

Systematic reading training in the family: Development, implementation, and initial evaluation of the Berlin Parent-Child Reading Program¹

Nele McElvany, Cordula Artelt

Abstract Drawing on knowledge about the development of reading comprehension, and empirical insights into the effects of training on students' learning strategies and reading-related metacognition, we developed a parent-child reading program for implementation in the home environment. The results of this first quasi-experimental evaluation study indicate that it is generally possible to implement a program of this kind within the family setting, but that participation is low and selective based on family background and children's achievement level. Nevertheless, participation in the program was found to have substantial effects on the development of vocabulary and on reading-related metacognition, indicating that family-based reading programs have considerable potential. The issues of selective participation and the lack of a program effect on text comprehension are discussed, and prospects for future research on systematic reading training in the family context are considered.

Keywords: *Family; Literacy training; Reading; Metacognition; Motivation; Reading comprehension; Vocabulary*

1. Introduction

Reading is a vital cultural tool in modern societies. The ability to read and understand continuous texts is crucial to success in educational, professional, and everyday settings. Proficiency in reading is a key target of schooling and a major prerequisite for learning, both within and beyond the context of formal education. Nevertheless, large-scale international studies have identified serious deficiencies in many students' reading skills (Mullis, Martin, Gonzales, & Kennedy, 2003; Organization for Economic Cooperation and Development., 2001), highlighting the need for effective measures to improve these skills. An impressive number of reading (comprehension) programs has been developed on the basis of theories of reading and memory development and knowledge of instructional design (see National Institute of Child Health and Human Development [NICHD], 2000, for a comprehensive overview). However, most of these programs are geared toward the classroom setting or intended as small group interventions (e.g., Guthrie et al., 2004; Palincsar & Brown, 1984) and do not involve the family. Given the variety of individual prerequisites for reading that are known to be developed rather implicitly through family interactions, such as vocabulary or metacognitive awareness of language (Hurrelmann, 2004; Snow, 1993; Weinberger, 1996; Wood, 2002), it seems worth exploring the potential of theory-based reading programs within the family context. Against this background, we developed and implemented the Berlin Parent-Child Reading Program, and evaluated both its practicability and its effectiveness in

enhancing students' reading skills. The present paper begins with a short overview of research on reading literacy and the family as a setting for the development of reading skills. We then describe the conceptual framework of our reading program, which builds on both lines of research, in more detail.

1.1. Reading literacy and target points of intervention

Reading is a complex process that takes place on the word, the sentence, and the text level. Theories of text comprehension assume that different mental representations of a text are established during the reading process (Johnson-Laird, 1983; van Dijk & Kintsch, 1983). According to the Kintsch model (1998; van Dijk & Kintsch, 1983), mental representations of texts are developed on three different levels on the basis of different reading processes. The verbatim representation reflects the surface features of a text, the propositional representation is based on information derived directly from the text, and the situational representation involves the integration of the content of the text with the reader's prior knowledge. Graesser, Millis, and Zwaan (1997) label these mental representations the surface code, the propositional textbase, and the referential situation model, respectively. They also identify two further levels of mental representation: the communication context and the discourse genre. Kintsch (1998) defines text comprehension as a combination of text-based construction and knowledge-based integration processes.

1 Accepted Author's Manuscript. This article was published in:
Learning and Instruction 19 (2009), 79-95 doi: 10.1016/j.learninstruc.2008.02.002

The determining factors for successful reading processes can be broadly classified into four different categories: characteristics of the text (e.g., coherence), activities of the reader during the reading process (e.g., use of metacognitive strategies), characteristics of reading tasks (e.g., remembering), and individual characteristics of the reader (e.g., vocabulary knowledge) (Artelt et al., 2005; Campione & Armbruster, 1985). Research focusing on the reader (characteristics and activities) has identified several individual prerequisites for successful reading and text comprehension. These include fluency and vocabulary, and thus speed of access to the lexicon (Rayner & Pollatsek, 1989), as well as competence in reading strategies and text-related metacognition (Collins Block & Pressley, 2002). Readers develop different reading strategies as a function of instruction and experience (Borkowski, Milstead, & Hale, 1988). These include metacognitive strategies of planning, monitoring, and regulating the reading process, strategies for elaborating the text in order to gain a deeper understanding of its content, and strategies for constructing and integrating new knowledge and supporting processes of memorization (Baker & Brown, 1984; Collins Block & Pressley, 2002). Weaker readers are known to make less frequent use of strategies than good readers (Artelt, Baumert, Julius-McElvany, & Peschar, 2003).

These general preconditions for successful reading and text comprehension processes (fluency, vocabulary, and the knowledge and use of metacognitive or elaborative strategies) can and should be targeted in reading interventions. Given the relevance of reading in day-to-day life, measures designed to foster reading skills should focus on the mastery of texts—i.e., on aspects such as text comprehension and communication about the text. Readers must be able to use strategies flexibly and appropriately, depending on the text's genre, difficulty, the purpose for reading, and other such aspects.

Grade 4 seems an appropriate time for a systematic intervention designed to foster reading comprehension. Fourth graders in German schools are typically aged 9 to 10 years, and it can be assumed that most of them have acquired basic reading skills and strategies by this point. Grapheme-phoneme correspondence is high in German (as compared to English) and students are able to decode relatively early (Klicpera & Gasteiger-Klicpera, 1993; Wimmer, Landerl, & Frith, 1999). By Grade 4, students are therefore receptive to text comprehension strategies that go beyond simple letter and word decoding (Baker & Brown, 1984; Schneider, 1989; Schneider & Pressley, 1997). Grade 4 is also an important transition point with expository texts being introduced alongside narratives.

1.2. The family as a setting for fostering reading literacy

The family has proved to be an important factor in fostering children's reading literacy both prior to and during schooling. As the earliest influence, it plays an important role in children's socialization into reading before systematic schooling begins, and it remains central to improving their reading literacy during the school years (Elley, 1994; Entwisle & Alexander, 1992). Positive aspects of reading socialization within the family include knowledge about language (e.g., phonological awareness or letter naming) and different content areas (prior knowledge that provides schemata, scripts, and word knowledge), and acquiring an appreciation of reading (Hurrelmann, 2004). Relevant processes in the family context include communication, cultural activities and resources, and parents' behavior as role models (McElvany, 2008). Additionally, families seem to have the potential for implementing systematic interventions designed to foster either reading or the prerequisites for reading in the home environment. The family is viewed as a promising educational setting because it enables parents and children to engage in adaptive, intensive interaction and learning. Some clear advantages of the family setting are (a) the intensity of the one-to-one interaction between parents and children (in contrast to the group or classroom setting), (b) the opportunity to establish a strong tradition of positive reading behavior within families, (c) the possibility of providing the learning child with direct feedback, and (d) the positive role model that can be provided by parents.

According to Vygotsky's (1978) socio-cultural theory, social interaction and communication with "more knowledgeable others" are the preconditions for all higher cognitive capabilities. Given that parents have more opportunity than teachers to work with individual children, and that this form of direct interaction, instruction, and feedback allows intensive practice, the family clearly has great potential as an educational setting. Parent-child interactions involving the intensive co-construction of meaning can be assumed to promote children's reading skills through mechanisms such as scaffolding (Wood, Bruner, & Ross, 1976) and cognitive apprenticeship (Collins, Brown, & Newman, 1989). Nevertheless, some potential disadvantages of parent-child reading programs also need to be considered. Parents are not teachers, in terms of either their didactic knowledge or their content knowledge about reading and teaching reading. Moreover, children and parents have a sensitive relationship that might be disrupted by conflicts and pressures arising from a "teaching-learning" situation (see Grolnick, 2003). Last but not least, families have busy schedules that may need to be adjusted to accommodate a reading program.

