Cooperative learning in third graders’ jigsaw groups for mathematics and science with and without questioning training

Elmar Souvignier* and Julia Kronenberger
University of Frankfurt, Germany

Background. There is much support for using cooperative methods, since important instructional aspects, such as elaboration of new information, can easily be realized by methods like ‘jigsaw’. However, the impact of providing students with additional help like a questioning training and potential limitations of the method concerning the (minimum) age of the students have rarely been investigated.

Aims. The study investigated the effects of cooperative methods at elementary school level. Three conditions of instruction were compared: jigsaw, jigsaw with a supplementary questioning training and teacher-guided instruction.

Sample. Nine third grade classes from three schools with 208 students participated in the study. In each school, all the three instructional conditions were realized in three different classes.

Methods. All classes studied three units on geometry and one unit on astronomy using the assigned instructional method. Each learning unit comprised six lessons. For each unit, an achievement test was administered as pre-test, post-test and delayed test.

Results. In the math units, no differences between the three conditions could be detected. In the astronomy unit, students benefited more from teacher-guided instruction. Differential analyses revealed that ‘experts’ learned more than students in teacher-guided instruction, whereas ‘novices’ were outperformed by the students in the control classes.

Conclusions. Even third graders used the jigsaw method with satisfactory learning results. The modest impact of the questioning training and the low learning gains of the cooperative classes in the astronomy unit as well as high discrepancies between learning outcomes of experts and novices show that explicit instruction of explaining skills in combination with well-structured material are key issues in using the jigsaw method with younger students.

* Correspondence should be addressed to Dr Elmar Souvignier, University of Frankfurt/M., Institute of Educational Psychology, Senckenberganlage 15, D-60325 Frankfurt am Main, Germany (e-mail: souvignier@paed.psych.uni-frankfurt.de).

DOI:10.1348/000709906X173297
There is broad agreement that cooperative learning methods should be promoted in teaching mathematics and science (Lampert, 1990; McCormick & Pressley, 1997; van Boxtel, van der Linden, & Kanselaar, 2000; Yackel, Cobb, & Wood, 1991). Learning activities like restructuring problems, integrating different points of view, giving explanations and analysing misconceptions, collaborative practice, high intensity of student activities and controversial discussions, which are typically used when engaging in cooperative methods, can be labelled as cognitive elaboration. It has been frequently argued that elaborative processes of integrating new information actively into one’s prior knowledge might be a fundamental explanation for the effectiveness of cooperative learning methods (Meloth & Deering, 1994; Slavin, Hurley, & Chamberlain, 2003; Webb, 1989).

Even if cooperative methods have shown to be effective for elementary students in general (Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003), Johnson (2003) states that less than 12% of the 754 studies analysed in his review had been conducted with students younger than 10 years. In this line, Slavin et al. (2003) point out that future research on effects and processes of cooperative learning in younger students should be expanded. Rosenshine, Meister, and Chapman (1996) note within their review on the effectiveness of the use of questioning strategies that too few attempts have been made to evaluate the effectiveness of this elaborative strategy in younger students.

Evaluating the cooperative method of Group Investigation in eleventh graders, Shachar and Fischer (2004) found that most of the students achieved higher scores, while motivation declined. As an explanation for this ambiguous result, the statement of one student with respect to the implementation of the new cooperative method seems to be very illustrative: ‘You have to start when students are young, now no one wants to make the effort . . .’ (Shachar & Fischer, 2004, p. 84). This line of arguments leads to the question, at what age students should start using cooperative methods. The answer seems easy: from the first day in school, cooperative learning methods have been shown to be feasible and effective (Stevens & Slavin, 1995). There is evidence that methods like peer tutoring, collaborative problem solving and especially the application of cooperative methods in order to review learning units foster achievement as well as social interaction in elementary schools (Fawcett & Garton, 2005; Fuchs, Fuchs, Kazdan, & Allen, 1999; Rohrbeck et al., 2003). However, most of the studies which have been evaluated with elementary school students make use of well-structured learning scripts or focus on review rather than on elaboration. Taking this as a starting-point, the aim of our study is to analyse whether elementary school students, who have only little prior educational experience in being responsible for their learning behaviour, are able to make use of more challenging cooperative methods. Therefore, we decided to implement the jigsaw method into third grade classrooms.

The jigsaw method (Aronson, Blaney, Stephan, Sikes, & Snapp, 1978) provides a cooperative learning environment which fosters learner activity, joint acquisition of content and mutual explaining. Students in a class are divided into teams, the so-called ‘home groups’. The teacher gives a short introduction of the subject matter and explains how it will be divided into subtopics. Each member of a home group chooses one particular subtopic. Now those students who have chosen the same subtopic meet in ‘expert groups’ in order to study the material and prepare to teach it to their home groups. As ‘experts’, students return to their home groups and teach their team mates, the ‘novices’, in their respective subtopics. Finally, all students are tested individually on the complete material that was covered. The external structure of jigsaw ensures that learners acquire knowledge in an autonomous and self-regulated way and produce
explanations for each other. This challenging character of the jigsaw method leads to the recommendation that jigsaw should be applied carefully with very young children (Aronson & Patnoe, 1997). As Slavin (1995) points out, a potential problem of the method is that learners may master their own subtopics, but nonetheless fail to teach it to their team mates in an adequate manner. As a consequence, students would show deficits in the subtopics they did not work on in their expert groups but which were taught to them by other experts. This assumption is in line with results presented by Rohrbeck et al. (2003). The authors found that the effectiveness of peer-assisted learning in elementary schools depends on the students’ tutoring competence. However, Borsch, Jürgen-Lohmann, and Giesen (2002) showed in an implementation of jigsaw in third grade classes that novices performed more poorly than experts but better than students in control classes with teacher-guided instruction.

