Learning Disabilities: A Contemporary Journal 16(1), 19-35, 2018

Fostering the Computation Competence of Low Achievers Through Cooperative Learning in Inclusive Classrooms: A Longitudinal Study

Elisabeth Moser Opitz*
Urs Grob
University of Zurich; Switzerland

Claudia Wittich
TU Dortmund University, Germany

Uta Häsel-Weide
University of Paderborn, Germany

Marcus Nührenbörger
TU Dortmund University, Germany

Fostering peer interaction and shared learning is an important aim of inclusive instruction. However, it has not been established whether it is possible to offer explicit and intensive support for low achievers in inclusive settings. This longitudinal study examined whether a structured program that includes cooperative learning fosters computational competence and flexible strategy use by low achievers in mathematics in inclusive classrooms. 126 persistent counters from 35 inclusive classrooms in grade 2 participated in a structured intervention lasting ten weeks, under three conditions: Cooperative learning (CL, students working in pairs during seatwork), individual learning (IND, students working individually during seatwork) and a control group (CG) with "business as usual." Even when there was substantial class-specific differential development during group-based interventions, between pre- and posttest, results showed no statistically significant main effects for either of the two intervention conditions relative to the control group. However, compared to the CG-group, the CL intervention showed a higher slope for pretest performance, indicating that students with a higher level of computation competence benefited relatively more from the CL intervention than students with similar preconditions in the CG. The results provide evidence that multiple approaches are needed; approaches which combine well-structured programs stimulating cooperative learning and shared learning situations with intensive and individualized measures are of benefit for students with very low mathematical competence.

Keywords: **Cooperative Learning, Computation Competence, Counting Strategies, Low Achiever in Mathematics, Inclusive Instruction, Intervention Study**

Introduction

The inclusive instruction of students with special educational needs (SEN) is a major aim of education in many countries. Research shows positive achievement

*Please send correspondence to: Elisabeth Moser Opitz, PhD, University of Zurich, Institute of Education, Freiestrasse 36, Zurich, 8032, Switzerland, Phone: 0041 44 634 31 36, Email: elisabeth.moseropitz@uzh.ch.

outcomes for SEN students in inclusive classrooms (e.g. Sermier Dessemontet & Bless, 2013; Ruijs & Peetsma, 2009) when they are compared to those in special classes. However, the best way to provide suitable support in inclusive classrooms for all students has yet to be clarified. Some scholars focus on social participation, demand full inclusion, and stress the importance of joint learning situations for students with and without SEN. This view is supported by a study by Wiener and Tardif (2004). They report negative social inclusion outcomes when SEN students are placed in Resource Rooms and Self-Contained Classrooms. Feldman, Carter, Asmus, and Brock (2015) emphasize that "students must be both present and in proximity to peers as prerequisites to fostering peer interaction and shared learning" (p. 204). Other authors, such as McLeskey and Waldron (2011), report research results showing that the quality of inclusive instruction with all students in one classroom is often low and not sufficiently explicit and intensive. However, there is no research on creating tailored programs for inclusive classrooms with shared learning situations that are simultaneously explicit and intensive. Our study aims to help to fill this research gap. We investigated whether a structured program, which includes peer learning (PL), is suitable for promoting the computation competence of low achievers[1] in mathematics in inclusive classrooms.

Peer Learning

When describing peer learning, researchers use several different terms and definitions: Peer assisted learning (PAL), peer-mediated learning, or cooperative learning (CL). According to Ginsburg-Block, Rohrbeck, and Fantuzzo (2006), peer assisted learning (PAL) is "an umbrella term that includes dyadic peer tutoring and small group cooperative learning interventions" (p. 732). PML involves pairs of students working collaboratively on structured and individualized activities and tasks (Kunsch, Jitendra, & Sood, 2007). In this study, we focused on cooperative learning. Smith (1996) defines CL as "the instructional use of small groups so that students work together to maximize their own and each other's learning" (p. 71). Johnson and Johnson (1986) delineate the following four characteristics of CL: Positive interdependence, or "the perception that one is linked with others in a way that one cannot succeed unless the others do," (p. 555); individual accountability, meaning that each member is aware that he or she must fulfill responsibilities for the success of the group; collaborative skills; and group processing, which gives the groups time to discuss if they are achieving their goals. An important aspect of CL is the "think-pair-share" principle (Lyman, 1998). First, students work individually on similar tasks. Second, they discuss their results, strategies and queries, and finally, they share their findings with the whole group. Another principle is "taking turns," where students take it in turns to solve the problem.

The research results of prior studies on the effectiveness of CL are inconsistent. A meta-analysis by Slavin (1995) reported significant effects for CL in two-thirds of the studies. O'Connor and Jenkins (2013) concluded that "CL is a blunt instrument that, depending on its form and implementation, may or may not help students

1 Due to a lack of consistent diagnostic criteria for identifying "learning disabilities" we use the term "low achievers in mathematics" and explain our criteria in the section "Method". If we refer to a specific paper, we use the authors' terms.

with learning disabilities " (p. 521). Gilles and Ashman (2000) demonstrated that students with learning disabilities who had participated in structured CL, performed significantly better on a comprehension questionnaire than their peers in unstructured groups. Kroesbergen and Van Luit (2003) found that compared to other mathematics interventions, CL is less effective for children with learning problems. The authors concluded that direct instruction seemed to be the most effective method for teaching basic math facts. Finally, Nührenbörger and Steinbring (2009) emphasized the importance of choosing and arranging mathematical problems and structuring the setting so that communication between students is initiated. In conclusion, it can be hypothesized that structured learning settings are crucial for successfully teaching low achievers in mathematics and for the implementation of CL.