To date, research and practice on home-based interventions has focused on early intervention programs, many of which take a broad approach that

is not necessarily specific to reading (Jordan, Snow, & Porche, 2000; Morrow & Young, 1997; Neuman & Gallager, 1994; Wasik, 2000; White, Taylor, & Moss, 1992). Studies evaluating parent involvement programs (Blok, Fukkink, Gebhardt, & Leseman, 2005; Fishel & Ramirez, 2005; Mattingly, Prislín, McKenzie, Rodriguez, & Kayzar, 2002) come to inconsistent results regarding the effectiveness of family interventions. This is partly due to methodological shortcomings of the studies, as well as variations in program approaches and dependent variables that were used. Furthermore, most of these studies have been conducted only in the USA.

Another focus in the area of research into family effects is the general home literacy environment (HLE), which has been shown to impact a variety of (pre-)literacy skills as well as more general cognitive abilities (Centre for Community Child Health, 2004; Gunn, Simmons, & Kameenui, 2000; Hurrelmann, 2004; Whitehurst & Lonigan, 1998). Studies on the HLE examine both the quantity and quality of literacy activities to which children are exposed in the home. For example, meta-analyses of Bus, van Ijzendoorn, and Pellegrini (1995), and Scarborough and Dobrich (1994) found significant effects of shared reading-frequency; Leseman and De Jong (1998) showed the effect of decontextualized talk and Bradley, Corwyn, Pipes McAdoo, and García Coll (2001) showed effects of socioeconomic status and ethnic background on the range of literacy activities provided. There are also ethnographic studies such as the one by Heath (1983).

Despite the clear and pressing need to foster students' reading competence during the school years and the potential value of parental involvement, there is a surprising lack of effective, systematic reading programs that can be implemented within the family context. The Berlin Parent-Child Reading Program was developed to fill this gap for students in Grade 4, drawing on theoretical and empirical evidence from educational psychology and literacy education research. In the next section, we outline the program's conceptual framework.

1.3. Conceptual framework of the Berlin Parent-Child Reading Program

The Berlin Parent-Child Reading Program focuses on the family as an environment facilitating direct, intensive interaction and training. It aims to foster important prerequisites of reading (vocabulary, word fluency in terms of fast and accurate word identification, knowledge and use of metacognitive strategies, and reading motivation), as well as text comprehension skills and the ability to elaborate text content in oral communication in dialogue with a partner. The program's regular reading sessions involve shared reading aloud of selected texts and discussion of elaborating questions in a social dialogue aimed at establishing and nurturing a productive family reading culture. Given the known

importance of fluency, vocabulary, metacognition, motivation, and knowledge and use of strategies as individual prerequisites for successful reading, the main elements of the program's conceptual framework are as follows: (a) guided reading aloud to support the development of decoding, fluency, and vocabulary; (b) systematic questions that monitor the understanding of what has been read to foster knowledge, use of metacognitive strategies, and the acquisition of new vocabulary, and (c) elaboration of text content in joint discussion to train reading strategies, metacognition, and vocabulary. (For a detailed account of procedures see subchapter 1.4. "Program description".)

A broad range of texts covering different genres and topics was selected to motivate the young readers; the social interaction involved in the program was also assumed to enhance their reading motivation (see the Method section for a detailed description of the program). Inasmuch as it provides training in a number of different strategies, the program is consistent with the recommendations of the US National Reading Panel. Based on its meta-analyses, the Panel concluded that simultaneous training in a variety of strategies is the most effective way to foster text comprehension. Some of the most successful approaches identified were training comprehension monitoring processes (metacognition), working on questions with direct feedback, generating questions on various aspects of a story, summarizing central ideas, and cooperative learning of strategies (NICHD, 2000). All of these approaches are implemented in the Berlin Parent-Child Reading Program.

In sum, the program's conceptual framework combines guided reading aloud with an implicit strategy-training component (strategies are used, but not discussed explicitly), thus bringing together two elements that have proved successful for reading training in other contexts (e.g., NICHD, 2000; Palincsar & Brown, 1984), and makes systematic use of the parents' function as role models. In helping the child to answer the basic and/or deeper elaborative questions and modeling behavior by answering questions him- or herself, the parent adopts the role of the "more knowledgeable other," and thus helps the child to reach his or her zone of proximal development (Vygotsky, 1978) in keeping with the ideas of scaffolding (Wood et al., 1976) and cognitive apprenticeship (Brown, Collins, & Duguid, 1989). Because most parents lack didactic training and reading-related content knowledge, however, their role in an instructional context is a rather difficult one. To counter this problem, the program sessions are highly structured and standardized. Furthermore, parents and children work as equal partners, taking turns to read aloud, answer questions, perform special tasks, and give their partner feedback. This approach fosters a positive working relationship and helps to prevent conflicts.

1.4. Program description

The program consists of 43 sessions, each one lasting 30 minutes, with three sessions being scheduled per week¹. The participating families received all program materials free of charge, along with accompanying instruction booklets informing parents and children about the session structure and providing guidance for program implementation and practical recommendations (e.g., on the time and place of sessions). The parent booklet further provided advice on how to deal with problems such as children’s reading errors. It was recommended that one parent should work through the whole program with the child, but it was also possible for fathers and mothers to alternate.

Essentially, each session consisted of two parts: (a) guided reading aloud (with the additional element of parents as models) and (b) an implicit strategy-training component that focused on metacognitive strategies and elaboration of text content, again with parents as models, and involves a scaffolding component, with parents helping children to answer questions or answering questions themselves. Each individual session was highly structured and standardized comprising: (a) Shared reading aloud of a text for about 15 min to support fluency and vocabulary; the materials indicated whether each text

should be read by the child or the parent. (b) A general metacognitive question to explicitly train comprehension and vocabulary monitoring: “Was there anything in the text that you didn’t understand? Or were there any new words? Try to clear up any problems together before going on.” (c) Three to four basic comprehension questions to ensure basic understanding (propositional text representation) and to implicitly induce metacognitive comprehension monitoring via comprehension monitoring questions (e.g., “Why was Monica [the main character] sad when she and her family moved to Romania?”); again, the materials indicated whether the questions are to be answered by the child or the parent. (d) Conversations based on elaborative questions and tasks (situational text representation; see Table 1) to train elaboration strategies, metacognitive strategies, and vocabulary. (e) Finally, a closing task targeting cognitive strategies or motivation, e.g., summarizing the text or drawing a picture. Comprehension instruction was therefore provided by a set of questions designed to guide children and parents through the process of negotiating the text. At the end of each session, the participating parent and the child each completed a short session protocol collecting basic information about the session, as well as process and motivational variables (see subchapter 2.3. “Measures”).

