors as well as tutees. The success of peer tutoring is considered to be based on the effectiveness of social interaction (Fogarty & Wang, 1982; Lumpe & Staver, 1995). This idea can be traced to Vygotsky for whom "… the mechanism of individual developmental change is rooted in society and culture" (Vygotsky, 1978, p. 7).

One of the study's aims is to investigate whether peer tutoring is an appropriate method to improve the instructional quality of science teaching. In our opinion, the known results cannot easily be transferred because science teaching has to keep in mind the dominant roles of students' misconceptions and that students have to undergo a conceptual change (Duit & Treagust, 2003). Therefore, the peer tutoring interventions presented in this study are based on certain concepts selected within a specific context. For this context, we predominantly looked on the academic outcomes by analyzing the achievement before and after the intervention. Furthermore, the goal of this project is to link the findings of educational research to the demands of instruction in science teaching and thus build a bridge between research and practice.

This study was initiated after the results of PISA 2009 (OECD, 2010) revealed the poor academic achievement of lower-secondary-level students, especially in science. The age of the students who participated differed widely from kindergarten to upper secondary level. We focus on students from the lower secondary level (grade 5 to 8, that is age 11 to 14 years), an age group that is hardly ever an object of research in science education in the context of peer tutoring. In this study, we will present cross-age peer tutoring as a method to teach science and evaluate its outcomes based on empirical analyses of the students' achievement in electricity.

Previous Findings

Peer and Cross-Age Peer Tutoring: Definition

Peer tutoring has a long tradition whose beginnings can be traced back to the ancient Greeks (Topping, 1996). There are various forms of peer tutoring and different reasons to practice this method, which results in the fact that there is no consistent definition for it in literature. On the one hand, for instance, peer tutoring programs were implemented during the Industrial Revolution in England to compensate for the shortage of teachers. On the other hand, peer tutoring is also used in order to give older children responsibility for younger ones (Fogarty & Wang, 1982). Evidently, the widespread practice of supporting low-achieving students through private tutoring can also be considered a form of peer tutoring.

For our project, we need a modern definition of peer tutoring which considers the full range of aspects taking place during the tutoring process and therefore use Topping (2005): "It involves people from similar social groupings who are not professional teachers helping each other to learn and learning themselves by so doing" (p. 631). This definition already indicates a shift of the research focus from tutees toward tutors. To distinguish between peer and cross-age tutoring, we use the following definition according to Gaustad (1993): "Peer tutoring occurs when tutor and tutee are at the same age whereas in cross-age peer tutoring the tutor is older than the tutee" (p. 1). Gaustad, however, does not mention how much older the tutors should be and whether the age corresponds to the school grade. The study by Robinson (2005) recommends that the age gap between tutors and tutees should not be too large because their interaction is based on friendship which is closer than the relationship between a teacher and a student. Sometimes the term *peer tutoring* is used parallel to the broader term *peer-assisted learning* for both cross-age and same-age tutoring in literature (Robinson et al., 2005). In our study, we use the term *cross-age peer tutoring* to emphasize that tutors and tutees differ in *grade level* but belong to similar social groups concerning type of school and social background.

Research Results About Cross-Age Peer Tutoring

Previous (meta-)studies on peer tutoring cover a large scale of different settings, different grades, and different subject matters and investigate either academic, attitudinal, or socioemotional outcomes. An early meta-analysis conducted by Cohen et al. (1982) compared 65 different studies with different foci using effect sizes. They emphasized positive effects concerning students' achievements, attitudes toward the subject matter, and self-concepts not only for the tutoring students, but for the tutees as well. A majority of studies showed a better examination performance for tutored students than for students in conventional classes. The average effect size was 0.4 though it was larger for shorter, more structured, cross-age, and substitute tutoring programs restricted to basic concepts. Math tutoring programs had higher average effect sizes than tutoring programs in reading and not clearly described "other." For tutors, the average effect size in achievement was 0.33. Also, attitudes toward subject matter and self-concept improved but with moderate to small effect sizes. The range of grade levels covered by this meta-study was from one to nine.

A recent meta-analytic review by Rohrbeck et al. (2003) evaluated 90 studies about peer-assisted learning in elementary schools. Tutoring sessions lasted at least 1 week, and the only limitation on the subject matter was that it had to be academic. The impulse for this meta-study was to search for a remedy for the poor academic outcome of a large number of U.S. students, especially among vulnerable groups of students (i.e. those who belong to minorities or grow up in urban poverty). They state that peer-assisted learning interventions are more effective for younger, urban, low-income, and minority students.

The review of the literature by Robinson et al. (2005) focused on minority students as well, especially on African–American students. The age range covered was primary and secondary level, and the subject matter was mainly limited to mathematics. They note that few studies focus on the sometimes surprising academic gains of tutors. These gains may be observed even if tutors lacked additional subject matter instruction. Among other results, Robinson et al. state that longer tutoring programs are not necessarily more effective than shorter ones. These studies both criticized the lack of detailed information given about demographic and interventional parameters such as academic level or gender composition in some of their reviewed studies.

There is another older study (Fogarty & Wang, 1982) which focused on the verbal interactions that take place during the tutoring process and on learning and attitudinal progress of both tutors and tutees. Their tutoring programs covered the subject matters remedial math and computer literacy with elementary-level students as tutees and intermediate-level students as tutors. The authors concluded that the success of cross-age peer tutoring is based on a give-and-take friendship and therefore qualitatively different from that of a teacher–student interaction.

In his review of literature, Topping (1996) reports on the effectiveness of peer tutoring in further and higher education. In a summary about trends in peer learning (Topping, 2005), he emphasizes that results are typically very good if peer tutoring is well organized and implemented with focus on appropriate contexts. Then, even in the case of classwide tutoring, both tutors and tutees can improve their achievement.

An independent approach to a similar form of cooperative learning was created by German didactics in the early 1980s, the so-called *Lernen durch Lehren* (LdL)—"learning by teaching" (Martin, 1998). There are hardly any studies examining the outcomes for science education in this context. One of these rare studies dealing with physics is a pilot study conducted by Zinn (2008, 2009). In his cross-age design, middle school students (10 years and older) worked with primary and preschool students on various topics in physics. Zinn emphasizes a significantly increasing interest in physics, especially for female students, and recommends further studies based on fixed learning contents.