Computation Problems of Low Achievers in Mathematics

Development of Computation Strategies in Addition and Subtraction. Computation fluency is an important aim of mathematics instruction in primary school. Children's first strategies for addition and subtraction are counting strategies, which develop into more advanced strategies, e.g. decomposition or retrieval, during the first years of school. However, empirical studies show that low achievers in mathematics frequently rely on immature, (finger) counting strategies even for easy computation problems (Andersson, 2008; Fuchs et al., 2010; Jordan, Hanich, & Kaplan, 2003; Ostad, 1999) and have problems developing more advanced strategies. The persistent use of counting strategies in higher grades can hinder a child's further arithmetical development (Baroody, 2006; Baroody, Purpura, Eiland, & Reid, 2014). For simple computation problems with small numbers, counting strategies may be effective. However, for bigger numbers and more complex problems – especially double-digit problems – counting strategies are inefficient and error-prone. According to Jordan, Kaplan, Ramineni, and Locuniak (2008), in Kindergarten, frequency of finger use was a reliable predictor of number combination accuracy. However, at the end of grade 2, a significant negative correlation between finger use and accuracy was found. Gaidoschik (2012) analyzed the development of computation strategies in grade 1 and observed that the "counting on" strategy (with fingers or manipulatives) seemed to persist. He assumed that the repeated use of counting on hindered the storage of basic facts in the child's long-term memory. Vasilyeva, Laski, and Chen (2015) investigated fluency with number facts using different math problems. They reported an increased use of counting strategies from single to mixed-digit problems. Interestingly, the study found a decreased use of counting strategies for double-digit problems. Students without learning problems appeared to instinctively understand the inefficiency of counting when solving complex problems. Gaidoschik (2012) and Lorenz (2015) emphasized that relying on counting by ones hinders insight into the structure of the base-ten system; persistent counters interpret numbers mainly as ordinal numbers (numbers on a number line) and not as cardinal numbers (how many?). This impedes insight into the relationships between quantities, (e.g. the part-whole-relationship), which are an important aspect of number acquisition (Krajewski & Schneider, 2009).

Several factors seem to account for the phenomenon of persistent counting strategies: It is assumed that the use of counting strategies is related to problems with fact retrieval caused by deficits in working memory (e.g., Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Jordan, Hanich, & Kaplan, 2003; Passolunghi & Siegel, 2004). Also, children appear to use counting strategies when they lack an understanding of numbers, relationships between numbers, and operations. Teaching methods may also influence whether children use counting strategies, for example, if teachers prompt children to use (finger) counting strategies or if they fail to explain and teach a strategy repertoire (Gaidoschik, 2012). All of these issues factor into determining if and how the development of flexible computation strategies can be supported. According to Siegler's Adaptive Strategy Choice Model - ASCM (1996), children use multiple strategies for a prolonged period of time, while simultaneously discovering new strategies and adding them to the repertoire. Shrager and Siegler (1998) and Häsel-Weide (2016) explained that generalization proceeded slowly and that old and new strategies are often used simultaneously. Thus, strategy use and its adaptivity are influenced by how accurately and quickly a strategy is executed for a particular item by a particular child "in comparison to other concurrent strategies available in the child's repertoire" (Verschaffel, Torebeyns, De Smedt, Luwel, & van Dooren, 2007, p. 19). Low achievers in mathematics use the same strategies as other children, but they execute more advanced strategies, such as decomposition, less accurately and more slowly (Verschaffel et al., 2007).

Intervention Studies on Computational Deficits. Even if the problem of the persistent use of counting strategies is widely acknowledged, the extent to which low achievers in mathematics should be taught flexible strategies, and which approaches would be the most suitable for teaching flexible strategies, are controversial issues (Cowan et al., 2011). Some authors (e.g. Powell, Fuchs, Fuchs, Cirino, & Fletcher, 2009; Fuchs et al., 2010) argued that it was best to first teach routines, such as memorizing basic skills, and then proceed to adaptive strategy use. They reported positive outcomes from the "routines first" approach. Others, e.g. Baroody (2006), emphasized the "number sense view," which relies on students discovering numerous patterns and relationships from the beginning. Cowan et al. (2011) provided empirical evidence for the number sense view. However, interventions that take strategy-use into account are rare. In a study by Moser Opitz (2002), persistent counting strategies in grade 1 could be reduced significantly by implementing a number-sense program that focused on the use of different addition and subtraction strategies with the support of a twenty frame[2]. Thus, the objective of this study was to adapt and evaluate such an approach for students who use persistent counting strategies in grade 2.