As mentioned before, one explanation which is frequently discussed for the effectiveness of cooperative methods has been labelled as ‘cognitive elaboration perspective’ (Slavin et al., 2003, p. 183). This perspective suggests that elaboration processes, such as the rephrasing of information, the integration within existing prior knowledge and the enrichment with examples, were major mediators of learning in cooperative settings. As a result of a review on 19 studies on verbal interaction in small groups, Webb (1989) states that effectiveness of learning in groups depends on the level of elaboration of the explanations given. In particular, processes of mutual explaining and questioning are regarded as effective ways of elaboration. At this point, it should be mentioned that even older students usually do not interact spontaneously on high levels of elaboration (Graesser & Person, 1994; King, 1999; van Boxtel et al., 2000). In line with this, Cohen (1994) states ‘if students are not taught differently, they tend to operate on the most concrete level. If teachers want high level operation, particularly verbal, the students will require specific development of skills for discourse’ (p. 7). Scripted Cooperation (Dansereau, 1988), Structured Oral Discussion (Yager, Johnson, & Johnson, 1985) and Guided Peer Questioning (King, 1994) are examples for instructional methods which explicitly support the elaboration of content. Cohen summarizes that ‘instructions that foster maximum interaction, mutual exchange, and elaborated discussions’ (p. 20) should be realized, when the teaching objective is learning for understanding. Guided Peer Questioning meets this requirement. This method provides learners with question stems they should use in group discussions. King (1994) differentiates between comprehension questions and connection questions. Comprehension questions ask for single facts or ideas in the material (‘What does . . . mean?’), whereas connection questions are more integrative and induce the learners to connect several facts or ideas (‘What is the difference between . . . and . . . ?’, ‘What are the strengths and weaknesses of . . . ?’). King (1994) describes the question stems as ‘thought provoking’ (p. 340) and therefore well suited to support elaboration of the learning material. Rosenshine et al. (1996) review 26 intervention studies evaluating the instruction of questioning strategies. They come to the conclusion that questioning strategies can be taught effectively and that they can positively influence the comprehension of learning material. The authors present an average effect size of $d = 1.12$ for questioning stems as ‘procedural prompts’. Against the background of these results, it is plausible to assume that the implementation of question stems could support elaboration in cooperative learning environments like jigsaw as well. To our knowledge, no studies with younger children than fourth graders have been conducted yet, using question stems. In their review of studies dealing with the effects of question generation, Rosenshine et al. explicitly encourage researchers to examine effects of this
approach in younger children. Our study aims to investigate whether third graders engaging in jigsaw benefit from instructional advice to make use of elaborative strategies.

**Research questions**
The goal of our study was to examine whether and under which conditions a challenging cooperative method like jigsaw can be applied in elementary schools. We expected that achievement in mathematics and science can generally be enhanced by cooperative learning. Nonetheless, the jigsaw method may promote only the learning of experts, while the lack of spontaneous elaboration may preclude these benefits from being extended to those in the novice role. As explicit teaching of how to ask questions and how to provide explanations to others proved to enhance elaborative thinking, an additional questioning training was implemented. Hence, achievement of experts and novices was analysed in two jigsaw conditions, a standard jigsaw group and a jigsaw group with the additional questioning training. Students receiving teacher-guided instruction were used as a control comparison.

**Method**

**Design**
Three classes learned in the condition of standard jigsaw, three classes in the condition of jigsaw with additional questioning training and three classes in the control condition of teacher-guided instruction. The entire investigation consists of four studies applying different learning units that can be interpreted as replications. The first three units deal with mathematics, the fourth unit with science. Each unit consists of six school lessons. Before the start of the first unit, tests were administered to control for differences in learning preconditions. Prior to and at the end of the six lessons of each unit, achievement tests were administered. To reduce testing frequency, delayed tests of two successive units were run together 2–3 months after the units. As Figure 1 shows, there were intervals of about 1 month between two units. Under the condition of jigsaw + questioning, the first two units started with the questioning training.

**Figure 1. Design.**
Participants
Nine third grade classes from three elementary schools in Frankfurt/Main, Germany, with a total of 208 students participated in the study. In each school, all the three different instructional conditions were realized in order to control school effects. Classes were assigned randomly to the three conditions with two exceptions: in two schools, teachers had chosen to teach cooperatively or teacher guided, respectively, before the start of the study. The mean age of the students was 8;11 years (SD = 0.56). As Table 1 shows, the three groups were comparable in age, proportion of girls and boys and language spoken at home. Demographic information indicates that 25% of the students reported that they do not speak German in their families at all and 38% say they speak German and another language in their families. These numbers are quite representative for schools in a large city like Frankfurt.