Implications from Previous Studies

The studies mentioned above give some implications for tutoring programs concerning age level, duration, or target group. In addition to that, though tutoring programs were originally understood to support the tutees' achievement, they have even more positive effects on the tutors' learning outcomes, which is emphasized by several studies (Fogarty & Wang, 1982; Robinson et al., 2005; Topping, 1996).

Robinson et al. (2005) discussed some more useful implications for the development of tutoring programs: Peer-assisted learning is a useful method to enhance both achievement in mathematics and attitudinal outcomes and need not be limited to high-achieving students as tutors. A combination of the roles of both tutor and tutee should be strongly considered. Concerning the tutor–tutee combination, the age gap should be modest to maximize the benefit for both groups. Single-sex tutoring dyads should also be considered. Fogarty and Wang (1982) emphasized the necessity of prior tutor training because of the tutors' limited instructional repertoire in the case of new contents.

Conceptual Change and Students' Conceptions

A unique situation is encountered with science education research and teaching: The students already have preinstructional conceptions in their minds, which originate from everyday life and language often labeled as *students' conceptions* (Duit, 2009; Wandersee, Mintzes, & Novak, 1994). They are partly contradictory to scientific conceptions and thus have to be fundamentally restructured during the learning process. This *conceptual change* has great influence on the way science is supposed to be taught. Conceptual change is not facilitated by just making students aware of their preinstructional conceptions, for they are usually very stable (Duit, Treagust, & Widodo, 2008). The best known approach to conceptual change is from Strike and Posner (1992) who claim four conditions that need to be fulfilled in order to facilitate this change: students' *dissatisfaction* with their prior concepts and a replacement conception which is *intelligible*, *plausible*, and *fruitful* (Duit & Treagust, 2003).

Therefore, it is necessary to create appropriate learning settings which support conceptual change in order to enable students to actively construct knowledge within the scope of the constructivist learning theory (Reusser, 2001). In their research about peer collaboration in physics Howe et al. (1995) report that peer tutoring facilitates conceptual change in the context of heating and cooling and can therefore be seen as an appropriate environment setting in the sense of Reusser (see above).

Research Questions

During the past few decades, a lot of research has been done on peer tutoring and other methods of cooperative learning (Treagust, 2007) with regard to various topics and different age groups. Although there are several studies about peer tutoring, there is one question that has not been sufficiently answered by recent research: Is this method also effective in the context of teaching physics? The answer to this question is not obvious because instruction in physics requires specific features in order to facilitate conceptual change on the basis of a constructivist learning theory. The abovementioned (meta-)studies deal with many different subjects, such as mathematics, reading, or computer literacy. Science topics are hardly covered explicitly in any study except that of Howe et al. (1995). Duit et al. (2008) propose that, based on a multi-perspective conceptual change approach, self-guided learning strategies support constructivist-oriented learning in science in a more suitable way than the traditional teacher-oriented instructional practice. This can be seen as an argument for the effectiveness of cross-age peer tutoring in the context of physics. Therefore, this study examines the achievement in the tested contexts in order to draw possible conclusions for the effectiveness of cross-age peer tutoring. However, caution should be exercised because tutees might possibly adopt their tutor's misconceptions. Consequently, these misconceptions might be even more ingrained.

Furthermore, there are few studies which examine secondary-level students in the context of peer-assisted learning. Robinson's meta-study (Robinson et al., 2005), as well as Rohrbeck's meta-study (Rohrbeck, et al., 2003), describe predominantly U.S. minority and at-risk students. Additionally, Rohrbeck claims that peer-tutoring interventions were more effective with younger elementary school students. Therefore, it is left open whether this method is appropriate in the European cultural context as well as for secondary-level students. One of the implications made in Robinson's study was to take into account that tutoring programs have positive academic outcomes for tutors as well. Accordingly, not only the tutees', but also the tutors', achievement has became one of our research goals.

The present study focuses on electricity because, according to the Austrian curricula, this topic is relevant for all of the participating classes. Also, students' preconceptions in electricity are well-known (Stork & Wiesner, 1981) and research on conceptual change teaches us to focus on specific conceptions. Thus, this study was guided by the following research questions:

Research Question 1: What Is the Achievement in Electricity for Secondary-Level Students According to Cross-Age Peer Tutoring?

This investigation aims to find evidence whether or not cross-age peer tutoring is an appropriate method for teaching electricity to secondary-level students who are rarely covered by previous studies. The tested hypothesis is that an achievement in electricity can be observed for all secondary-level students within the context of electricity.

Research Question 2: Is There an Increased Achievement in Electricity for Tutors?

Older studies have reported that there is a considerable benefit not only tor tutees but also for those students who act as tutors (Cohen et al., 1982; Fogarty & Wang, 1982; Topping, 1996). Therefore, our hypothesis contains a considerable benefit for tutors in the context of electricity.

Research Question 3: What Are the Differences in Achievement in the Posttests due to the Three Different Roles (Tutor, Tutee, Both) that Students Had During the Tutoring Process?

If an overall achievement in electricity can be found, further analysis will be needed. The hypothesis, deduced from results of previous studies (Cohen et al., 1982; Hattie, 2009; Topping, 1996), is that there are differences in the achievement between these three groups. Comparisons between the groups will be conducted by planned contrast analyses. Another point of interest concerns possible effects due to the treatment's different durations for the subgroups. According to the existing literature (Robinson et al., 2005), no such effects are expected. Therefore, the research hypothesis is that there are no differences in the achievement due to the treatment's different durations.

Research Question 4: What Are Relevant Predictors to Estimate the Posttest Scores?

The idea is to create a model by means of multiple linear regression (MLR) that estimates the posttest scores of students after cross-age peer tutoring. On one hand, this model summarizes the results from research questions 1 to 3. On the other hand, the relevance and the individual contribution of the pretest and the role as well as the demographic parameters gender, first language, and grade in physics for the achievement in electricity are also examined. Our hypothesis is that first language and gender influence the outcomes, whereas the grades in regular classes do not contribute to the posttest scores.