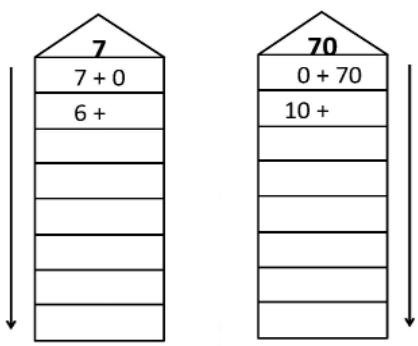
An Approach to Foster the Use of Non-Counting Strategies. The approach aimed to organize the learning processes of the students on the level of content and on the level of the setting (Nührenbörger & Steinbring, 2009). The content focused on relations between numbers, recognizing "simple" addition and subtraction problems (e.g. 6 + 6, 18 – 8) and the process of decomposition (Table 1). The overall aim was to develop students' mental representations of numbers, operations, and relations.

2 A grid that contains two rows of 10 dots each; within each row, dots are organized in two groups of five.

Table 1. ***Topics of the Intervention***

Week	Topic	Content
1	Whole-Part Relationship	Number conservation (“Always 7”); recognizing number sets “at a glance”
2		Number (de)composition
3		Representations of numbers on the twenty-frame
4	Counting competence	Counting by groups
5		Number sequences and number relations at the number line
6	Representation of Addition and Subtraction	Adding up to tens and subtracting to tens
7		Doubling
8	Computation number relations	Addition problems: recognizing related problems
9		Recognizing simple subtraction problems and using them for derivation
10		Addition and subtraction problems on the empty number line

The setting was organized by providing detailed lesson plans for the teachers. The lessons always followed the same structure: introduction (whole class), seat work, reflection, and discussion. To test the effectiveness of cooperative learning (CL), two different settings for the working phase were developed, CL and individual learning (IND). In the CL-condition, the intervention involved pairs of children with different computation competences who worked together during the whole intervention. The pairs were selected by the teacher. During the working phase, two cooperative settings were implemented (think and pair, and taking turns), and it was possible to work on math problems at different levels of difficulty (Figure 1; Häsel-Weide et al. 2017). In the IND setting, students, supported by the teacher if necessary, worked individually on worksheets with the same mathematical problems as those given to students in the CL setting. This was so that the work element of the CL protocol would be comparable to that in a more intensive, individualized setting.



Complete your number house.
Compare! Which calculations belong together? Colour in the same colour.

Figure 1. **Task for a heterogeneous group in the CL condition.**

Research Questions and Hypothesis

In summary, existing research on the development of non-counting strategies and on CL raises the following research questions:

Q1. *Does an intervention program that focuses on number sets and number relations carried out in inclusive classrooms lead to increased success in solving arithmetical problems without (finger) counting strategies among low achievers in mathematics who use persistent counting strategies?*

The program was implemented under two different conditions: Cooperative learning (CL), which creates shared learning situations, and individual learning (IND). Therefore, the second research question is:

Q2. Do the results of the CL and IND intervention significantly differ?

To answer these research questions, the following hypotheses were tested:

H1. Participants in intervention condition CL are able to solve more computation problems with non-counting strategies than students who did not take part in the intervention.

H2. Participants in intervention condition IND are able to solve more computation problems with non-counting strategies than students who did not take part in the intervention.

Due to the inconsistent results from CL, we expected this intervention to be as effective as the IND-intervention.

H3. Participants in intervention conditions CL and IND do not differ in the ability to solve computation problems with non-counting strategies.

Method

Participants

The participants were students with below-average mathematics achievement who used persistent (finger) counting strategies at the end of grade 1and were in inclusive classrooms. The initial sample (sample$^{\text{INITIAL}}$) consisted of 831 students from 38 classes in 19 schools in Germany. Teachers voluntarily agreed to participate in the study and students taking part had written parental consent. Trained test administrators gave the students both an IQ test and a mathematics test at the end of grade 1. From the initial sample, 126 persistent counters (sample$^{\text{COUNT}}$) from 35 classes were selected for the evaluation of the intervention (see Instruments for criteria).

Instruments

Mathematics achievement in the sample$^{\text{INITIAL}}$ was measured at the end of grade 1 (t1) using DEMAT1+, a curriculum-based standardized test (Krajewski, Rüspert, Schneider, & Visé, 2002). Students with weak mathematical achievement (first terzile, n = 326) in the DEMAT1+ test sat a second test developed by the research group, (test$^{\text{STRAT}}$, Wittich, 2017), which allowed for the assessment of computation strategies. 126 students were selected and defined as persistent counters and were assessed at three points in time using test$^{\text{STRAT}}$: At the start (t1), after the intervention (t2), and three months later, at the end of the school year (t3). The test included 13 computation problems for the pretest and 19 items for both the posttest and the follow-up (cf. table 2).