Table 1. Sample description

<table>
<thead>
<tr>
<th>Instruction</th>
<th>N</th>
<th>Girls/boys</th>
<th>Age</th>
<th>German</th>
<th>German &amp; other</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher guided</td>
<td>70</td>
<td>37/33</td>
<td>8;10 (0;53)</td>
<td>23</td>
<td>26</td>
<td>21</td>
</tr>
<tr>
<td>Jigsaw</td>
<td>67</td>
<td>34/33</td>
<td>9;0 (0;59)</td>
<td>22</td>
<td>30</td>
<td>15</td>
</tr>
<tr>
<td>Jigsaw + questioning</td>
<td>71</td>
<td>37/34</td>
<td>9;0 (0;55)</td>
<td>32</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>208</td>
<td>108/100</td>
<td>8;11 (0;56)</td>
<td>77</td>
<td>79</td>
<td>52</td>
</tr>
</tbody>
</table>

As Table 2 shows, all the three groups were also comparable according to the preconditions of learning (one-way ANOVAs).

Table 2. Means and standard deviations of preconditions for learning

<table>
<thead>
<tr>
<th></th>
<th>Vocabulary</th>
<th>Reading comprehension</th>
<th>Mathematics</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher guided</td>
<td>11.87 (3.65)</td>
<td>13.47 (3.59)</td>
<td>21.16 (7.37)</td>
<td>8.06 (3.01)</td>
</tr>
<tr>
<td>Jigsaw</td>
<td>11.39 (4.07)</td>
<td>13.00 (3.65)</td>
<td>21.08 (8.29)</td>
<td>7.58 (3.90)</td>
</tr>
<tr>
<td>Jigsaw + questioning</td>
<td>11.56 (4.05)</td>
<td>12.29 (4.17)</td>
<td>21.68 (7.33)</td>
<td>8.24 (3.86)</td>
</tr>
<tr>
<td>One-way ANOVA</td>
<td>F(2, 205) = 0.27</td>
<td>F(2, 205) = 1.71</td>
<td>F(2, 205) = 0.13</td>
<td>F(2, 205) = 0.61</td>
</tr>
</tbody>
</table>

Learning units and instructional conditions
It was decided to choose mathematics as the main subject for the study. Within the content of geometry, which is part of the usual curriculum for third graders, students could be given the opportunity to deal with ‘hands-on’ material. However, the instructional material gave students a clear structure in order to guide the process of mutual teaching. The three geometry units covered the contents of (1) geometric solids, (2) symmetry and (3) topology. For science, one unit about astronomy was prepared. This unit offered more opportunities for exploring the material that was provided and steps of teaching were not specified as much as in the geometry units. According to the requirements of jigsaw, each unit was divided into five separate subtopics (Table 3). Varying subject and structure between units 1–3 and 4 offered the chance of exploring whether jigsaw would bring about comparable effects under different conditions. In general, we wanted to find out, whether findings will be replicated over the three comparable geometry units. The goal of this replication over three units was to enhance
the external validity of the study. Since not more than four units could have been
realized within our project, we decided to accept the imbalance of three vs. one unit and
use the fourth unit to explore whether transfer of cooperative competence will occur.

**Table 3. The four learning units and their subtopics**

<table>
<thead>
<tr>
<th>Mathematics</th>
<th>Science</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric solids</td>
<td>Symmetry</td>
</tr>
<tr>
<td>Cuboid</td>
<td>Mirror</td>
</tr>
<tr>
<td>Cube</td>
<td>Laying</td>
</tr>
<tr>
<td>Pyramid</td>
<td>Folding and cutting</td>
</tr>
<tr>
<td>Cone</td>
<td>Shifting</td>
</tr>
<tr>
<td>Cylinder and sphere</td>
<td>Turning</td>
</tr>
</tbody>
</table>

All four subject matters are part of the usual curriculum for third graders. Learning
units were developed by undergraduate university students as part of their senior paper
which they are required to prepare for their teacher exams in the last year of studies in
cooperation with the researchers. Orientation for the contents and design of the units
was taken from usual textbooks for third graders. Each unit covered six lessons of 45
minutes for teaching and two additional lessons for achievement tests. While the
astronomy unit had been used during a previous study (Borsch et al., 2002), the math
units were newly developed. In several pre-studies, all units were administered to at
least two classes. Learning processes in these classrooms were observed and the units
were revised in order to meet the demands of everyday teaching situations. To optimize
the revision process, experienced teachers of the pre-studies classes were asked for
their advice. Teachers in the main study had the opportunity to read the materials
carefully before the beginning of the study. They all judged the materials in sum as good
and adequate for third grade. All classes obtained the same learning materials.