Research Design

Description of the Sample

This study was conducted in Austrian schools in and around Vienna within the context of a larger project on *Cross-Age Peer Tutoring in Physics* (CAPT). For the whole project, the age of the children and students ranged from 5 years (kindergarten) up to 17 years (upper secondary level). According to the research questions, in this study, we confined ourselves to students aged 11 to 14 years, which corresponds with grades 5 to 8. This restriction results in a total sample size of $n = 172$ students investigated. The questionnaires used were optimized for this specific age group. The grade level for the tutors was 6 to 8 (ages 12 to 14) and that of the tutees was 2 to 7 (ages 8 to 13) resulting in an average age gap of 2.82 years and a standard deviation of 1.99. In two cases, the age gap was 0 years (same-age tutoring, but different types of schools), and in one case, when tutees were in kindergarten, it was 6 years. According to Robinson et al. (2005) and Fogarty & Wang (1982), the age gap between tutors and tutees should not be too large, which is the case in most situations. Beyond that, hardly anything is found in previous studies about the influence of the age gap. A gap of 0 years, which means same-age tutoring, at most reduces the observed effect sizes (Hattie, 2009) but does not bias the analyses in any other way.[1]

Due to organizational reasons, it was not possible to draw a more homogenous sample. Therefore, the circumstances of the project made it necessary to conduct a field study.

The analyses of the pretests showed that, based on this sample, it was still possible to answer the research questions because these tests did not vary significantly over all participating classes. The focus on lower-secondary-level students was chosen because the sample size of upper secondary-level students was too small.

The most important criterion that participating classes had to meet was that cross-age peer tutoring could be accomplished without too much organizational effort. Thus, four schools were selected in which different types of schools were located within the same building. Besides this criterion, the four schools were chosen in order to get the most representative sample possible. School 1 is a private school in the center of Vienna, attended by hardly any students with migration background; the second one is a public school, located also in the center of Vienna, with classes of up to 90 % students with migration background, according to the definition of OECD (Breit, 2009). The third one is a private school as well, but is located at the periphery of the city. The participating teachers of all three schools were involved in the training of the preservice teachers. The fourth school is situated in a more rural environment, 20 km outside of Vienna. All in all, there were seven urban classes and six rural classes. However, we tried to include different types of schools with different sociocultural backgrounds. The reader can claim that the sample is biased because of the overrepresentation of private schools, but we would like to mention that, in Austria, all schools follow the same curriculum and all teachers, in private as well as in public schools, are equally paid by the community.

Most of the classes chosen for this study belonged to a type of compulsory school which qualifies their students for a job rather than for higher education such as colleges and universities. Therefore, the attending students' abilities in science, mathematics, and reading are significantly below the mean for this age group (OECD, 2010).

Although there is evidence found in literature, as reported above, that cross-age peer tutoring works in various situations, it might be different and more difficult when trying to induce conceptual change. If it is possible that cross-age peer tutoring enhances the achievement and thus initiates a conceptual change in the tested low-achieving population, it is likely to work even better for different populations of higher-achieving students.

All students tested and monitored, as well as their parents, had given their informed written consent.

Subject Matter

Following Zinn (2009), we based the interventions on fixed learning contents. According to literature (Duit & Rhöneck, 1998; Shipstone, 1984), we selected basic topics within the context of electricity. The interventions and the learning materials had been designed based on findings of research about students' conceptions and instructional strategies (Shaffer & McDermott, 1992). For each class, there was a specific plan about which students' conceptions would be addressed depending on the students' grade level. The learning materials were selected accordingly. The main students' conceptions that were addressed included the following: electric circuits must be closed; current has a certain direction; current is constant; and there is a connection between current, resulting resistance, and the brightness of equal bulbs. The resulting resistance in parallel and series circuits was also discussed. For the instruction of primary students by secondary-level students, the main focus was on the basic concept that electric circuits have to be closed.

Despite the fact that the interventions took place at different times of the year and at different grade levels, we asked the regular class teachers not to give the students previous or additional instructions in electricity. This seems crucial, especially during the time between the tutor training and the tutoring. There was only one class that already had instruction in this subject matter 1 year earlier. As a precursor to the results, this class did not perform significantly better in the pretest anyway.

Mentoring and Tutoring

Before the tutoring sessions, the tutors were trained in a so-called *mentoring*. According to Fogarty and Wang (1982), it makes sense to do this if the subject matter is completely new to both the tutors and the tutees. Each mentoring was conducted by the same researcher involved in the project. A mentoring has to fulfill several purposes. First of all, the soon-to-be tutors should become familiar with the subject matter and the associated scientific concepts (which in some cases will be contrary to their own). Secondly, the students get the opportunity to realize and to reflect on their own conceptions in order to construct new ones which match the scientific ones. This helps the tutors in their ability to monitor the tutees' conceptions, to regard these conceptions as a basis of the learning process, and to find a way to reconstruct them in order to lead to scientifically correct ideas. Another aspect of the mentoring is that students get involved in the preparation and selection of suitable learning materials for their teaching purposes. Tutors were asked to base their instructions on certain student conceptions.

For the mentoring, the tutors received a list containing a variety of tasks and suggested experiments (materials available in German only on our webpage—reference hidden due to blind review process). Additionally, they were provided experimental materials such as bulbs, wires, batteries, and small electric motors.

Firstly, tutors started with completing these tasks[2] on their own or in a dyad with their neighbor. Students were asked to *predict* the possible result, then to *observe* the experiment and finally *explain* the observed results according to the POE strategy (White & Gunstone, 1992). The following group discussion was moderated by the researcher. The correct answer as well as possible misconceptions was discussed. The discussion about the students' own misconceptions is seen as a fruitful basis to prepare them for their tutees' conceptions.

Afterwards, students were asked to select those tasks and experiments that, in their opinion, were most suitable for the tutoring with respect to the tutees' grade level. It was even possible for the tutors to invent new exercises or experiments, if they matched their teaching purposes. This was accomplished in few cases. With regard to the students' motivation, they were told that from that moment on they were the experts for teaching the younger students. Due to the timetables in the individual schools, the duration of the mentoring varied from 60 to 80 min. It took place 2 to 3 weeks prior to the tutoring.