Table 1: Categories of elaborating questions/tasks

| | Texts | | |
|--|-----------|------------|-------|
| | Narrative | Expository | Total |
| Describing personal emotions or actions from the perspective of the characters in the text | 29 | 4 | 33 |
| Transferring text content to own life (friends, family) | 18 | 18 | 36 |
| Understanding reasons for actions or opinions | 28 | 0 | 28 |
| Detecting special features of the text, actions, or main characters | 20 | 3 | 23 |
| Checking a hypothesis against the text content | 2 | 8 | 10 |
| Philosophical or abstract questions | 3 | 0 | 3 |
| Judging actions or solutions, in the light of their consequences, if applicable | 19 | 0 | 19 |
| Finding, weighting up, and evaluating alternative actions | 16 | 4 | 20 |
| Evaluating the text | 19 | 8 | 27 |
| Thinking about the author’s intention | 19 | 2 | 21 |
| Proposing changes to the text | 3 | 6 | 9 |
| Activating prior knowledge | 13 | 30 | 43 |
| Giving examples of things mentioned in the text | 6 | 9 | 15 |
| Illustrating text content, transforming text content (to a different medium) | 2 | 6 | 8 |
| Formulating questions to the author | 2 | 16 | 18 |
| Predicting the next part of the text | 12 | 0 | 12 |
| Continuing a storyline | 14 | 0 | 14 |

Once the first set of basic comprehension questions had ensured a basic understanding of the text, participants tackled the second set of (open-ended) questions or tasks, which required elaboration and metacognitive monitoring and regulation strategies.

The questions and tasks, which implicitly guided readers to use strategies, included categories such as “understanding reasons for actions or opinions”, “checking a hypothesis against the text content”, “using prior knowledge” or “analyzing the author’s

¹ Example sessions, session protocols, and parent and child manuals are available from the first author.

intention". Table 1 gives an overview of the question categories and how often they were used in the program. Additionally, later on in the program, parents and children were instructed to formulate and answer basic comprehension and elaboration questions of their own in order to support their individual and independent use of the reading-related strategies.

A total of 28 narrative texts (e.g., fairy tales, detective stories, and adventures) and 15 expository texts (on topics such as nature, the universe, and animals) were selected for the program. All texts came from age-appropriate school and children's books. A broad range of topics was covered, such that children could likely draw on prior knowledge for some texts, but not for others. The average text length increased over the course of the program, starting with 452 words in the first 15 sessions, increasing to 550 words in sessions 16 to 30, and to 637 words in sessions 31 to 43. Two different versions of each text (one short version, one long version) were offered to the families to choose from in order to accommodate different reading levels.

1.5. Research questions

Research on home-based early literacy programs indicates that the effectiveness of voluntary programs conducted by parents who do not typically have training in content-related aspects of text comprehension or reading literacy is largely dependent on implementation variables (Olds & Kitzman, 1993; Wagner & Clayton, 1999). We therefore addressed two major areas in the present research: the implementation of the program and its effectiveness.

1.5.1. Implementation

Given the voluntary nature of participation in the program and the special role assigned to parents, the central research questions regarding its implementation were as follows: (a) Are families willing to participate in a voluntary reading program in their free time at home and do they actually complete the program? (b) Is it possible to implement the parent-child program as intended in the conceptual framework in terms of accuracy of implementation, conditions of implementation, and implementation processes?

Because the program built on everyday reading practices, and because the program-specific elements are highly structured, we assumed that it was possible for parents and children to implement the program as intended. Likewise, we expected the discussion of interesting texts within a harmonious social interaction to support families' motivation to continue with the program over time.

1.5.2. Effectiveness

In terms of effectiveness, we investigated whether the parent-child reading program succeeded in fostering

the development of individual facets of reading competence in children of elementary school age (Grade 4). Our central research question was as follows: Does the parent-child reading program have positive effects on key prerequisites for reading in terms of students' (a) fluency, (b) reading-related metacognition, (c) vocabulary, (d) reading motivation, as well as (e) on students' text comprehension?

The program aimed and was expected to foster all of these prerequisites for reading literacy. The development of vocabulary was promoted by the range of texts provided, discussion of unknown words after each text, and basic comprehension questions that checked understanding; fluency was fostered by reading texts aloud at the beginning of each session; and reading-related metacognition was supported by a direct metacognitive question after each text and elaborating questions that stimulated metacognitive processes in joint discussion. Additionally, the program aimed to foster reading motivation by providing interesting, age-appropriate texts, stimulating and guiding deeper-level text elaboration, and embedding reading within a process of social interaction and communication with a parent. As a result of these interventions, the program was expected to promote reading comprehension in general.

1.6. Hypotheses

Apart from expecting children in the treatment group to make greater progress than control children in all the above mentioned prerequisites of reading literacy (Hypothesis 1), we also expected the program—like many programs focusing on training strategies (NICHD, 2000)—to have compensatory effects. In other words, we expected students with particularly low pretest scores on the dependent variables investigated to draw particular benefit from participation. By contrast, we did not expect to observe a Matthew effect (Stanovich, 1986); in other words, we did not expect students who scored higher at pretest to draw greater benefit from program participation, because we assumed that these good readers would already be using various reading strategies effectively (Hypothesis 2).

2. Method

2.1. Design

The study ran from August 2003 to January 2004 (see Figure 1). A quasi-experimental design was implemented with students from 32 fourth-grade classes in Berlin, Germany, and their families. The classes were located in different parts of Berlin, ensuring a broad range of social backgrounds. A family program can only be implemented with the consent of the families; therefore, the families of all children in these classes were informed about the program in writing and invited to participate. All families volunteering for the program were accepted

in the treatment group, with the others forming the control group. Children in the treatment group participated in the parent-child reading program in the family; children in the control group (from the same classes) were not exposed to any special treatment. Pre- and posttests assessing reading literacy and prerequisites of reading were administered to both groups directly before and after the four-month implementation phase.

Additionally, parents and children in the treatment group completed a session protocol after each reading session and post-questionnaires at the end of the program, and one session per parent-child dyad was videotaped at the beginning and the end of the intervention period (for a detailed description of the video analyses see McElvany, 2008). Students' teachers were not involved in the program, and they were not informed about their students' participation.

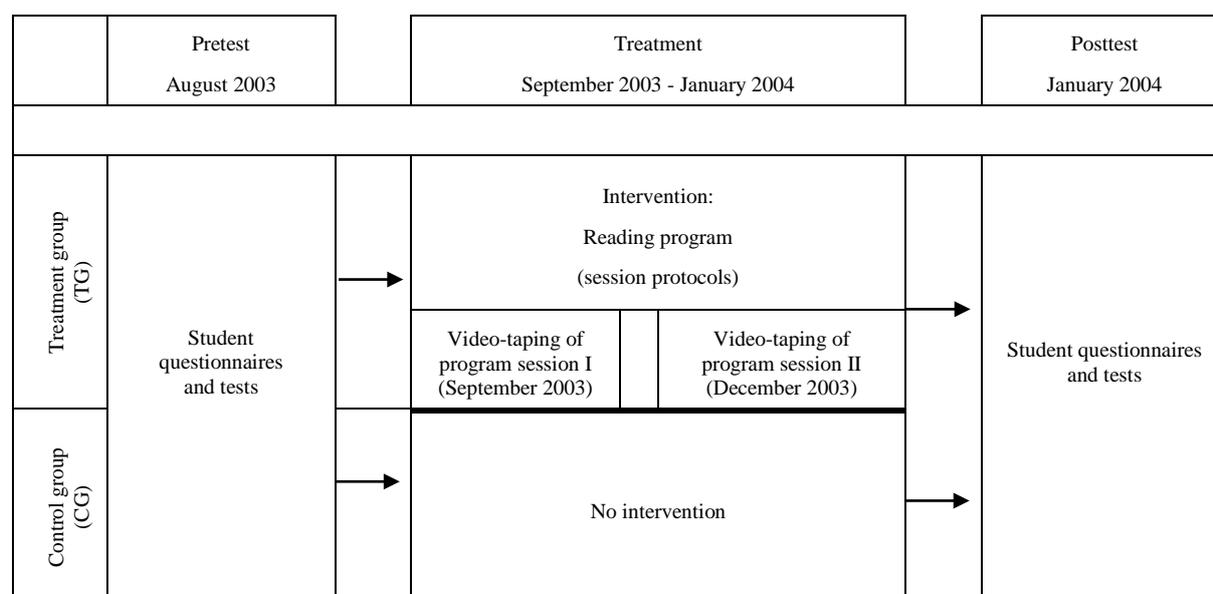


Fig. 1. Design overview.