The teachers in all nine classes were supported by student assistants who took care
of all organizational aspects, administered the tests and engaged in classroom
observations (see below). For each class, one student assistant was responsible from
the beginning until the end of the study. The assistants were students preparing for teacher
exams and students preparing for psychology diploma in their last study year. All of
them were familiar with basic literature about cooperative and teacher-guided learning
and with the learning material (this was assured by assistant training sessions guided by
the researchers). Most of the assistants used the experiences in this study to write their
exam thesis. Additionally, one psychology student was responsible for the questioning
training. She was involved in the development of the questioning training from the
beginning and was familiar with the relevant literature.

Under the condition of teacher-guided instruction, a booklet with the tasks of the
home groups and all visual aids were given to the teachers. In the three teacher-guided
classes, a time sample of the lessons was observed systematically to assess the
instructional methods used. These observations revealed that about 75% of the
instructional time was used for teacher presentations, classroom discussions and silent
seat work. Only 4% was used for group work. The rest of the time was used for student
presentation, instructional games and phases of classroom organization processes.
Hence, the teacher-guided (control) condition mirrors a rather typical way of this kind
of instruction.
In order to reach high experimental control, two general rules were given in the jigsaw conditions: (1) Instructional time had to be divided into two periods of three lessons each. During the first three lessons, expert groups worked on their materials and in the following three lessons, students taught each other in their home groups. (2) Students were given no help concerning the learning contents. Only organizational aids were allowed to be given by the teachers. The student assistants took care for the compliance with these two rules. Deviations from the rules were reported to be rare by the student assistants. Teachers generally met the planned time schedules. Interventions in the group process remained at a minimum and were restricted to groups with severe problems in understanding.

Children in all conditions knew from the beginning that each learning unit would end with an individual achievement test. In the jigsaw conditions, teachers discussed with their classes several times the positive interdependence emerging from the fact that students would have to answer questions about the expert topics of their team colleagues.

Students in the ‘jigsaw + questioning’ condition were taught to use five question stems from King’s Guided Peer Questioning (1994): (a) ‘What does . . . mean?’, (b) ‘Explain why . . . !’, (c) ‘Explain how . . . !’, (d) ‘How are . . . and . . . similar?’ and (e) ‘What is the difference between . . . and . . . ?’. Since it is very demanding for third graders to internalize the questioning routines, two lessons of practice were carried out before starting the first jigsaw unit. An additional training lesson in questioning was conducted at the beginning of the second jigsaw unit. During the first lesson, the sequence of listening–asking–responding was established, the five question stems and their functions were explained to the students, and the application was practiced as part of a game. In this game, students had to draw cards with one question stem and to ask appropriate questions with regard to a given text. The game was repeated during the second lesson. Further on, cards (for each table) with the question stems were prepared to support the use of the questions, work-sheets contained prompts to ask questions, and at the beginning of each lesson, students were reminded of the question stems.

**Measures**

Before the start of the first unit, four tests were administered to control for differences in learning preconditions. Since the independent work under cooperative conditions requires a lot of reading and explaining, achievement in vocabulary and reading comprehension (AST 2, Rieder, 1991) was tested. These two subtests consist of 20 items each in a multiple choice format and are constructed for the last 3 months of second grade. General math achievement examined by the DEMAT 2+ (Krajewski, Liehm, & Schneider, 2004) and general science achievement examined by the subtest ‘science’ from the AST 3 (Fippinger, 1991) served as measures for domain-specific knowledge. The DEMAT 2+ consists of nine subtests for different mathematical areas with different answering formats and is constructed for the first 8 weeks of third grade. The subtest ‘science’ from the AST 3 contains 20 items in a multiple choice format and is constructed for the second half of third grade. All tests of learning preconditions have satisfying internal consistencies in the sample of the study (Cronbach’s α: .70–.90).

For each learning unit, an achievement test was designed which represented the unit’s main contents. Most answers had to be given in an open format requiring to produce a term, a number, a short answer, a drawing or an arithmetical operation.
The tests consisted of 10 to 20 items with an equal number of questions about each subtopic. The tests included items asking for basic knowledge (e.g. ‘What is the name of this geometric solid?’ ‘Which perspectives do you know?’) as well as items asking for deeper understanding (e.g. ‘Do you know one difference and one similarity between a cone and a cylinder?’ ‘Draw a sketch of a lunar eclipse!’) and for procedural skills (e.g. ‘Draw the symmetric axis in this figure!’ ‘Draw this body as if you were looking at it from above!’). The tests were given as pre- and post-tests prior to and at the end of the learning units. Short forms of these tests, using a subsample of the items, were used as delayed tests. Since we assumed that the students would have forgotten the more specific information after 2 or 3 months, we chose the easier items for the delayed tests. Points were given according to a standardized scoring schema. Inter-rater correlations of $r = .98$ between two different raters revealed high agreement. Maximum possible scores of the four achievement tests were 20.5 (math 1 and math 2), 31 (math 3) and 38 (astronomy), and, respectively, 9.5 (math 1), 13 (math 2), 18 (math 3) and 12 (astronomy) for the delayed tests. Internal consistencies (Cronbach’s $\alpha$) were acceptable for the complete tests (math 1, $\alpha = .79$; math 2, $\alpha = .78$; math 3, $\alpha = .70$; astronomy, $\alpha = .83$), and were returned somewhat lower for the shorter delayed versions (math 1, $\alpha = .69$; math 2, $\alpha = .76$; math 3, $\alpha = .53$; astronomy, $\alpha = .56$).