Just before the tutoring itself took place, the soon-to-be tutors practiced another 20 to 30 min for the tutoring session. They prepared their experimental materials and also received cue cards to scaffold the tutoring process. These cue cards included questions and experiments they had selected for their tutees. Finally, there was a brief instruction on *how* to teach. We told the tutors that they are not supposed to do everything themselves, but rather to have the tutees create their own experiences. Their instructions should be based on the POE strategy as well.

The tutoring started with the distribution of experimental materials like batteries, bulbs, and cables. In addition, the tutors also received their cue cards including exercises, questions, or experiments. These were the exercises they had selected at the end of the mentoring for their teaching purposes. As a little hint in case a tutor forgot the correct answer, there were keywords on the back of each card. The whole tutor class and the whole tutee class met in a classroom; therefore, we call it classwide tutoring. We preferred the more intense one-to-one setting. So, *each* student of the tutor class instructed one student of the tutee class. In some rare cases, when the number of students in the classes differed, it was necessary that two tutors worked with one tutee or the other way around. The tutor–tutee dyads were assigned randomly. Therefore, the tutor–tutee pairs were coeducational because in Austria it is not common to have single-sex classes and gender aspects do not cover the research focus.

Moreover, we operated on a cross-ability basis. It was not intended that the tutor group consisted solely of gifted students and that tutees were underachievers or at-risk students.

The tutors started the intervention by asking the tutees about the tasks and inviting them to probe their assumptions by experiment. This process should lead to further discussions.

The actual instructing process lasted for 30 to 45 min, depending on the students' eagerness. Tutoring was stopped when either the time ran out or the students reported that they have completed all exercises, experiments, and discussions.

A typical sequence of mentoring and tutoring for a *tutor-only* group (class A) and a *tutee-only* group (class B) is shown in Fig. 1. Class A received a mentoring and then the tutoring took place. Three of the classes (type A) had an additional tutoring with another class.

Another typical setting is shown in Fig. 2. Students of *class B* switched roles and first acted as tutees and then as tutors following mentoring 2. Thus, class B received a tutoring and a mentoring and ran another tutoring.

One may claim that the classes spent differing times on tasks and, therefore, their achievement in the posttest would be different. While Hattie (2009, p. 184) explains that the quality of learning is essential, the influence of the time on a task is less than average. Cohen et al. (1982) and Robinson et al. (2005) report that shorter programs are more effective, presuming them to be more structured. Our tutoring program is perceived to be highly structured and thus effective, and the influence of the time on a task, on the other hand, is subordinate. This matches preliminary results from MLR where the *time of intervention*, which does not mirror the quality of intervention, did not turn out to be a significant predictor.

Achievement in Electricity

For this study, we preferred a pretest–posttest design rather than compare treatment with control groups. This was done because it seemed a priori impossible to tell which parameters are relevant and thus had to be controlled.

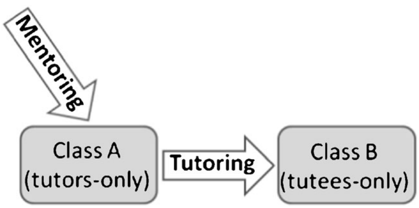

Fig. 1. Sequence of mentoring and tutoring

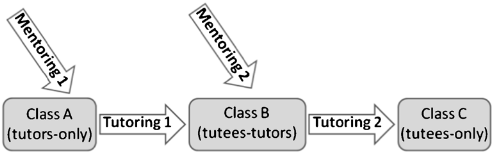

Fig. 2. Mentoring and tutoring with switching roles

For the tutor-only group and the tutee–tutor group, the pretest was conducted immediately before the mentoring and after the (second) tutoring. For the tutee-only group, the testing points were before and after the tutoring. Regular class teachers were asked to give no additional instruction to the students before the mentoring and between mentoring and tutoring in order not to bias the effects of cross-age peer tutoring. Except for one class, all classes had had no previous instruction in electricity.

The test instrument used in this study was a multiple choice conceptual knowledge test which was developed on the basis of the test of Rhöneck (1986) and on DIRECT—Determining and Interpreting Resistive Electric Circuit Concepts Test—by Engelhardt and Beichner (2004). It is based on students' conceptions mainly using two-tiered items (Urban-Woldron & Hopf, 2012). On the first tier, students have to answer a certain question, e.g. "How bright will the bulbs in the figure be?" In the second tier, they tick an explanation for why they had chosen a certain answer. The distractors in this test were based on previous interviews with students about their conceptions in electricity.

We selected the first five items of this test which corresponded to the students' conceptions addressed in the tutoring process. Due to the brief intervention and the limited number of students' conceptions addressed, the items were counted as sum scores.

Students were asked to complete the tests on their own as accurately as possible. They were told that the test was anonymous and had no influence on their grades. It was also important to tell them that researchers were not interested in their individual performance but in how well cross-age peer tutoring works as a teaching method. Observations during the data collection showed that students were very honest and eager while filling out the test items.

Results

Before reporting the results, we want to clarify two details about the analyses: Single missing values in the knowledge test were interpreted as *not known*. If a whole questionnaire was missing, this was interpreted as *student was absent*. Concerning the reliability of the knowledge tests on electricity, Cronbach's α for the pretest was found to be .78 and, for the posttest, α=.85.

Sample Characteristics

In the following, the characteristics of the sample are described as accurately as possible in order to build the basis of comparable conclusions. This is also done because, at first glance, the sample seems to be rather inhomogeneous regarding the wide range of grades (5 to 8).

The total number of students investigated was n=172. There were 62 % boys and 38 % girls. The ratio of boys to girls is representative for this type of school in Austria. This ratio was not supposed to be of statistical importance, but we had to be more careful with the interpretation of the results according to research question 4. There were nine secondary-level classes involved in this project. Regarding the different roles, the distribution was 94 tutors-only (55 %), 43 tutees–tutors (25 %), and 35 tutees-only (20 %). For 68 % of the students, their first language was the same as the language of instruction (German), and 32 % had a different first language. The grades of the students who were actually tested in this study ranged from 6 to 8 (M=6.47, SD=0.501).