Table 2. ***Items of the Test^{Strat}***

Addition	Subtraction
8 + 7	9 – 6
9 + 5	11 – 3
6 + 12	19 – 4
17 + 4	23 – 6
3 + 15	14 – 5
7 + 7	20 – 13
27 + 9*	16 – 8
8 + 23*	26 – 19*
46 + 15*	32 – 16 *
	42 – 39*

*Items used in the posttest and follow-up only

In the test^{STRAT}, students were asked to carry out a "dual task" during computation (Grube, 2006; Thomas, Zoelch, Seitz-Stein, & Schumann-Hengsteler, 2006): they had to tap their hand regularly on the table. Persistent counters are expected to stop the tapping or to do it irregularly, and/or to switch to counting strategies. Each child was tested individually in a quiet room at school. The test administrator started the tapping at a frequency of 120 beats per minute. The student joined in, tapping at the same rate. An empty slide was presented on a screen (15" and 17"), after a verbal signal ("now") a computation problem (e. g. 8 + 7 =, presented in Arial pt. 96) appeared on grey background. The student was advised to solve the problem as quickly as possible and say the result out loud. The test administrator pressed a key when the student answered the problem. Processing time was calculated (appearance of the problem through spoken solution). If the processing time was more than 3 seconds or the tapping was irregular or stopped, the students were asked the following question: "How did you solve the problem?" The strategies were assessed according to three categories: Retrieval (answer within 3 sec), decomposition (e.g. 8 + 7 → 7 + 7 + 1), and counting. The criteria for assessing a strategy as a counting-strategy were: a) student answered "I solved it by counting", b) student heard counting out loud, c) hidden counting strategies (e.g. moving the lips when rehearsing a number sequence), d) moving the fingers. Items were scored 0 for an incorrect answer, 1 for a correct answer with a counting strategy, and 2 for a correct answer by retrieval or decomposition. Students who solved more than five problems with counting strategies at t1 were defined as persistent counters (mean of counting strategies = 10.0, SD = 1.8, min = 6, max = 13).

For the sample of the 326 students with weak mathematics achievement (first terzile in DEMAT1+), Cronbach's alpha for the pretest^{STRAT} was .85. For the sample^{COUNT} (n = 126), Cronbach's alpha was .68[3] for the pretest, .84 for the posttest

3 The rather low alpha coefficient in the persistent counters sample at t1 results from a reduced variance in non-counting computation performance.

and .86 for the follow-up. Intelligence was tested at t1 with CFT 1 (Weiß & Osterland, 1997). The IQ cut-off was set at 75. Data on background factors such as gender, age, and first language were collected through a questionnaire given to the teachers. The 126 persistent counters were randomly assigned to the three groups: cooperative learning (CL), individual learning (IND), and a control group (CG; description see below). Table 3 gives an overview of the demographic and cognitive characteristics of this sample.

Table 3. ***Sample Properties by Treatment Group***

		CL	IND	CG	Total	*p*
n		53	34	39	126	
sex	female (*n*, percent)	32 (60 %)	21 (62 %)	31 (79 %)	84 (67 %)	
	male (*n*, percent)	21 (40 %)	13 (38 %)	8 (21 %)	42 (33 %)	.123[a]
First language	german (*n*, percent)	48 (91 %)	27 (96 %)	31 (79 %)	106 (88 %)	
	other (*n*, percent)	5 (9 %)	1 (4 %)	8 (21%)	14 (12 %)	.082[a]
Age in years t1 (*M*, *SD*)		7.3 (0.4)	7.2 (0.5)	7.2 (0.4)	7.2 (0.4)	.271[b]
IQ t1 (*M*, *SD*)		105.3 (13.2)	103.5 (13.0)	102.1 (12.7)	103.8 (13.0)	.510[b]
Math pretest score (*M*, *SD*)		7.8 (3.8)	7.3 (3.0)	6.9 (3.3)	7.4 (3.4)	.460[b]

Notes. IND = individual learning group; CL = cooperative learning group; CG = control group.
[a] error probability of chi-square test, [b] error probabilities of one-factor ANOVAs

The three groups, CL, IND, and CG did not differ significantly in sex ratio ($Chi^2 = 4.20$, $df = 2$, $p = .123$), first language ($Chi^2 = 5.00$, $df = 2$, $p = .082$), mean age ($F_{(2, 123)} = 1.32$, $p = .271$), mean IQ ($F_{(2, 123)} = 0.677$, $p = .510$), or mean pretest computation score ($F_{(2, 123)} = 0.78$, $p = .460$). Table 4 documents uni- and bivariate statistics of the key variables.

Table 4. ***Descriptive Statistics and Correlations Among the Key Variables***

Variables	*M* (*SD*)	Sex	Language	Age	IQ	Pretest score	Posttest score	Follow up score
Sex	1.67 (0.47)	1	.03	-.14	.07	-.10	-.18*	-.12
Language	7.2 (0.4)		1	.00	.13	.19*	-.02	-.01
Age	7.2 (0.4)			1	-.21*	-.06	.03	-.01
IQ	103.8 (13.0)				1	.27***	.29***	.22**
Pretest score	7.4 (3.4)					1	.59***	.48***
Posttest score	15.8 (7.3)						1	.69***
Follow up score	20.2 (8.3)							1

Notes. N=126; Sex; 1 = female, 2 = male. First language: 0 = German, 1 = other. * $p < .05$, ** $p < .01$, *** $p < .001$

There was a statistically significant linear dependence of the computation scores on the students' IQ at all three points in time. Additionally, the pretest score correlated significantly with first language and the posttest score with sex. These correlations suggest that some of these variables should be included as covariates in the multilevel regression models, despite the nonsignificant differences between the treatment groups (cf. table 2). Student age did not contribute to explaining non-counting computation competence[4], especially when controlling for IQ. Therefore, age was not included in the multilevel analyses. Also, student first language was omitted (very limited and redundant contribution, 6, 5% data is missing).