Classroom observations in the cooperative condition

In the cooperative learning classes, several expert groups and home groups were videotaped. A sample of 56 groups was analysed (10 minutes each). Videotaping took place in the units math 1, math 3 and astronomy. Sampling was balanced over phases (expert vs. home groups), classes and units. In the group discussions, questions on four quality levels were identified and counted (unspecific questions, short answer comprehension questions, long answer comprehension questions and connection questions) and as an additional category helping questions (‘didactic questions’ posed by a child who apparently knows the answer). Furthermore, the group discussions were rated for level of explanation (no content discussion, answers without explanations, partial explanations and full explanations). For five groups (questions), respectively ten groups (level of explanation) discussions were rated by two independent raters with a satisfying inter-rater reliability (questions, Cohen’s Kappa $\kappa = .71$; level of explanation, Spearman correlation $\rho = .75$). The same two raters judged the rest of the sample. Methods of observation and analysis as well as results are thoroughly discussed in Kronenberger and Souvignier (2005).

Results

As Table 4 shows, there are large gains from pre- to post-test in all groups and all units. $t$-tests on dependent samples reveal that students improved significantly under all conditions and in all learning units (all tests $p < .001$). While preconditions for learning proved to be similar for the three groups, all four pre-tests yield a superiority of the teacher-guided group compared with the two jigsaw conditions taken together (math 1, $t(206) = 3.44, p < .01$; math 2, $t(206) = 4.50, p < .01$; math 3, $t(206) = 3.13, p < .01$; astronomy, $t(206) = 1.66, p < .10$).
Group comparisons of learning gains

Due to differences in the unit-specific knowledge prior to instruction, the analyses of raw post-test and delayed test scores would not have been satisfactory. The use of gain scores as measure for the learning amount ascertains fair comparisons. In spite of some argumentation against the reliability and fairness of gain scores (Cronbach & Furby, 1970), Rogosa (1988) states that gain scores are reliable and fair if there are real differences between individuals in changes which can be assumed in the context of a learning intervention (see also Hager, 1995). For the analysis of the dependent measures, gain scores between post-test and pre-test of a unit and between delayed test and pre-test of a unit were used. To obtain gain scores for the post-tests, the entire pre-tests were used, whereas to obtain gain scores for the delayed tests shorter versions of the pre-tests were used, containing the same items as the delayed tests themselves.

Besides the unit-specific prior knowledge, a broader domain-specific prior knowledge and general verbal abilities could influence the learning outcome of the instructional units. To reflect this, general knowledge in mathematics or science, respectively as well as vocabulary and reading comprehension were used as covariates in the analyses of learning amount.

Due to the use of different tests in each learning unit and due to the short versions in the delayed tests, gains cannot be compared between the four units and between immediate and delayed achievement. Therefore, separate analyses for the units and for immediate and delayed achievement are conducted. To analyse differences in the amount of learning between the three instructional conditions, gain scores from pre- to post-test and from pre-test to the delayed test were compared for each unit by one-factorial ANCOVAs (with instructional condition as factor) controlling for vocabulary, reading comprehension and general knowledge in mathematics or in science, respectively, as covariates.

No differences in the immediate amounts of learning could be detected between the three groups in any of the math units. As Figure 2 shows, gains from pre- to post-test appeared to be rather similar (math 1, $F(2, 202) = 0.95$; math 2, $F(2, 202) = 1.54$; math 3, $F(2, 202) = 2.62$). With the exception of math achievement in the second unit ($F(1, 202) = 9.43; p < .01$), preconditions for learning had no significant impact on learning.1

Table 4. Means and standard deviations of pre- and post-tests

<table>
<thead>
<tr>
<th></th>
<th>Teacher guided</th>
<th>Jigsaw</th>
<th>Jigsaw + questioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math unit 1</td>
<td>Pre 7.79 (3.16)</td>
<td>6.24 (3.17)</td>
<td>6.14 (3.18)</td>
</tr>
<tr>
<td>Geometrical shapes</td>
<td>Post 13.85 (3.34)</td>
<td>12.80 (3.26)</td>
<td>12.88 (3.63)</td>
</tr>
<tr>
<td>Math unit 2</td>
<td>Pre 4.94 (2.30)</td>
<td>4.30 (2.27)</td>
<td>2.57 (2.10)</td>
</tr>
<tr>
<td>Symmetry</td>
<td>Post 14.20 (4.32)</td>
<td>12.42 (3.68)</td>
<td>10.87 (4.20)</td>
</tr>
<tr>
<td>Math unit 3</td>
<td>Pre 11.44 (5.94)</td>
<td>9.10 (5.73)</td>
<td>8.68 (5.10)</td>
</tr>
<tr>
<td>Topology</td>
<td>Post 20.23 (5.88)</td>
<td>16.68 (6.31)</td>
<td>18.32 (5.97)</td>
</tr>
<tr>
<td>Science unit on</td>
<td>Pre 8.58 (3.65)</td>
<td>7.40 (3.85)</td>
<td>7.87 (4.09)</td>
</tr>
<tr>
<td>Astronomy</td>
<td>Post 20.81 (7.45)</td>
<td>16.73 (6.44)</td>
<td>16.46 (6.50)</td>
</tr>
</tbody>
</table>