Since the classes participating in the study appeared to be diverse, the first important thing was to find out their preinstructional knowledge in electricity. In order to do this, the pretest scores were compared in an attempt to find out whether classes differ significantly in their pretests. This hypothesis had to be rejected, because ANOVA shows no significant difference between the classes (p=.489; $F_{(8,163)}$=0.936). Therefore, the classes had apparently been drawn from the same population although they differed in grade and social background and one of them had already had previous instruction in electricity prior to the study. The homogeneity of the sample in the pretest scores builds the essential basis for the interpretation of the posttest scores.

Another question concerning the assignment of the tutor and tutee roles to the classes was that of randomness. It might have happened that classes with better-performing students were primarily suggested for tutoring by their class teachers. In this case, ANOVA again shows no significant differences in the pretest scores for the three groups tutor-only, tutee–tutor, and tutee-only (p=.113; $F_{2,169}$=2.205) so that therefore the assignment can be considered random.

In order to answer research question 1, the pretests and posttests of all students were compared. In this first analysis, we did not distinguish between roles (Table 1). Analyses therefore represent the intraindividual gain in knowledge. Both tests were available from n=164 students. Pretests as well as posttests were rated as sum scores, with a maximum score of 9 points. As described above, we expected observable gains in electricity.

The t test on the paired differences (Table 2) indicates a highly significant difference between the mean scores of pre- and posttests within the population ($t_{163}=5.826$, $p<.001$). The overall effect size due to the treatment is 0.46 with a statistical power of .99.

In order to find an answer to research question 2, the investigated sample was reduced to those students who acted as tutors-only. Those who acted in the tutee–tutor role were not taken into account yet, in order to distinguish between the results according to the tutoring process and the duration of the treatment. The comparison of the two active groups was carried out in the analyses related to research question 3. The number of tutors for whom pre- *and* posttests were available shrank from n=94 to n=91 because of student absences during either pre- or posttest. The sum scores of pre- and posttests for this subgroup were compared. A t test on paired samples indicated a highly significant gain for tutors ($t_{90}=3.976$; $p<.001$) within the population with an overall effect size of 0.41 and a statistical power of .95.

As explained in the description of the sample, there are no significant differences in the pretests according to the classes and to the role assignment. The following analysis, according to research question 3, tested whether there are differences in the posttest scores if the sample is split up by the roles.

This was done by an ANOVA.[3] Additionally, analyses were carried out based on different assumptions for the reasons of possible differences in the posttests. Regarding the results from literature, it could be expected that the role is the only parameter which splits the sample into differing subgroups. Neither the grade level nor the affiliation to a certain class reveals significant differences in the posttests.

TABLE 1

Descriptive statistics of pre- and posttests

Paired-samples statistics				
		Mean	*N*	*SD*
Pair 1	Pretest	4.41	164	2.59
	Posttest	5.84	164	2.85

TABLE 2

Comparison of pretest with posttest

Paired-samples test						
		Paired differences				
		Mean	*SD*	*T*	*df*	*Sig. (2-tailed)*
Pair 1	Pretest–posttest	−1.421	3.123	−5.826	163	.000

For a sample size of $n=164$, ANOVA endorses highly significant differences in the posttest scores between a grouping based on the role ($F_{2,162}=6.716$; $p=.002$) with an effect size of 0.66 and a statistical power of approximately 1. Figure 3 displays the mean posttest scores due to the students' roles including a 95 % confidence interval.

Further analyses by planned contrasts were carried out to find out more details about the differences. The most interesting contrast appeared when students who took part in the active role (tutor-only, tutee–tutor) were compared to those in the passive role (tutee-only). *T* statistics showed the significance of this contrast indicating a much better performance for the active group in the posttests ($t_{49.8}=3.320$; $p=.002$).

A point that has not been answered sufficiently by these analyses is whether or not the duration of the treatment has an effect on the outcome. The tested hypothesis was based on the assumption that students who took part in the active *and* the passive role had spent more time on the subject matter than those in a single role. This could influence their achievement. For this analysis, a planned contrast compared the two subgroups tutor-only and tutee–tutor. *T* statistics showed no significance of this contrast ($t_{162}=-1.461$; $p=.146$), which implies that there is no effect due to the duration of the treatment in the population. This means that a longer treatment does not necessarily lead to better performance in the posttest.

The predictors that turned out to be significant from the above analyses, i.e. pretest score and role, were selected for a MLR to sum up the results from the analyses above, to tell about the predictors' individual contributions, and thus to construct a model for the students' posttest scores (research question 4). Additionally, new predictors, based on the theoretical framework, were entered into the model. Several models had been tested in an attempt to estimate the values of the posttest scores by the simplest one, which shows the highest amount of variation in the posttest scores that is accounted for by the model.

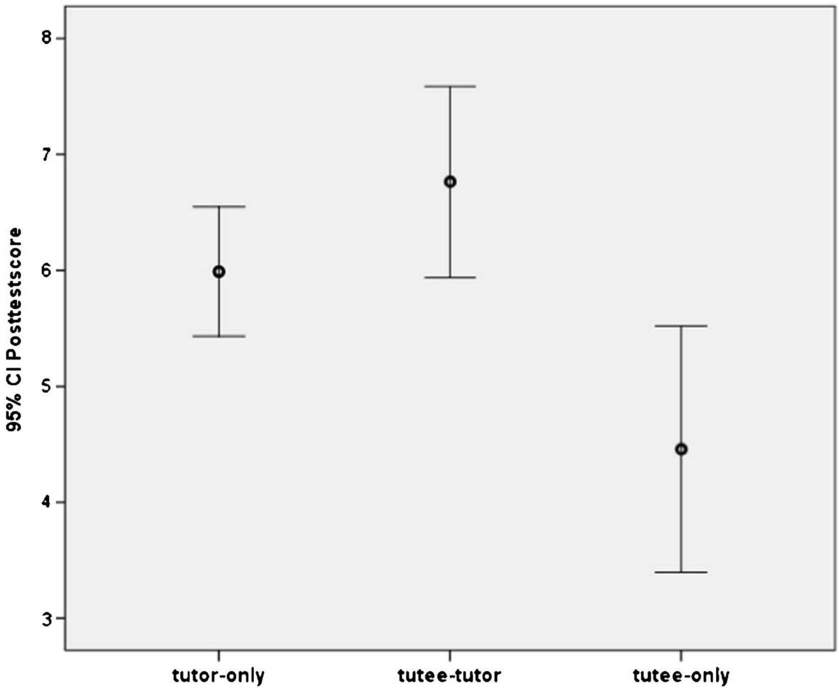

Fig. 3. Mean posttest scores and 95 % confidence interval according to the three roles. The *circles* indicate the mean, and the *bars*, the confidence interval