Instructional Design

Over a period of 10 weeks, the teachers of the group CL and IND carried out the program twice a week for 30 minutes. The control group attended classroom instruction with "business as usual." To guarantee fidelity, the teachers had a manual with standardized lesson plans which was introduced in two teacher training meetings lasting three hours each. In addition, in each classroom one program-unit of 30 minutes was videotaped. The analysis showed that all of the teachers implemented all three phases. As a further check, teachers were asked to regularly submit protocol sheets, detailing their use of the lesson plans. For an additional video study (Häsel-Weide, 2016), five persistent counters and their partners from the CL group were videotaped ten times (whole program unit) during the intervention.

Data Analysis Procedures

A series of two-level hierarchical regression models were estimated in order to test the statistical hypotheses and to answer the research questions (Bryk & Raudenbush, 1992; Hox, 2011). First-level units were the 126 students of the sample$^{\text{COUNT}}$. Second-level units were the 35 classes. The mean number of persistent counters per class was 3.6 (SD = 2.4). The multi-level analyses were performed using HLM 7.01 (Raudenbush, Bryk, & Congdon, 2013). For the three conditions – CL, IND, and CG – two Level 2 dummy variables (CL: yes/no, IND: yes/no) were created. Their effects were estimated in separate models: relative to the control group and, including a dummy for CG, relative to each other. In order to assess the effects of the treatment variable and the controls (gender, age, IQ, all on level 1) on the development of the computation competence between t1 and t2 and between t1 and t3, computation competence at pretest was included as a level 1 predictor. A possible dependency between the effect of the treatment conditions and the level of computation competence at t1 was tested by modeling cross-level interactions between the two treatment dummy variables and the pretest score.

Results

For computation competence, varying degrees of intra-class correlation (ICC) were found. The ICC of the pretest scores was only 3 %, most certainly due to the homogenization resulting from selecting low achievers from each class. In contrast, ICC was 25 % in the posttest and 13 % in the follow-up. This suggests a

4 In the following, we use the term "computation competence" as a synonym for "non-counting computation competence."

substantial amount of computation-related class specific differential development during group-based intervention between t1 and t2. The actual mean values of computation competence, by group, for the three measurement points are listed in table 5. The increase over time is slightly larger in the CL condition.

Table 5. ***Descriptive Statistics of the Dependent Variables of the Sample^{COUNT}***

	CL (*n* = 53)	IND (*n* = 34)	CG (*n* = 39)	Total (*n* = 126)
Pretest mathematics score	7.8 (3.8)	7.3 (3.0)	6.9 (3.3)	7.4 (3.4)
Posttest mathematics score	17.6 (8.2)	14.5 (7.5)	14.6 (5.2)	15.8 (7.3)
Follow up mathematics score	21.8 (8.8)	18.3 (9.0)	19.7 (6.6)	20.2 (8.3)

*Notes.*Values presented as *M* (*SD*). CL = cooperative learning group; IND = individual learning group; CG = control group.

Table 6 shows the results of two hierarchical linear models with the computation posttest at t2 as the dependent variable. Model 1 contains the following predictors: two dummy variables for the treatment condition on level 2, and sex, IQ, and the pretest computation score on level 1. In Model 2, an additional cross level interaction for the effect of the two dummies on the slope of the computation pretest is included.

Table 6. ***Multilevel Regression for Posttest Mathematics Score***

	Model 1			Model 2		
	b	*SE*	*p*	*b*	*SE*	*p*
Level 2						
Intercept	18.94	1.88	.000	18.75	1.89	.000
CL	1.45	1.39	.303	1.56	1.37	.264
IND	-0.92	2.10	.665	-0.68	2.11	.748
Level 1						
Sex	-1.92	0.97	.051	-1.94	0.96	.046
IQ	0.07	0.03	.034	0.08	0.04	.042
Pretest	1.14	0.14	.000			
Cross-level interaction: Treatment group effect on pretest slope						
Intercept				0.72	0.09	.000
CL				0.63	0.21	.003
IND				0.43	0.26	.105

Notes. N=126; no missing data. Sex: 1 = male, 2 = female; CL = cooperative learning group; IND = individual learning. Reference category: control group.
* $p < .05$, ** $p < .01$, *** $p < .001$