1 Taking into account the differences on pre-test scores, we additionally analysed raw post-test scores adjusted for pre-test values and learning preconditions. In general, we found the same pattern of results as by comparing gain scores. However, effects of the ‘jigsaw without questioning training’ condition proved to be inferior in the second and the third math unit (math 2, $F(2, 201) = 8.01, p < .01$; math 3, $F(2, 201) = 4.81, p < .01$) taking this analytical strategy. Differences between the jigsaw + questioning condition and teacher-guided instruction remained negligible.
Contrary to our expectations, students benefited more from teacher-guided instruction in the astronomy unit ($F(2, 202) = 5.95; p < .01$). Learning gains in this unit proved to be enhanced somewhat by vocabulary ($F(1, 202) = 10.88; p < .01$). Effect sizes of $d = -0.28$ (jigsaw) and $d = -0.44$ (jigsaw + questioning) underline a modest advantage for the teacher-guided condition.²

As Figure 3 shows, the delayed tests return a rather comparable picture. Differences between the three groups appear to have somewhat increased in math 2, but the other three results remained about the same (math 1, $F(2, 202) = 0.37$; math 2, $F(2, 202) = 5.49; p < .01$; math 3, $F(2, 202) = 1.16$; astronomy, $F(2, 202) = 4.57; p < .05$). In math 2, the learning success was affected by general mathematical achievement ($F(1, 202) = 4.44; p < .05$). Besides that, none of the tested learning preconditions had significant impact on the long-term learning results. Table 5 shows the raw scores of the (reduced) pre-tests and delayed tests.

Comparisons between experts and novices
As pointed out in the introduction, one argument against the jigsaw method is that especially younger - students will only understand their own expert subtopic, because the quality of reciprocal teaching in the home groups will be too poor. In jigsaw, every child is an expert in one subtopic of the unit and a novice in all other subtopics. In order to analyse the differences between experts and novices, we divided all tests into parts according to the subtopics, obtaining five different short versions for each test. Since these short versions cannot be compared directly, we calculated effect sizes to ascertain the difference between experts and novices for each subtopic. In the same way, experts and students in the teacher-guided condition as well as novices and students in the teacher-guided condition were compared.

² Following the helpful suggestion of a reviewer, we added an analysis to check, how the variables which were included in the ANCOVA as covariates contributed to the improvement. Within a hierarchical regression, treatment was included in step 1 and cognitive preconditions as well as ethnic origin in step 2. In general, we found that improvement appeared to be quite independent from any of these predictors. Only in the second math unit, general math achievement (DEMAT 2 + ) turned out to be a significant precondition for gains in learning ($\beta = 0.23; p < .01$). In the science unit, we found that beyond differences between the treatment groups ($\beta = -0.20; p < .01$) - achievement was only influenced by vocabulary ($\beta = 0.33; p < .01$).
Table 6 reveals the means of the effect sizes. The first contrast examines the differences between experts and novices in the two jigsaw conditions. While in math experts show a small advantage as compared with the novices, differences between experts and novices appear to be rather large in the astronomy unit. The experts’ learning gains were even somewhat larger than those of the students in the teacher-guided condition. The novices only showed a small disadvantage as compared with the teacher guided condition in math, but in the science unit, they were outperformed by about half a standard deviation.

Classroom observations in the cooperative condition
Questions and explanations in the group discussions were analysed and serve as indicators for elaborative processes. Analyses in general reveal a low elaboration level. Children in both cooperative conditions and both jigsaw phases (home groups and expert groups) posed many questions (9–15 in 10 minutes), but few higher-level questions (1–2 long answer comprehension and connection questions in 10 minutes). In the home groups, 1–2 helping questions in 10 minutes were posed, nearly all of them by the experts. No group was rated with the category ‘no content discussion’, meaning that all groups discussed the learning material in some way. In 43% of the groups, no explanations were given (children stated only solutions without explaining them any further). In another 41% of the groups, partial explanations occurred. Full explanations
were observed in only 16% of the groups. Small positive effects for the questioning training concerning amount and quality of questions occur in the home groups (amount of questions, \( t(29) = 1.34, p < .10, d = 0.48 \); amount of high-quality long answer comprehension and connection questions, \( t(29) = 1.58, p < .10, d = 0.57 \)). Overall, the level of explanation was not optimized by the training.