In order to get a first impression concerning the influence of the demographic variables, Pearson correlations with regard to posttest scores were tested (Table 3). The only correlation which is significant, besides the pretest score and the active role (represented by the variable tutor; coded with 1 for both tutor-only and tutee–tutor, coded with 0 for tutees), is the *first language* (coded with 1 if the instructional language (German) is the same as the first language, else 0): The correlation between first language and the posttest score is $r=.182$ with $p=.015$.

The correlation of posttest scores to gender showed that being a girl (coded with 1) means having a slightly, but not significantly, lower posttest score. The last grades correlate significantly with the posttest scores. The correlation is negative because in Austria “1” is the highest grade and “5” is the lowest one. There are, however, as demonstrated later, good reasons to exclude the last grade from MLR because frequency distributions show that 70 % of the students had grades “1” or “2” (the best ones). This indicates that this variable does not account much for the variance in the posttest. According to the results of correlation analyses, the variables were entered simultaneously into MLR

TABLE 3

Pearson correlations to posttest score

Variable	*r*
Pretest score	.426***
Tutor	.309***
First language	.182*
Gender	−.069
Last grade	−.151*

*$p<.01$; **$p<.005$; ***$p<.001$

by the method *forced entry*. Table 4 shows the summary of four different MLR models displaying constants and standardized coefficients beta.

Model 3 is the preferred one because it includes relevant parameters deduced from theory and it is the simplest model with the highest adjusted R^2. Also, it explains approximately 25 % of the variance. All included predictors are significant, and the statistical power of .34 indicates a rather strong effect (Buehner & Ziegler, 2009).

Model 4 includes the additional independent variables *gender* and *last grade* without increasing R^2 and thus explaining more of the effects of tutoring. The variable first language loses significance because it is highly significantly correlated to last grade ($r=-.493$; $p<.001$). Although model 4 is not the best, we mention it here because it shows that neither the last grade in physics nor gender has significant influence on the outcomes of cross-age peer tutoring. This point seems to be interesting with regard to further implementation of this method in regular classes.

TABLE 4

Summary of different MLR models

Model	*1*	*2*	*3*	*4*
Constant	3.808***	2.327**	1.428*	2.234**
Pretest score	0.426***	0.379***	0.360***	0.349***
Tutor		0.232***	0.263**	0.281***
First language			0.195**	0.141*
Gender				−0.048
Last grade				−0.103
N	142	142	142	142
Adjusted R^2	.175	.222	.254	.253
F stat	30.970***	21.104***	17.004***	10.573***
Effect size	0.21	0.29	0.34	0.34

*$p<.01$; **$p<.005$; ***$p<.001$

Discussion

The inferences made within this study are based on the idea that the sample is drawn from the same population. First of all, this assumption has to be made plausible because the available sample varied over a certain range of grade levels and first languages and it included different age gaps. Although it would have been better to have more precise conditions, field research has to be flexible enough to handle such a variety, if necessary. ANOVA shows clearly that inferences to the population can be drawn on the assumption of homogeneity in the tested parameters. For the content knowledge variable *pretest*, this might be a surprising result because, as reported above, one of the classes had had instruction in electricity based on the regular curriculum a year before the study. This class was expected to perform significantly better in the pretest, but it did not. This finding underpins the well-known results of research (Duit & Treagust, 2003) that render science teaching ineffective unless teachers explicitly focus on students' conceptions and thus conceptual change.

Statistical tests carried out in order to answer research question 1 show two things. The first is that there is an enhanced achievement in electricity for secondary-level students. This age group has not been covered in previous studies. Results from these studies describe positive effects of peer and cross-age peer tutoring mainly for younger or elder students. The present study closes this gap. With an effect size of 0.46, this effect, according to the standards in research, is estimated to be a medium effect (Cohen, 1988; Häußler, Bünder, Duit, Gräber, & Mayer, 1998). In our case, this is regarded as satisfactory considering the rather short duration of the intervention of approximately 1 h.

The second finding is that an enhanced achievement *in physics* is observable for this age group. Concerning the special situation in physics teaching, this result indicates that cross-age peer tutoring supports conceptual change and therefore leads to higher achievement in the posttest. Perhaps the detailed mechanisms of how this conceptual change takes place and what the role of cross-age peer tutoring within this process actually is could be investigated further by means of qualitative research.

The analyses above give an overview of the effects of tutoring but do not differentiate between the different roles students took part in. Therefore, further analyses were carried out in order to investigate the relation of outcome and role. According to the results of previous studies, it was recommended to focus on the tutors' outcomes (Fogarty & Wang, 1982; Robinson et al., 2005; Topping, 1996). With an effect size of 0.41, the effect of cross-age peer tutoring on the tutors' posttest scores (when compared to the pretest scores) is slightly smaller than the overall effect, but still satisfying. The reason for the smaller effect size might very likely be the smaller sample (n=92) since this analysis was carried out with students of the group tutor-only (not including the group tutee–tutor) in order to avoid the influence of possible dose effects. This result does not only confirm what has been reported about the tutors' outcomes in literature (Fogarty & Wang, 1982; Robinson et al., 2005; Topping, 1996), but also extends these results to physics instruction at the secondary level.