The two level 1 covariates, sex and IQ, had only a limited impact on the development of computation competence between t1 and t2 (cf. models 1 and 2). Not only are the error probabilities close to the conventional statistical significance threshold of $p = .05$, but also the effect sizes are very small with β values below .15. This means that individual progress between pretest and posttest does not substantially depend on the individuals' sex or intelligence. The effect of the pretest computation score itself (Table 6, model 1) is large ($\beta = .53$, $p = .000$), which indicates a high level of continuity in computation performance. Contrary to expectations, the treatment conditions CL and IND did not have a significant impact on the progress in computation relative to the control group. No significant difference was found between the CL and the IND condition, either ($b = 2.37$, $p = .289$). As a consequence, H1 and H2 have to be rejected, H3 (CL condition as successful as the IND condition) can be accepted. There was, however, a significant cross-level interaction between the treatment condition and the pretest computation score (Table 6, model 2): In the CL condition, computation performance in the posttest was influenced to a significantly higher degree by the pretest score than in the CG condition ($b = 0.63$, $p = .003$). This means that students with higher computation scores at t1 tend to make more progress relative to those with a lower computation score when participating in the CL intervention compared to the CG. Given the symmetric character of interactions, this can be interpreted as a significantly stronger and more positive effect of the CL intervention, relative to the CG condition, on computation competence development in students with a higher competence in the pretest compared to those with lower competence. Figure 2, showing mean gains for CL vs. CG condition by pretest score level[5], supports and illustrates these findings.

Students with pretest computation competence above the median had higher benefits ($M = 11.5$) from the CL intervention than students with below median pretest scores ($M = 7.9$). In contrast, under the CG condition students with pretest computation competence above the median did have a slightly lower benefit ($M = 7.0$) compared to students with below median pretest scores ($M = 8.2$).

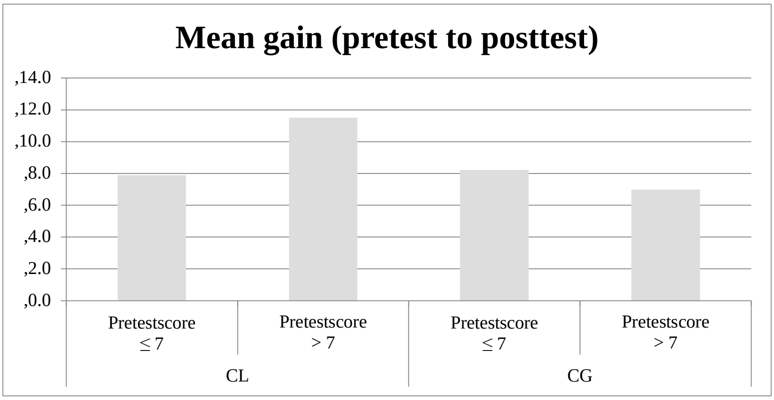


***Figure 2.* Mean gain in CL vs. CG condition by pretest computation competence**

5 Pretest scores were split at the overall median of 7.

Effects on Computation Competence in the Follow-Up (t3)

Analogously to the short term models 1 and 2 above (Table 6), the impact of the two intervention conditions on computation competence at t3 (follow-up) was analyzed in models with and without cross-level interaction. There were no statistically significant differences between the intervention groups and the control group (CL-CG: $b = 0.22$, $p = .904$; IND-CG: $b = -2.76$, $p = .244$)[6]. Also, the effects of the two treatment conditions did not differ (CL-IND: $b = 2.98$, $p = .215$). Therefore, H1 and H2 have to be rejected for the follow-up, too. H3 can be accepted. The cross level interaction for computation competence at t2 was not found in the follow-up (t3). Therefore, the intervention did not have a long-term impact.

Discussion

Our study aimed to investigate if a structured and intensive program, which includes cooperative learning in inclusive classrooms, is suitable for fostering the computation competence of low achievers in mathematics who rely persistently on counting strategies. Analysis of the general impact of the intervention (research question 1), does not show clear confirmation of the benefits of the intervention. Even though there is evidence for substantial class-specific differential development between t1 and t2 in the group-based intervention, results revealed no statistically significant main effects for any of the two interventions over the control group, for either the posttest or the follow-up. However, just after the intervention, a significant cross-level interaction of the CL-intervention and the computation competence pretest was identified. Compared to the CG, the CL intervention showed a higher slope for the pretest performance, indicating that students with a higher level of computation competence benefited relatively more from the CL intervention than students with similar preconditions in the CG. But, this differential effect did not last for the follow-up.

How can these results be explained? First – and this is a limitation of the study due to the constraint of the annual school calendar – a ten week intervention was probably too short (Ise, Dolle, Pixner, & Schulte-Körne, 2012). Slavin (2008) demonstrated that in order for an intervention to be successful, it had to last for at least 12 weeks. Second, Shrager and Siegler (1998) explained that old and new computation strategies are often used simultaneously. This was corroborated by the results of the video study by Häsel-Weide (2016), carried out with our sample. That analysis showed that persistent counters participating in the study used existing counting alongside newly acquired non-counting strategies. For example, they tended to interpret relations between numbers within the ordinal paradigm, but also sometimes used cardinal or relational interpretation. However, the students did not notice the relationship between math problems, as the following example shows. By explaining the relationship between 16 – 6 = 10 and 16 – 7 = 9, the students did recognize the difference between 6 and 7 in the subtrahend, and between 10 and 9 in the result. However, they did not take into account that the difference will be smaller when the subtrahend increases (Häsel-Weide, 2016).