**Table 6. Mean effect sizes for the comparison of experts, novices and teacher-guided learners**

<table>
<thead>
<tr>
<th></th>
<th>Experts–novices</th>
<th>Experts–teacher guided</th>
<th>Novices–teacher guided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jigsaw</td>
<td>0.18</td>
<td>0.08</td>
<td>-0.15</td>
</tr>
<tr>
<td>Science</td>
<td>0.79</td>
<td>0.48</td>
<td>-0.42</td>
</tr>
</tbody>
</table>

Discussion

In all classes, the children understood the procedure of the jigsaw method and worked independently on their group tasks most of the time. Looking for a general pattern in the results of the four learning units, at first glance only negligible differences between the three instructional approaches seem to exist. It is not surprising that the students in the jigsaw condition did not outperform those in the teacher-guided instruction. Meta-analyses (Johnson, Johnson, & Stanne, 2000; Slavin, 1995) show that achievement gains appear to be rather small for jigsaw – mean effect sizes range between \( d = 0.12 \) and 0.29. However, the study compared jigsaw with the control condition under very strict conditions. Except for a few cases in which the groups did not manage to master the specific material at all, teachers did not provide any help concerning the learning contents. (These rare cases were restricted to one or two groups out of four or five groups in a class and these few ‘problem groups’ were equally distributed between the conditions.) The ‘non-intervention rule’ was intended to prevent teachers from engaging in teacher-guided instruction with jigsaw groups in order to assure the comparability of the instruction in all jigsaw classes. It is likely that the decision to restrict the role of the teachers to providers of organizational and social aids generated a disadvantage for the jigsaw conditions which should be taken into account when interpreting the effects. Despite all difficulties, our study shows in sum that third graders are able to use the jigsaw method with comparable learning results as those achieved in teacher-guided classrooms. It is likely that the decision to restrict the role of the teachers to providers of organizational and social aids generated a disadvantage for the jigsaw conditions which should be taken into account when interpreting the effects. Despite all difficulties, our study shows in sum that third graders are able to use the jigsaw method with comparable learning results as those achieved in teacher-guided classrooms. Taking into account that cooperative learning in a team and presentation of one’s knowledge as an expert are valuable learning objectives in themselves, this seems to us to be a sufficient justification for the integration of jigsaw in the educational programme of a class even if additional training should be necessary.

Two other results need more clarification. Firstly, we did not anticipate the barely modest impact of the questioning training. Secondly, the low effects in the astronomy unit stand in clear contrast to earlier findings with the same learning materials where even the novices outperformed the teacher-guided group (Borsch et al., 2002).

The aim of the questioning training was to enhance the elaboration of the learning material which was expected to lead to better learning results. The question stems were supposed to induce a more intensive processing of the content, verify whether learners
understand the problems to be solved, and foster the production of detailed explanations. However, the analysis of the questioning and explaining behaviour shows that the questioning training had only marginal effects on these processes (Kronenberger & Souvignier, 2005). Observations during the training sessions revealed that most of the children were able to pose questions using the question stems, but the transfer to the jigsaw environment proved to be difficult. As the number of challenging long answer comprehension and connection questions in the group work remained rather small, the intended ‘thought-provoking’ effect of the questioning training could not be established. As an interesting finding from the classroom observations, a slightly differential pattern of effects of the questioning training between the learning processes in expert groups and home groups was found: working in the expert groups, students seemed to make only little use from the questioning stems, whereas during the phase of mutual teaching in their home groups, more higher level questions and even some ‘didactic questions’ were posed. However, results of the math units suggest that the questioning training had a small positive impact on experts’ learning results (see Table 6). Even if classroom observations show that the questioning training does not lead to more high-quality questions during the expert-group phase, we find a slight trend towards higher levels of explanation while acquiring expert knowledge in this group. Hence, a possible explanation for the higher merits of experts under the questioning condition might be that the question stems help to focus on conceptual knowledge while preparing collaboratively in the expert groups. King’s (1994) study shows that in principle, young children can profit from training in questioning strategies, but there are two main differences between her and our investigation. First, in King’s study, the question stems served to practice skills and deepen understanding of concepts which had been taught in teacher-guided instruction before, whereas in our study the entire acquisition process of the content took place in autonomous and self-regulated learning groups. Second, students in our investigation were 1–2 years younger than the fourth and fifth graders in King’s study. Third graders might need more than three training lessons to adopt the questioning strategy. In addition, it might have been more adequate to guide the students slowly from simple short answer questions to challenging comprehension and connection questions instead of beginning with questions on a rather complex level.

The disappointing learning results of the cooperative learning classes in the science unit correspond with higher differences between experts and novices in this unit. Interestingly, it is possible to establish a clear sequence of learning results. On the one hand, experts learn more in their subtopics than students in teacher-guided instruction. On the other hand, novices in the cooperative learning conditions learn less than children in the control classes. This pattern becomes particularly apparent in the science unit, whereas the differences between the three groups in the mathematics units are much smaller. Summarizing the results, it is safe to say that the acquisition of new content in the expert groups was effective in all learning units, whereas the mutual peer tutoring in the home groups proved to be more problematic, especially in the science unit.