The knowledge gain for tutors as a result of the tutoring process seems interesting when planning further implementations of this method in regular classes and regular teaching. Considering the *beliefs* of the teachers who participated in this study, the knowledge gain was expected to be solely on the tutees' side; the tutors were supposed to be at a disadvantage because of the time needed for tutoring. Therefore, an open question was what to offer tutors in order to compensate them for their involvement. These teachers' beliefs can clearly be rebutted by our research, which shows a knowledge gain and thus a benefit for tutors as well. Their benefit lies in *learning by teaching* and thus improving their own abilities. Altogether, the active role seems to be crucial in the tutoring process. What exactly happens cannot be answered in detail with this research design and might be a topic for further investigation.

What has been left open until now is a more detailed analysis of where the differences in the posttests come from. Comparing the posttests of the three different groups (tutors, tutees, both), ANOVA is significant. However, ANOVAs were not significant when proving different factors that could have affected the posttests. Thus, there is a sample whose performance is homogenous in the pretests. All students showed improved performance in the posttest. The investigations of the posttest revealed that students' achievements differed highly significantly if the sample was split up by the factor *role*. This is underpinned by literature (e.g. Robinson et al., 2005) and by the analyses according to research question 2. All additional factors tested did not split the sample into highly significantly differing subgroups.

Further analyses were carried out by theory-based planned contrasts rather than by post hoc tests because there were already strong hints that the active role is the more powerful one. Figure 3 also indicates that the differences can be found between the active (tutor-only and tutee–tutor) and the passive role (tutee-only). Statistical computations revealed that this contrast, in fact, showed the most powerful statistics ($t_{49.8}$=3.320;

p=.002). Therefore, we infer that the posttests showed highly significant differences insofar that tutors and tutee–tutors in their active roles were more affected by the treatment than the passive tutees.

The active group itself contains two differently treated subgroups, the tutor-only group and the tutee–tutor group. They might differ because the tutor-only and tutee–tutor groups spent different times on the subject matter (compare Fig. 1 to Fig. 2). Considering Robinson et al. (2005), no such dose–effect relations are expected. In fact, t statistics were not significant in testing this contrast ($t_{162}=-1.461$; p=.146). This can be interpreted to mean that the effect of tutoring does not depend on its duration. This result may not be surprising since the same students' conceptions were addressed in nearly every mentoring or tutoring. Perhaps there would have been an effect if, together with a longer intervention, more conceptions had been addressed. This was not the case but may be of interest for further research.

Multiple linear regressions were carried out in order to estimate posttest scores as best as possible and to quantify the entered predictors' contribution. Based on correlation analyses, the parameters shown in Table 3 have been put in by the method "enter." Model number 3 shows a small constant (1.428). The constant gives the estimated posttest score independent of the pretest, the role, or the first language, and therefore, models with small constants are preferred. The additional gains can be explained by the predictors pretest, tutor, and first language. Comparing model 3 to model 4, the variable first language loses significance when the variable last grade is entered. Still, the last grade in physics seems to be a consequence of the first language and not the other way around. Therefore, first language is regarded as the more powerful predictor. The intrinsically interesting thing indicated by model 4 is that gender does not significantly influence the estimated posttest score. Considering the reported low interest of girls in physics, this result is in effect pleasing, though we should be careful with the interpretation because the boys were overrepresented in this sample.

Finally, MLR allows us to estimate the posttest score for different types of students (Table 5):

Case 2, for example, displays that the student's pretest score was 7 points, the student acted as tutor (tutor=1), and his first language was German (coded 1). The score of his posttest estimated by model 3 is 8.1. Whereas, case 6 describes a student with the same pretest score (7) and role (tutor=1), but the first language is not German, meaning it is not the language of instruction. For this case, the estimated score of the posttest is 7.2.

Conclusion

All in all, this study supports previously accepted outcomes from literature on peer and cross-age peer tutoring. Moreover, it closes the gap in research regarding the age level, as our study extends findings to the age group of secondary-level students. Another important point is that all tests carried out within this study show knowledge gains in physics. This is not a priori clear because teaching physics appears to be an especially demanding situation, which requires opportunities to facilitate a conceptual change. The results indicate that cross-age peer tutoring provides an appropriate learning environment to learn physics. Moreover, the findings reveal that the active role is the crucial one in the context of achievement. Considering the different sequences within the tutoring process, no effects on the duration on the outcome could be found.

Our MLR demonstrates that relevant parameters that explain the increase in achievement in electricity turned out to be the pretest, the role, and the first language. Based on these results, it will be interesting for further studies to compare classes instructed by the cross-age peer tutoring method to conventional ones.

Effects on attitudinal outcomes like motivation or the correlations of attitudinal and cognitive outcomes have not been covered by this work. The finding that there are no dose effects in the achievement does not necessarily mean that there are no effects in the attitudinal outcomes. This should be the subject of further analyses.

Another open question concerns the intrinsic mechanism of how the conceptual change takes place or what exactly happens during the tutoring process. This question is likely to be answered by additional qualitative research to support the quantitative findings. Another possible approach is to control the tutor–tutee assignments and accurately monitor

TABLE 5

Estimated posttest scores for different pretest scores, roles, and first languages

Case	*Pretest*	*Tutor*	*First language*	*Model 3*
1	1	0	1	3.4
2	7	1	1	8.1
3	7	0	1	5.7
4	1	1	1	5.8
5	1	1	0	5.0
6	7	1	0	7.2

the students' conceptions, which has not yet been done by the present research. Many questions concerning conceptual change cannot be answered sufficiently by this kind of research. There are hints that this method facilitates it, but it is left open what exactly supports the conceptual change and how it takes place within the tutoring process.

The overrepresentation of boys in the sample does not affect the research questions, but inference from MLR has to be done carefully. However, results indicate that gender has no significant influence on the outcomes of the peer tutoring process. In order to underpin these results, single-sex dyads in a sample where the ratio of boys to girls is approximately the same should be considered.

All in all, we regard this study as a contribution to the present research on learning environments in the sense of constructivist learning theories that support learning in science. Still, there is much work left to identify the mechanisms of how exactly it works.