2.2. Participants

Students were defined as program participants if the protocols returned showed that they had completed more than 15 of the 43 intended sessions. This cut-off point was chosen because we wanted to ensure that the families included had participated in the program for at least one month (at the rate of three sessions per week). Of the initial 665 Grade 4 students participating in the study, 116 were in the treatment group and 393 in the control group. A further 156 of the 665 students were excluded from the effectiveness analyses comparing the treatment group and the control group, either because they dropped out of the program after completing less than a third of the

sessions ($n = 44$) or because, although their families had indicated interest in participating in the program and received program material, they did not send back any protocols and therefore could not be clearly categorized as belonging to either of the two groups ($n = 112$). (See also the Results section for a comparison of the different groups; e.g., in terms of the families' socioeconomic status.) Finally, 12 students from the treatment group had to be excluded after the treatment check (see below).

The mean age of the students was 9.31 years ($SD = .61$). Boys were slightly over-represented in the control group (53%) and the treatment group (52%). For the sample overview see Table 2.

Table 2: Sample overview

| | Treatment group | Control group | Unidentified/excluded group | Total |
|-----------------------------|------------------|---------------|-----------------------------|------------|
| <i>N</i> | 104 ^a | 393 | 168 | 665 |
| Age: <i>M</i> (<i>SD</i>) | 9.12 (.58) | 9.38 (.62) | 9.25 (.53) | 9.30 (.60) |
| Girls: percentage | 45 | 48 | 52 | 48 |

^a *N* for analyses of effectiveness (after treatment check); initial total treatment group: $n = 116$.

2.3. Measures

2.3.1. Implementation

Families' willingness to participate and dropout rates were measured on the basis of registration figures and the session protocols in which participants recorded whether or not each session was conducted. Accuracy of implementation and implementation conditions were measured by reference to the session protocols and post-program questionnaires, which provided data on the number, frequency, and duration of sessions, session participants, and implementation processes. The following three scales from the session protocols assessed implementation processes.

2.3.1.1. Collaboration/working process from the parental perspective. After each session, parents rated the quality of collaboration and working process on a 5-point Likert-scale, with three items measuring engagement in joint discussion, the child's active cooperation, and the quality of the joint working process. The reliability² was ICC(2) = .94 (M = 4.12, SD = .46). An ICC of this magnitude indicates that it is possible to calculate the average scale score for all sessions rated for each family.

2.3.1.2. Structures and processes from the child's perspective. In their session protocols, children rated five items on a 5-point Likert-scale, providing data on the quality of the joint discussion, the comprehensibility of parental statements, and children's satisfaction with the answers given to their own comprehension queries, as well as the answers that they and their parents gave to the text comprehension questions, ICC(2) = .96 (M = 4.51, SD = .40).

2.3.1.3. Difficulties. To assess any difficulties encountered during the session, parents rated three items measuring the difficulty level of the text and the questions/tasks, as well as any problems they had understanding the child's verbal contributions, on a 5-point Likert-scale, ICC(2) = .92 (M = 1.69, SD = .38).

2.3.2. Effectiveness

In our pre-post design, students' reading-related skills were measured directly before and after implementation of the reading program (i.e., in August 2003 and January 2004). At each point of measurement, vocabulary was assessed using 15 items from the Vocabulary subtest of the German version of the Culture Fair Intelligence Test (Weiß, 1987; Cronbach's α = .70 and .74 for the two testing occasions, respectively).

Fluency was evaluated by a speeded 70-item multiple-choice test (Würzburger Leise Leseprobe [Würzburg Silent Reading Test]; Küspert & Schneider, 1998) that required one of four pictures to be matched to a given word, e.g., "foot" or

"thermometer". Different item sets were administered at pre-and posttest. For test security reasons, two versions of each test were used at each point of measurement, differing only in the order of the items.

Metacognition was measured in terms of declarative metacognitive knowledge about reading strategies, using six of the nine reading-related items from the Würzburger Testbatterie für Deklaratives Metagedächtnis (Würzburg Test Battery for Declarative Metamemory; Schlagmüller, Visé, & Schneider, 2001). In one task students rated three proposed strategies for dealing with a reading-related task by grading each strategy on a scale from 1 to 6 (1 being the best). Their ratings were scored based on the rating scheme provided by the test publishers. The scoring rationale is based on comparative instead of absolute correctness of student ratings. Grading the better of two compared strategies with better grade was scored with 2 points, grading the strategies as the same was scored with 1 point, and grading the better as worse was given no point. For example, regarding the task to remember a story just read, rating "Paying attention to the sentences which seem important to you" as better than "Writing the first word of each sentence in your notebook" was scored with 2 points. In two other tasks students had to choose the best strategy for dealing with a reading-related task out of three strategies proposed. Points were given for choosing the correct best strategy and additionally for not choosing the worst strategy. The number of items was reduced to achieve acceptable reliability. Cronbach's α was .70 and .74 for the two testing occasions, respectively.

Reading motivation was assessed by using a newly developed scale comprising five items such as "I have fun reading". Agreement with each of the items was rated on a 4-point self-report Likert scale from 1 (disagree completely) to 4 (agree completely). The mean of all items was calculated for the scale. Cronbach's α was .85 and .89 for the two testing occasions, respectively.

Text comprehension was assessed by using 20 multiple-choice items from the Text Comprehension subtest of the Hamburger Lesetest für 3. bis 4. Klassen (HAMLET 3-4) (Hamburg Reading Test for Grades 3 to 4; Lehmann, Peek, & Poerschke, 1997). Five texts with questions requiring skills ranging from simple decoding to independent reasoning were administered at each occasion of measurement; here again, the order of presentation of the texts was varied across participants. Texts and items were selected on the basis of the item difficulty documented in the HAMLET 3-4 test manual, allowing a broad range of difficulty to be covered. Achievement scores were estimated using Item Response Theory (IRT), as in the original HAMLET 3-4 test. IRT scaling makes it possible to establish a common metric for both times

² Intra-class correlation (ICC) was used as a measure of reliability.

of measurement (Hambleton, Swaminathan, & Rogers, 1991). We used the single parameter variant (Rasch model or latent trait model) with fixed item difficulty in ConQuest (Wu, Adams, & Wilson, 1998). The resulting logits for the person parameter were later transformed to allow the absolute change between the measurement points to be interpreted relative to a mean of 500 (SD = 100).