Taking this into account, it is helpful to take a closer look at differences in the structure of the learning materials between the units. Whereas the tutoring process in the home groups was prepared step by step in the mathematics units, the home group material in the science unit lacked such a clear structure. In mathematics, each child was supposed to work out concrete problems for all subtopics, while the experts helped and guided the problem-solving process. In the science unit, however, the
experts had to explain the content in an open form. A possible explanation for the lower learning results in the science unit could therefore be that the children, once used to the clear structure of the mathematics jigsaw materials, were not able to manage the more open and therefore more challenging process of peer tutoring in the science unit. Since third graders dealt very successfully with this challenge in a previous study (Borsch et al., 2002), the alternation, particularly, between cooperative arrangements with varying levels of structure might have constituted a considerable problem for the young students.

The social structure of the school classes in our study could imply an additional problem that might have been responsible for some of the disappointing results: a high percentage of the students do not come from native German-speaking backgrounds. On the one hand, these students perhaps needed more clarification of texts and learning tasks, which is normally given by the teacher and was missing in the jigsaw learning environment. On the other hand, maybe the questioning training was too demanding for students who speak German as a second language. Students who only speak German at home were in fact significantly better in vocabulary and reading comprehension (AST 2) than students speaking other languages. Nonetheless, there was no interaction between these language skills and the condition of instruction. In all the three learning environments, students with higher language skills performed better than students with lower language skills. So, we cannot conclude that the jigsaw method or the questioning training put the non native German-speaking children in a special disadvantage.

In our study, jigsaw and questioning training were implemented in third grade classes, hence the children were younger than in most studies concerning comparable interventions which implied the risk that they would not yet be able to cope with a complex and demanding learning environment (c.f. Aronson & Patnoe, 1997). Nevertheless, we decided to work with these young children because we wanted to introduce cooperative learning methods as soon as possible in order to make them a natural part of the school day. This is in line with the claim of Slavin et al. (2003) for the evaluation of cooperative methods for younger children, which is supported by the general finding of positive effects for using peer-assisted learning methods with elementary school students (Rohrbeck et al., 2003) and by promising results of teaching self-regulation skills in younger children (Schunk, 1998). Additionally, our decision to study third grade was supported by positive experiences with jigsaw in this grade in earlier studies (Borsch et al., 2002). Although we used some of the learning materials which had brought about high learning achievement for cooperative learning in an earlier study, the effects in the current study proved to be much smaller. As mentioned before, reasons for this different pattern of results might be the lower cognitive (especially verbal) competence of the current sample and the contrast between the different challenges inherent in the cooperative material.

The different learning results in the several classes show the particular importance of the peer-tutoring process in the home groups. Whereas low improvement scores between pre- and post-tests correspond with high discrepancies between experts and novices, successful cooperative classes are characterized by good learning results of novices. This pattern underlines that for an effective implementation of the jigsaw method, the design of the home group phase, in particular, is very important. Well-structured material or an explicit instruction of explaining skills – in parallel with the questioning training – seem to be promising approaches.

Which consequences can be drawn from these findings for the use of the jigsaw method with elementary schoolchildren? (1) It is possible and seems worth to
implement this demanding cooperative method in elementary schools’ classrooms. Even under the restriction that teachers did not give any support concerning the learning contents, the two jigsaw conditions showed gains comparable to those children receiving teacher-guided instruction. (2) As the Stevens and Slavin (1995) study shows, young children need time to familiarize themselves with cooperative learning methods. On the one hand, this is a strong argument to begin early by teaching with cooperative methods in order to utilize the high potential of these instructional approaches. On the other hand, children of this age need support to cope with the challenges of independently preparing new learning contents and mutually teaching these topics to their team mates. (3) Since young children still need to develop their peer-tutoring competence, explicit instruction on how newly compiled information can be structured to teach it to the members of the home groups seems to be mandatory. Therefore, the primary goal of the expert group phase is to make sure that the children feel well prepared to explicate their expert themes. (4) Questioning training is one way to support the quality of questions and the level of explanations given (King, 1994; Rosenshine et al., 1996). In fact, the question stems used in our study seemed somewhat too demanding for the third graders. More than three training sessions and easier to handle question stems probably would have been more appropriate. (5) Learning materials undoubtedly are a crucial factor to support cooperative learning activities. In the present study, students showed difficulties by changing from well structured to more open tasks. Beyond this, it seems necessary to make sure that the different expert themes share some common aspects: children need to have the opportunity to discover similarities between their expert themes and work on tasks that require an integration of different experts’ information. Preparing learning materials this way will encourage the experience of positive interdependence.

In sum, cooperative learning in younger children needs as well explicit (preparing experts as teachers; questioning and explanation training) as implicit (fostering interdependence by adequate learning materials) support. Results of the present study suggest that independent learning and instruction by using the jigsaw method can already be realized in elementary classrooms. Given the above outlined conditions, it seems very likely that cooperative learning in young children will lead to superior learning effects.

Acknowledgements
This research was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) (Gi 134/4-1)

References
Cooperative learning in jigsaw groups


*Received 6 January 2006; revised version received 3 November 2006*