ACKNOWLEDGMENTS

This study was funded by *Sparkling Science* (number of project deleted due to review process)—a program of the Austrian Federal Ministry of Science and Research (BMWF).

NOTES

[1] The correlations between the age gap, pretest, posttest, and first language were not significant and above .17, matching the results from multiple linear regression analyses: standardized betas for the variable *age gap* were below .1 and not significant.

[2] For an example task, see also the Electronic Supplementary Material.

[3] Theoretically, these analyses could have been replaced by a repeated-measures ANOVA with test time as within-subject factor and role as between-subject factor. The main effects are test time and role. Rm ANOVA gives similar results, but unfortunately, the interpretation of the interaction effects is problematic because the main effects cannot be clearly interpreted. The rank order of the roles with respect to the test time is not retained. Therefore, in this study, we abandoned this kind of data analysis and replaced it by the more flexible MLR analysis.

REFERENCES

Breit, S. (2009). Sozialisationsbedingungen von Schülerinnen und Schülern mit Migrationshintergund. In C. Schreiner & U. Schwantner (Eds.), *PISA 2006* (pp. 136–145). Graz: Leycam.

Buehner, M. & Ziegler, M. (2009). *Statistik fuer Psychologen und Sozialwissenschaftler*. Muenchen: Pearson.

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

Cohen, P. A., Kulik, J. A. & Kulik, C. L. C. (1982). Educational outcomes of tutoring—A meta-analysis of findings. *American Educational Research Journal, 19*(2), 237–248.

Duit, R. (2009). Bibliography—students' alternative frameworks and science education. Retrieved 13/08/2012, 2012, from http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html.

Duit, R., & Rhöneck, C. (1998). Learning and understanding key concepts of electricity. *Connecting research in physics education with teacher education*, 55-62.

Duit, R. & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*(6), 671–688.

Duit, R., Treagust, D. F. & Widodo, A. (2008). *Teaching science for conceptual change: Theory and practice*. New York: Taylor & Francis.

Engelhardt, P. V. & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics, 72*, 98–115.

Fogarty, J. L. & Wang, M. C. (1982). An investigation of the cross-age peer tutoring process: Some implications for instructional design and motivation. *The Elementary School Journal, 82*(5), 451–469.

Gaustad, J. (1993). *Peer and cross-age tutoring*. Oregon: ERIC Clearinghouse on Educational Management Eugene.

Hattie, J. A. C. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London, New York: Routledge.

Häußler, P., Bünder, W., Duit, R., Gräber, W. & Mayer, J. (1998). *Naturwissenschaftsdidaktische Forschung: Perspektiven für die Unterrichtspraxis*. Kiel: Institut für die Pädagogik der Naturwissenschaften.

Howe, C., Tolmie, A., Greer, K. & Mackenzie, M. (1995). Peer collaboration and conceptual growth in physics: Task influences on children's understanding of heating and cooling. *Cognition and Instruction, 13*(4), 483–503.

Lumpe, A. T. & Staver, J. R. (1995). Peer collaboration and concept development: Learning about photosynthesis. *Journal of Research in Science Teaching, 32*(1), 71–98.

Martin, J. P. (1998). Das Projekt: "Lernen durch Lehren"–fachdidaktische Forschung im Spannungsfeld von Theorie und selbsterlebter Praxis. In M. Liedtke (Eds.), Gymnasium - Neue Formen des Unterrichts und der Erziehungpp. 151-166).

OECD (Producer). (2010, 2012-09-05) PISA 2009 Ergebnisse: Zusammenfassung.

Reusser, K. (2001). *Unterricht zwischen Wissensvermittlung und Lernen lernen*. Donauwörth: Auer.

Rhöneck, C. (1986). Vorstellungen vom elektrischen Stromkreis. *Naturwissenschaften im Unterricht Physik/Chemie, 34*(13), 10–14.

Robinson, D. R., Schofield, J. W. & Steers-Wentzell, K. L. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17*(4), 327–362.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W. & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95*(2), 240–257.

Shaffer, P. S. & McDermott, L. C. (1992). Research as a guide for curriculum development: An example from introductory electricity. Part II: Design of instructional strategies. *American Journal of Physics, 60*(11), 994–1013.

Shipstone, D. M. (1984). A study of children's understanding of electricity in simple DC circuits. *European Journal of Science Education, 6*(2), 185–198.

Stork, E., & Wiesner, H. (1981). Schülervorstellungen zur Elektrizitätslehre und Sachunterricht. *Sachunterricht und Mathematik in der Primarstufe, 9*(6).

Strike, K. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice*, 147-176.

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education, 32*(3), 321–345.

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631–645.

Topping, K. J. & Bryce, A. (2004). Cross-age peer tutoring of reading and thinking: Influence on thinking skills. *Educational Psychology, 24*(5), 595–621.

Treagust, D. F. (Ed.). (2007). *General instructional methods and strategies*. New York: Taylor and Francis.

Urban-Woldron, H., & Hopf, M. (2012). Testinsrument zum Verständnis in der Elektrizitätslehre. *Zeitschrift für Didaktik der Naturwissenschaften, Jg, 18*, 229–259.

Vygotsky, L. S. (1978). *Mind in society*. London: Harvard University Press.

Wandersee, J. H., Mintzes, J. J. & Novak, J. D. (1994). *Research on alternative conceptions in science*. New York: Macmillan.

White, R., & Gunstone, R. (1992). *Probing understanding*. London, New York: RoutledgeFalmer.

Zinn, B. (2008). *Physik lernen, um Physik zu lehren*. Kassel: Universität Kassel.

Zinn, B. (2009). Ergebnisse einer Pilotuntersuchung zur Unterrichtsmethode "Lernen durch Lehren". *Zeitschrift für Didaktik der Naturwissenschaften, Jg, 15*, 325–329.

Marianne Korner
Austrian Educational Competence Centre Physics
University of Vienna
Porzellangasse 4/2/203, A-1090, Vienna, Austria
E-mail: marianne.korner@univie.ac.at

Martin Hopf
Austrian Educational Competence Centre Physics
University of Vienna
Porzellangasse 4/2/206, A-1090, Vienna, Austria