2.4. Multiple imputation of missing values

The percentage of data missing from the total sample at pre- and post-test was 13.1% and 13.0%, respectively. Values representing an implausibly high gain or loss in achievement scores (criterion: ± 2 SD), indicating that these children did not comply properly with the tests, were also set as missing. To generate five imputations for all relevant indicators NORM 2.03 (Schafer, 2000) was used. The advantage of multiple imputation over single imputation is that it avoids the problem of limited variance of estimates (Graham, Cumsille, & Elek-Fisk, 2003). The model specified took into account variables relating to achievement, motivation, and metacognition, as well as gender, age, language spoken at home, socioeconomic background, and group membership (treatment group vs. control group). All subsequent analyses using imputed data were performed with Mplus (Muthén & Muthén, 2001), making it possible to simultaneously analyze all five data sets and to produce a combined set of results according to Rubin (1987).

2.5. Treatment check

Before analyzing the program effects, we had to be certain that the total variability in the implementation of the program elements was small enough to be defined as a single treatment, despite the individual variations in implementation in the parent-child dyads (assumption of homogeneity). We checked that the following criteria were met by reference to the session protocols and videos: (1) compliance with the general program structure, (2) completion of at least one-third of the sessions, (3) average session length of at least 20 minutes (cf. intended length: approx. 30 minutes), and (4) sufficient command of the German language by parents and children. Twelve parent-child dyads were excluded from the analyses of effectiveness as a result of this treatment check, giving a sample of 104 students in the treatment group (see Table 1).

3. Results

3.1. Implementation

The first research question addressed was whether families were willing to participate in a voluntary reading program in their free time and whether they actually completed it. The overall participation rate turned out to be low: 34% of all families approached by letter registered for participation, and only 13% of all families conducted the program as intended.

Comparison of the treatment group and the control group (families who did not register) revealed a number of differences. Compared to the control group, children in the self-selected treatment group came from families with higher socioeconomic status, $t(495) = 5.04$, $p < .05$, Cohen's $d = .61$. They also scored significantly higher on measures of fluency, $t(495) = 2.91$, $p < .05$, Cohen's $d = .31$, of vocabulary, $t(495) = 2.23$, $p < .05$, Cohen's $d = .24$, of text comprehension, $t(495) = 4.51$, $p < .05$, Cohen's $d = .53$, at pretest, and reported spending significantly more of their free time reading, $t(495) = 3.19$, $p < .05$, Cohen's $d = .35$. In view of these differences, the pretest scores had to be taken into account in the analyses of effectiveness. Propensity score matching methods were applied to investigate whether it was legitimate to compare the treatment group and control group despite their differences at pretest (see below).

In a second group comparison, we investigated potential differences between the treatment group and the group that had to be excluded from the analyses of effectiveness due to the treatment check ($n = 12$), or because the families either dropped out of the program ($n = 44$), or could not be clearly categorized as belonging to the treatment group or the control group because they had not submitted session protocols ($n = 112$). No significant differences were found between the excluded group and the treatment group in terms of fluency, metacognition, vocabulary or reading motivation. There was, however, a small effect for text comprehension in favor of the treatment group, $t(270) = 2.62$, $p < .05$, Cohen's $d = .34$. Free time reading, $t(270) = 3.07$, $p < .05$, Cohen's $d = .38$, and socioeconomic background, $t(270) = 2.88$, $p < .05$, Cohen's $d = .38$, were significantly lower in the excluded group, while the percentage of single-parent homes was higher (36 % in the excluded group vs. 26 % in the reading group). Analysis of session protocols did not provide any indication that the implementation processes and acceptance of the program were significantly less positive in the 44 families who dropped out of the program than in the group of continuers.

We now turn to our second research question on program implementation: Is it possible to implement the parent-child program as intended in the conceptual framework in terms of accuracy of implementation, conditions of implementation, and implementation processes? With respect to accuracy of implementation, analyses of session protocols and post-program questionnaires indicated that, on average, families conducted 35 of the 43 sessions scheduled (SD = 8)—an implementation rate of 81.4%. Moreover, 31% of families did not skip a single session. In line with the recommendations given in the program manual, families spent an average of 30.09 minutes (SD = 8.38) on each session. In contrast to the recommendations, however, 77.9% of the families reported that they did not conduct three sessions per week. Instead, most families tended to complete two sessions per week,

thereby extending the total duration of the program. Regarding the conditions of implementation, children worked with one parent as recommended in 94.8% of cases. However, 23.3% of families reported that other persons were often present during the sessions, although they had been advised to find a time and place for the child and one parent to work together alone. The three scales in the session protocols assessing the implementation processes indicated (a) that the collaboration/ working process was judged to

be successful ($M = 4.12$, $SD = .46$) on a scale from 1 (not at all) to 5 (very) from the parental perspective, (b) that, on average, the children perceived the structures and processes of the program positively ($M = 4.51$, $SD = .40$) on a scale from 1 (not at all) to 5 (very), and (c) that few difficulties were reported by parents ($M = 1.69$; $SD = .38$) on a scale from 1 (not at all) to 5 (very). Figure 2 illustrates the average ratings of the 43 sessions on these three scales.

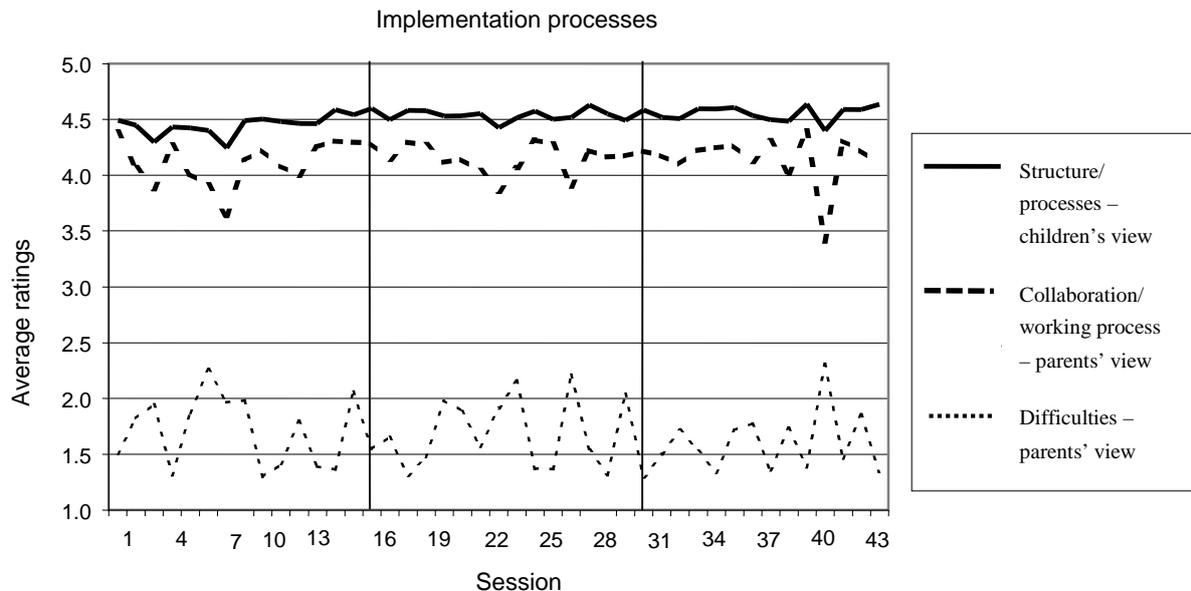


Fig. 2. Participants' ratings of implementation processes (average ratings of the 43 sessions) on a scale from 1 (not at all) to 5 (very).