6 To save space these findings are presented without a table.

Nevertheless, our results give some new insights into the effectiveness of well-structured and intensive CL measures. First, no significant negative difference was found compared to the IND-condition. Second, the result of the significant cross-level interaction effect in favor of the CL group compared to the CG is promising. Persistent counters with more prior knowledge had better learning gains than students in the control group. This is especially encouraging because the intervention was tested in real-life situations with high ecological validity. This shows that it is possible to implement well-structured and intensive CL measures – and shared learning situations for students with different achievement levels – successfully in inclusive classrooms. Verschaffel et al. (2007) mention that the sociocultural context, in the sense of what is defined as appropriate in a certain context (e. g counting or not counting when solving computation problems), is a factor that influences the development of flexible computation strategies. If an intervention could be carried out over a longer period of time, non-counting strategy use could become more evident. Despite the positive outcome for students in the CL group with higher prior competence, it should be noted that students with low prior computation competence did not benefit from the CL-intervention. It may be that those students need more individualized, intensive measures. The meta-analysis from Ise et al. (2012) has shown an advantage of one-to-one interventions compared to small group and classroom interventions for students with poor mathematics achievement. For these students, classroom interventions alone do not seem to be sufficient.

Further limitations of the study must be mentioned. Even if the fidelity of the intervention was controlled in several ways, we do not know exactly how the teachers used the materials, and how adaptively they guided the students. In addition, we do not have any information about the teaching in the CG. Nor do we know whether the intervention material designed for use between posttest and the follow-up was used. Finally, our test^{STRAT} was possibly not sensitive enough to measure the development of non-counting strategies.

In conclusion, even if the main hypotheses have to be rejected, the results give evidence that it is worth conducting more research on well-structured CL-programs that include shared learning situations suitable for fostering social participation. Second, multiple approaches are needed to support students with very poor mathematics achievement in inclusive classrooms. These approaches should offer both well-structured programs stimulating cooperative learning and shared learning situations, and intensive, individualized measures.

References

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology, 100*, 48–66.

Baroody, A.J. (2006). Why children have difficulties mastering the basic number combinations and how to help them. *Teaching Children Mathematics*, *8*, 22−31.

Baroody, A. J., Purpura, D. J., Eiland, M. D., & Reid, E. E. (2014). Fostering first graders' fluency with basic subtraction and larger addition combinations via computer−assisted instruction. *Cognition and Instruction, 32*, 159–197.

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Cowan, R., Donlan, C., Shepherd, D. L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*, 786–803.

Feldman, R., Carter, E. W., Asmus, J., & Brock, M. E. (2015). Presence, proximity, and peer interactions of adolescents with severe disabilities in general education classrooms. *Exceptional Children, 82*, 192–208

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learning and Individual Differences*, *20*, 89–100.

Gaidoschik, M. (2012). First-Graders' Development of Calculation Strategies: How Deriving Facts Helps Automatize Facts. *Journal für Mathematik-Didaktik, 33,* 287–315.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disabilities. *Child Development, 78,* 1343–1359.

Gilles, R. M., & Ashman, A. F. (2000). The effects of cooperative learning on students with learning difficulties in the lower elementary school. *Journal of Special Education, 34,* 19–27.

Ginsburg-Block, M.D., Rohrbeck, C.A., & Fantuzzo, J.W. (2006). A meta-analytic review of social, self-concet, and behavioral outcomes of peer-assisted learning. *Journal of Educational Psychology, 98,* 732–749.

Grube, D. (2006). *Entwicklung des Rechnens im Grundschulalter [Development of mathematical skills in elementary school age].* Münster, Germany: Waxmann.

Häsel-Weide, U. (2016). *Vom Zählen zum Rechnen. Struktur-fokussierende Deutungen in kooperativen Lernumgebungen. [From counting to calculating].* Wiesbaden, Germany: Springer Spektrum.

Häsel-Weide, U., Nührenbörger, M., Moser Opitz, E., & Wittich, C. (2017). *Ablösung vom zählenden Rechnen. Fördereinheiten für heterogene Lerngruppen. [Replacing counting calculating]*. Seelze, Germany: Klett Kallmeyer.

Hox, J. J. (2011). *Handbook of advanced multilevel analysis.* New York, NY: Routledge.

Ise, R., Dolle, K., Pixner, S., & Schulte-Körne, G. (2012). Effektive Förderung rechenschwacher Kinder: Eine Metaanalyse [Effective treatment of children with poor math abilities: Results of a meta-analysis]. *Kindheit und Entwicklung, 21*, 181–192.

Johnson, D.W. & Johnson, R.T. (1986). Mainstreaming and cooperative learning strategies. *Exceptional Children, 52,* 553–561.

Jordan, N. C., Kaplan. D., Ramineni, C., & Locuniak, M. N. (2008). Development of number combination skill in the early school years: When do fingers help? *Developmental Science, 11*, 662–668.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). A longitudinal study of mathematics competencies in children with specific mathematics difficulties versus children with comorbid mathematics and difficulties. *Child Development, 74*, 834–850.