3.2. Effectiveness

3.2.1. Method of analysis

As shown by the results of the implementation analyses, the problem of endogeneity in the present self-selected sample led to differences between the treatment and control groups at pretest in terms of reading-related skills and important background variables. Two approaches were utilized to take these differences into account: (a) A propensity score matching approach was used to investigate whether the treatment and control group populations matched in terms of their likelihood to participate in the program based on their reading skills and background variables. (b) Regression analyses were used to analyze program effects.

3.2.1.1. Regression analyses. Because the treatment group and the control group are considered as two different populations, we specified a sequence of regression models with and without control variables rather than repeated measures analysis of variance (cf. Campbell & Kenny, 1999; Maxwell & Delaney, 2004; Plewis, 1985; for a critical discussion of the ANCOVA approach see Miller & Chapman, 2001). In contrast to the repeated measures of variance

approach, which focuses on differences (gains or losses) in measures of change, the focus of the regression approach is on predicting changes at posttest that go beyond what can be predicted on the basis of the pretest score. Adjusted means of the posttest are compared, controlling for different pretest scores, thereby also avoiding the regression towards the mean effect, by directly comparing the post-program scores of the treatment and control group rather than comparing the pre- and posttest differences between the two groups.

In the first regression model (Model I) only the pretest score and the dichotomous group variable (treatment/control) as predictors of the dependent variable were taken into account. In the second regression model (Model II) pretest text comprehension scores and children's reading behavior (amount of reading in their free time) were further included as control variables. To estimate the regression coefficients and to determine their significance at the 5% level Mplus (Muthén & Muthén, 2001) was used. The R^2 statistic represents the explained variance in the dependent variable at posttest in each model. In addition to p-values, we report the conservative effect sizes (d) as the ratio

between the nonstandardized regression coefficient of the group variable and the root of the posttest variance of the dependent variable (Long, 1997) to allow better evaluation of the effects calculated. In this context, a positive algebraic sign indicates a more positive development for the treatment group than for the control group³.

In a next stage, the interaction term of group membership and the pretest score on the dependent variable under investigation was added to regression Models I and II as a means of investigating whether the program had differential effects for children with low vs. high pretest scores on the dependent variable (Models III and IV; cf. Aiken & West, 1991). A

significant interaction term would indicate that the benefit children draw from program participation differs depending on their individual pretest score on the dependent variable.

3.2.1.2. Results of propensity score matching. The propensity score matching analysis indicated that findings cannot be generalized across the whole population, given the different reading skills and family backgrounds of the participating children. Children with a low propensity score for likelihood of participation (the lowest quintile of the distribution) were underrepresented in the treatment group (see Table 3).

Table 3: Number of children per quintile of the propensity score matching in the treatment vs. control group

| Quintile | Treatment group | Control group |
|----------|-----------------|---------------|
| 1 | 5 | 94 |
| 2 | 9 | 91 |
| 3 | 14 | 85 |
| 4 | 25 | 75 |
| 5 | 50 | 49 |

This means that the results of the analyses of program effectiveness that follow cannot be generalized to students with low propensity scores, i.e., to students with the lowest achievement scores at pretest and with a less favorable family background than their peers.

3.3. Program participation and vocabulary

The results of the regression models showed the expected main effect of participation in the reading program on vocabulary skills. The nonstandardized regression coefficient (*B*) of the dichotomous group variable was significant, for Model II $t = 3.34$, $p < .05$ (see Table 4).

Table 4: Regression models for vocabulary

| | Model I | | Model II | | Model III | | Model IV | |
|--------------------------|----------|-----------------|------------------|-----------------|-----------|-----------------|----------|-----------------|
| | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) |
| <i>Program effect</i> | | | | | | | | |
| Group | 1.14*** | .25 | .84*** | .25 | 1.14*** | .25 | .85*** | .25 |
| <i>Pretest score</i> | | | | | | | | |
| Vocabulary | .71*** | .03 | .58*** | .04 | .71*** | .04 | .58*** | .04 |
| <i>Interaction term</i> | | | | | | | | |
| Group x pretest score | | | | | -.00 | .09 | -.04 | .09 |
| <i>Control variables</i> | | | | | | | | |
| Text comprehension | | | .01*** | .00 | | | .01*** | .00 |
| Reading behavior | | | .21 ⁺ | .12 | | | .21 | .12 |
| <i>d</i> (group) | | .49 | | .36 | | .49 | | .36 |
| <i>R</i> ² | | .53 | | .56 | | .53 | | .56 |

Vocabulary scores range from 0 to 15; *B* = nonstandardized regression coefficient; SE (*B*) = standard error of *B*; *R*² = multiple determination coefficient; *** $p < .001$; ⁺ marginally significant at the 10% level.

³ Due to the multiple imputation we applied on the data using MPlus, only nonstandardized (*B*) values are reported along with *SE* of *B*.

Children who participated in the program showed higher gains in vocabulary over the first half of the fourth grade than children in the control group, even when control variables were taken into account ($d = .36$). No significant interaction term of the group variable with the vocabulary score at pretest was found, meaning that the positive program effect on vocabulary was comparable for children with different levels of pretest knowledge, and that there were no differential effects.

3.4. Program participation and fluency

Contrary to our prediction, no differences were detected between the two groups in terms of development in word fluency, for Model II, $t = -0.30$, $p > .05$ (see Table 5). The nonstandardized regression coefficient (B) of the dichotomous group variable was not significant when controlling for text comprehension and reading behavior at pretest.

Table 5: Regression models for fluency

| | Model I | | Model II | | Model III | | Model IV | |
|--------------------------|---------|------------|----------|------------|-----------|------------|----------|------------|
| | B | SE (B) | B | SE (B) | B | SE (B) | B | SE (B) |
| <i>Program effect</i> | | | | | | | | |
| Group | .44 | .86 | -.26 | .85 | .32 | .88 | -.36 | .87 |
| <i>Pretest score</i> | | | | | | | | |
| Fluency | .92*** | .03 | .85*** | .04 | .91*** | .03 | .84*** | .04 |
| <i>Interaction term</i> | | | | | | | | |
| Group x pretest score | | | | | .05 | .08 | .05 | .08 |
| <i>Control variables</i> | | | | | | | | |
| Text comprehension | | | .02*** | .00 | | | .02*** | .00 |
| Reading behavior | | | .58 | .40 | | | .55 | .40 |
| d (group) | .06 | | -.04 | | .04 | | -.05 | |
| R^2 | .69 | | .70 | | .69 | | .70 | |

Fluency scores range from 0 to 15; B = nonstandardized regression coefficient; SE (B) = standard error of B ; R^2 = multiple determination coefficient; *** $p < .001$.

These findings indicate that the program did not succeed in significantly enhancing fourth graders' fluency in decoding words. A differential, compensatory effect was not observed either.

3.5. Program participation and reading-related metacognition

In terms of reading-related metacognition (see Table 6), the regression model showed a significant effect of program participation, with students in the treatment group showing more pronounced gains ($d = .27$). The effect was still marginally significant, for Model II, $t = 1.64$, $p < .10$, after controlling for text comprehension and reading behavior at pretest, with a reduced effect size of $d = .15$. The analyses with the interaction term ($B = -.30$) and with control variables ($B = -.28$) indicated that the program's effects on reading-related metacognition were based on its power to foster weaker students' skills, that is, students with lower reading-related metacognition at pretest drew particular benefit from the program.