Krajewski, K. & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties. Findings from a four-year longitudinal study. *Learning and Instruction, 19*, 513–526.

Krajewski, K., Rüspert, P., Schneider, W., & Visé, M. (2002). *Deutscher Mathematiktest für erste Klassen (DEMAT 1+) [German Mathematics Test for First Graders]*. Göttingen, Germany: Beltz.

Kroesbergen, E. & van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: a meta-analysis. *Remedial and Special Education, 24*, 97–114.

Kunsch, C.A., Jitendrea, A.K., & Sood, S. (2007). The effects of peer-mediated instruction in mathematics for students with learning problems: A research synthesis. *Learning Disabilities Research & Practice, 22,* 1–12.

Lorenz, J. H. (2015). Fingerrechnen: Aspekte aus didaktischer Sicht [Calculating with the fingers seen from a didactic perspective]. *Lernen und Lernstörungen, 4*, 195–207.

Lyman, F. T. (1998). The responsive classroom discussion: The inclusion of all students. In A. Anderson (Ed.), *Mainstreaming Digest* (pp. 109–113). College Park, MD: University of Maryland Press.

McLeskey, J. & Waldron, N. L. (2011). Educational programs for elementary students with learning disabilities: Can they be both effective and inclusive? *Learning Disabilities Research & Practice, 26*, 48–57.

Moser Opitz, E. (2002). Mathematical knowledge of children with special needs in their first year of school. In H.G. Weigand, A. Peter-Koop, N. Neil, K. Reiss, G. Törner, & B. Wollring (Eds.), *Selected papers from the annual conference on didactics of mathematics, Potsdam 2000* (pp. 99–108). Hildesheim, Germany: Franzbecker.

Nührenbörger, M. & Steinbring, H. (2009). Forms of mathematical interaction in different social settings – Examples from students', teachers' and teacher-students' communication about mathematics. *Journal of Mathematics Teacher Education, 12*, 111–132.

O'Connor, R.E. & Jenkins, J.R. (2013). Cooperative Learning for students with learning disabilities. In H.L. Swanson, K.R. Harris & S. Graham (Eds.), *Handbook of Learning Disabilities* (pp. 507–525) (2nd edition). New York, NY: The Guilford Press.

Ostad, S. A. (1999). Developmental progression of subtraction strategies: a comparison of mathematically normal and mathematically disabled children. *European Journal of Special Needs Education, 14*, 21–36.

Passolunghi, M. C. & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*, 348–377.

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learning Disabilities Research & Practice, 24*, 1–11.

Raudenbush, S. W., Bryk, A., & Congdon, R. (2013). HLM 7.01 for Windows [Computer software]. Lincolnwood, IL: Scientific Software International.

Ruijs, N. M. & Peetsma, T. T. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review, 4*, 67–79

Sermier Dessemontet, R. & Bless, G. (2013). The impact of including children with intellectual disability in general education classrooms on the academic achievement of their low-, average-, and high-achieving peers. *Journal of Intellectual & Developmental Disability, 38,* 23–30.

Shrager, J. & Siegler, R.S. (1998). SCADS: A model of children's strategy choices and strategy discoveries. *Psychological Science, 9*, 405–410.

Siegler, R.S. (1996). *Emerging minds. The process of change in children's thinking.* New York, NY: Oxford University Press.

Slavin, R. E. (1995). *Cooperative learning. Theory, research, and practice* (2nd ed.). Boston, ILL: Allyn and Bacon.

Slavin, R. E. (2008). Perspectives on evidence-based research in education. What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, *37*, 5–14.

Smith, K. A. (1996). Cooperative learning: Making "groupwork" work. *New Directions for Teaching and Learning, 67,* 71–82.

Thomas, J., Zoelch, Ch., Seitz-Stein, K., & Schumann-Hengsteler, R. (2006). Phonologische und zentral-exekutive Arbeitsgedächtnisprozesse bei der mentalen Addition und Multiplikation bei Grundschulkindern [Phonological and central executive working memory processes in children's mental addition and multiplication]. *Psychologie in Erziehung und Unterricht, 53,* 275–290.

Vasilyeva, M., Laski, E.V., & Shen, C. (2015). Computational fluency and strategy choice predict individual and cross-national differences in complex arithmetic. *Developmental Psychology, 51*, 1489–1500.

Verschaffel, L., Torbeyns, J., De Smedt, B., Luwel, K., & Van Dooren, W. (2007). Strategy flexibility in children with low achievement in mathematics. *Educational & Child Psychology, 24*, 16–27.

Weiß, R. & Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1) [Culture Fair Intelligence Test, Scale 1].* Göttingen, Germany: Hogrefe.

Wiener, J. & Tardif, C. Y. (2004). Social and emotional functioning of children with Learning disabilities: Does special education placement make a difference? *Learning Disabilities Research & Practice, 19*, 20–32.

Wittich, C. (2017). *Mathematische Förderung durch kooperativ-strukturiertes Lernen [Fostering math skills through cooperative-structured learning].* Wiesbaden, Germany: Springer.