3.6. Program participation and reading motivation

As regards reading motivation, no significant program effects emerged in the regression models, whether or not the control variables were taken into account, for Model II, $t = -.02$, $p > .05$ (see Table 7). Reading motivation remained stable in both groups over time, as shown by t-tests for the adjusted means at pre- and posttest. A significant differential effect was not found either.

3.7. Program participation and text comprehension

Again, no significant program effects emerged in the regression models, whether or not the control variables were taken into account, for Model II, $t = .63$, $p > .05$ (see Table 8), and no significant interaction term was found. Contrary to our prediction, then, no treatment effect was found in terms of a more pronounced increase in text comprehension in students who participated in the program.

Table 6: Regression models for metacognition

| | Model I | | Model II | | Model III | | Model IV | |
|--------------------------|----------|-----------------|------------------|-----------------|-----------|-----------------|----------|-----------------|
| | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) |
| <i>Program effect</i> | | | | | | | | |
| Group | .78** | .28 | .45 ⁺ | .27 | .89** | .28 | .57* | .27 |
| <i>Pretest score</i> | | | | | | | | |
| Metacognition | .51*** | .04 | .43*** | .04 | .57*** | .04 | .49*** | .05 |
| <i>Interaction term</i> | | | | | | | | |
| Group x pretest score | | | | | -.30* | .09 | -.28*** | .08 |
| <i>Control variables</i> | | | | | | | | |
| Text comprehension | | | .01*** | .00 | | | .01*** | .00 |
| Reading behavior | | | .00 | .14 | | | -.01 | .13 |
| <i>d</i> (group) | | .27 | | .15 | | .30 | | .19 |
| <i>R</i> ² | | .33 | | .38 | | .35 | | .39 |

Metacognition scores range from 0 to 15; *B* = nonstandardized regression coefficient; SE (*B*) = standard error of *B*; *R*² = multiple determination coefficient; *** *p* < .001.

Table 7: Regression models for motivation

| | Model I | | Model II | | Model III | | Model IV | |
|--------------------------|----------|-----------------|----------|-----------------|-----------|-----------------|----------|-----------------|
| | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) |
| <i>Program effect</i> | | | | | | | | |
| Group | .05 | .06 | .00 | .07 | .05 | .07 | .00 | .07 |
| <i>Pretest score</i> | | | | | | | | |
| Metacognition | .69*** | .04 | .53*** | .05 | .69*** | .04 | .54*** | .05 |
| <i>Interaction term</i> | | | | | | | | |
| Group x pretest score | | | | | -.02 | .10 | -.06 | .10 |
| <i>Control variables</i> | | | | | | | | |
| Text comprehension | | | .00 | .00 | | | .00 | .00 |
| Reading behavior | | | .18*** | .04 | | | .18*** | .04 |
| <i>d</i> (group) | | .08 | | .00 | | .08 | | .00 |
| <i>R</i> ² | | .42 | | .45 | | .42 | | .45 |

Motivation scores range from 0 to 15; *B* = nonstandardized regression coefficient; SE (*B*) = standard error of *B*; *R*² = multiple determination coefficient; *** *p* < .001

Table 8: Regression models for text comprehension

| | Model I | | Model II | | Model III | | Model IV | |
|--------------------------|----------|-----------------|----------|-----------------|-----------|-----------------|----------|-----------------|
| | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) | <i>B</i> | SE (<i>B</i>) |
| <i>Program effect</i> | | | | | | | | |
| Group | 5.94 | 6.94 | 4.44 | 6.90 | 8.03 | 7.02 | 6.44 | 6.97 |
| <i>Pretest score</i> | | | | | | | | |
| Text comprehension | .53*** | .04 | .51*** | .03 | .55*** | .04 | .53*** | .04 |
| <i>Interaction term</i> | | | | | | | | |
| Group x pretest score | | | | | -.08 | .07 | -.07 | .07 |
| <i>Control variables</i> | | | | | | | | |
| Reading behavior | | | 9.37** | 3.38 | | | 9.31** | 3.37 |
| <i>d</i> (group) | | .08 | | .06 | | .11 | | .09 |
| <i>R</i> ² | | .44 | | .45 | | .44 | | .45 |

Motivation scores range from 0 to 15; *B* = nonstandardized regression coefficient; SE (*B*) = standard error of *B*; *R*² = multiple determination coefficient; *** *p* < .001

4. Discussion

We begin by summarizing our findings on program implementation. Overall, families' willingness to participate in the voluntary reading program was low (34%), and only one-third of the families who registered to participate actually conducted the program as intended. Moreover, participation was selective based on family background and reading skills. Nevertheless, analyses of the session protocols and the post-program questionnaire indicated that, if small modifications to the procedure are made, it is generally possible to implement a parent-child reading training program that is successful in terms of accuracy of implementation, conditions of implementation, and implementation processes.

Our findings on the development and implementation of the Berlin Parent-Child Reading Program add to the existing research on approaches to foster reading in the family setting. The program was designed to foster specific skills, and both quantitative and qualitative measures were used to evaluate its effects in the areas of skills, motivation, and family reading culture. Our evaluation of the program's effectiveness showed that participation promoted the development of vocabulary and reading-related metacognition. Contrary to Hypothesis 1, however, no significant main effects were found for fluency (decoding speed), motivation, or text comprehension. Additional analyses of the effects on vocabulary and metacognition showed that there was a differential and compensatory effect for reading-related metacognition, with children with low metacognition scores at pretest benefiting most from program participation. No differential effect was found for vocabulary development, however. These findings partially confirmed Hypothesis 2.

Effect sizes between $d = .15$ and $d = .49$ indicate that the gains resulting from program participation were substantial. Relative to average Grade 4 achievement gains, the additional gains observed are equivalent to three-quarters of a school year for vocabulary and approximately half a school year for metacognition, as comparison with the development of the control group in the first half of Grade 4 showed (McElvany, 2008; McElvany & Artelt, 2007). Let us not forget, however, that propensity score matching indicated that the generalizability of the analyses of effectiveness for students with very poor achievement scores and from less privileged socioeconomic backgrounds is limited because of their low participation in this study. Given the voluntary nature of the program, moreover, the results are limited to those families where adults would volunteer. This first evaluation of the program (see also McElvany, 2008) identified some of the strengths and weaknesses of voluntary training programs in the home context. One of the major issues to be discussed in the light of the present results is the problem of selective participation. Self-selection of families occurs at program registration, and again when families decide at some point either to continue with the program or to drop out. Our methodological approach to pre-program differences between the treatment and control groups

was to use a regression model with control variables. Based on the present results, the likely participation and dropout rates of families in future training programs can be estimated, and a waiting-group design implemented, with families being randomly assigned to treatment or control groups to ensure comparability of groups and internal validity. Another important issue for future research, namely to raise participation levels, is sensitivity to the children's social and cultural backgrounds when designing family program materials.

From a normative point of view, it might be argued that, given the problems of selectivity in program participation, the parent-child reading program in its present form failed to meet the important challenge of supporting students with reading deficits, but instead catered for proficient readers from socially privileged backgrounds, thereby increasing social disparities in competence acquisition and educational attainment (see Bourdieu, 1984). However, this argument is qualified by the finding that children and families from less privileged individual or social backgrounds drew equal or even greater benefit from program participation than children and families from more privileged backgrounds. The question to be asked, therefore, is how more families with less favorable socioeconomic characteristics can be motivated to participate in programs of this kind and how dropout rate can be reduced. Ideas that might be considered include target-group-specific (initial) contact, target-group-specific materials, the provision of incentives and advice for participating families, supervision during the program, parental training prior to the program, and the involvement of schools. Although the program elements of extensive text reading, discussion of unknown words, and elaborative parent/child dialogue led to an increase in vocabulary, as expected, the lack of a program effect on text comprehension warrants discussion. The treatment was designed to help students acquire strategy routines for dealing with written texts by checking understanding (metacognitive questions after the text and basic comprehension questions) and elaborating content (elaborative questions and tasks). We had therefore expected text comprehension skills to increase, but they did not. One possible explanation for this finding is that the skills measured by the reading test administered at the end of the program (HAMLET 3-4; Lehmann, Peek, & Poerschke, 1997) were not addressed directly in the reading program. The program focused on the implicit training of elaboration strategies, with additional questions to ensure the children's basic understanding of the text. Over the course of the program, children and parents learnt to communicate verbally about texts and to elaborate their content. As such, the tasks completed during the joint reading sessions differed markedly from those administered in the final comprehension test, which required the children to answer multiple-choice questions on their own and under time pressure. Hence, our global indicator of text comprehension might not in fact be suitable for evaluating changes resulting from the reading program. The numerous

differences between the program tasks and the test tasks—elaboration (program) versus local coherence building (test), oral communication (program) versus written work (test), social interaction (program) versus individual work (test), free allocation of time (program) versus time limit (test)—increase the transfer skills required during the evaluation of text comprehension. Furthermore, we cannot exclude the possibility that students not participating in the program, who were aware that their classmates had been given extra reading training, might have made a special effort in the posttest to counter this “disadvantage.” Another reason for the lack of an effect could be the thematic breadth of the texts used or the range of elaboration strategies introduced (see Table 1). Moreover, it seems worth investigating potential mid- or long-term effects of program participation on the development of text comprehension, in particular. These may occur as individual reading processes become habitualized, or as the families’ reading culture is nurtured by the program. Potential mid-term effects on the development of reading motivation and reading behavior should also be investigated. It can be assumed that program effects in these areas only develop some time later, when children are no longer compelled to read on a regular basis as part of the program.

Overall, our results provide important insights into the effectiveness of the newly developed parent-child reading program, indicating that its conceptual framework (guided reading aloud and implicit strategy instruction in social interaction between parent and child) is a promising basis for designing training programs that can be successfully and effectively implemented within the family context and can promote the development of important prerequisites for reading literacy. In a two-year follow-up study, we are currently investigating whether the program had effects on the development of the families’ reading culture (shared reading time, parents’ reading behavior, parents’ sense of responsibility as regards their children’s reading development), as well as on the children’s cognitive skills and motivational attitudes towards reading.

Future research in the area of systematic family reading programs needs to focus first of all on ways to tackle the problem of selective participation and on the systematic evaluation of individual components of the program (e.g., implicit or explicit strategy instruction or the choice of strategies introduced). A combination of field and laboratory research settings is therefore recommended for future studies, as is the investigation of mid- and long-term effects of participation in family-based training programs.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life: Student approaches to learning. Results from PISA 2000*. Paris: OECD.
- Artelt, C., McElvany, N., Christmann, U., Richter, T., Groeben, N., Köster, J., Schneider, W., Stanat, P., Ostermeier, C., Schiefele, U., Valtin, R., & Ring, K. (2005). *Förderung von Lesekompetenz: Eine Expertise [Promotion of reading competence: An expert report]*. Bonn: Bundesministerium für Bildung und Forschung.
- Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, M. Kamil, R. Barr, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353 - 394). New York: Longman.
- Blok, H., Fukkink, R. G., Gebhardt, E. C., & Leseman, P. P. M. (2005). The relevance of delivery mode and other programme characteristics for the effectiveness of early childhood intervention. *International Journal of Behavioral Development, 29*(1), 35-47.
- Borkowski, J. G., Milstead, M., & Hale, C. (1988). Components of children’s metamemory: Implications for strategy generalization. In F. E. Weinert & M. Perlmutter (Eds.), *Memory development: Universal changes and individual differences* (pp. 73-100). Hillsdale, NJ: Erlbaum.
- Bourdieu, P. (1984). *Distinction: A social critique of the judgement of taste* (R. Nice, Trans.). London: Routledge.
- Bradley, R. H., Corwyn, R. F., Pipes McAdoo, H., & García Coll, C. (2001). The home environments of children in the United States Part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*(6), 1844-1867.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-41.
- Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*(1), 1-21.
- Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artefacts*. New York: Guilford.
- Campione, J. C., & Armbruster, B. B. (1985). In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Acquiring information from texts: An analysis of four approaches, Vol. 1* (pp. 317-362). Hillsdale, NJ: Erlbaum.
- Centre for Community Child Health. (2004). *Let’s read*. Literature review. Melbourne, Australia: Author. Retrieved 8.01.08 from the World Wide Web: http://www.rch.org.au/emplibrary/ccch/LetsRead_LitReview.pdf
- Collins, A., Brown, J. L., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Collins Block, C., & Pressley, M. (2002). *Comprehension instruction: Research-based best practices*. New York: Guilford.
- Elley, W. B. (1994). *The IEA Study of reading literacy: Achievement and instruction in thirty-two school systems*. Exeter, UK: Pergamon.
- Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57*, 72-84.
- Fishel, M., & Ramirez, L. (2005). Evidence-based parent involvement interventions with school-aged children. *School Psychology Quarterly, 20*, 371-402.

- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual Review of Psychology*, 48, 163-189.
- Graham, J. W., Cumsille, P. E., & Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology*, Vol 2 (pp. 87-114). New York: Wiley.
- Grolnick, W. S. (Ed.). (2003). *The psychology of parental control: How well-meant parenting backfires*. Mahwah, NJ: Erlbaum.
- Gunn, B. K., Simmons, D. C., & Kameenui, E. J. (2000). *Emergent literacy: Synthesis of the research*. Eugene, OR: National Center to Improve the Tools of Educators.
- Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., Scaffidi, N. T., & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, 96(3), 403-323.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.
- Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge, UK: Cambridge University Press.
- Hurrelmann, B. (2004). Informelle Sozialisationsinstanz Familie [The family as an informal socialization agent]. In N. Groeben & B. Hurrelmann (Eds.), *Lesesozialisation in der Mediengesellschaft: Ein Forschungsüberblick* (pp. 169-201). Weinheim, Germany: Juventa.
- Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.
- Jordan, G. E., Snow, C. E., & Porche, M. V. (2000). Project EASE: The effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly*, 35(4), 524-546.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.
- Klicpera, C., & Gasteiger-Klicpera, B. (1993). *Lesen und Schreiben. Entwicklung und Schwierigkeiten. Die Wiener Längsschnittuntersuchung über die Entwicklung, den Verlauf und die Ursachen von Lese- und Schreibschwierigkeiten in der Pflichtschulzeit [Reading and writing. Development and difficulties. The Vienna longitudinal study on the development, trajectory, and causes of difficulties in reading and writing in compulsory education]*. Bern: Huber.
- Küspert, P., & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP) [Würzburg Silent Reading Test]*. Göttingen, Germany: Hogrefe.
- Lehmann, R. H., Peek, R., & Poerschke, J. (1997). Hamburger Lesetest für 3. bis 4. Klassen (HAMLET 3-4) [Hamburg Reading test for Grade 3 to 4 (HAMLET 3-4)]. Weinheim, Germany: Beltz.
- Leseman, P. P. M., & de Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation, and social-emotional quality predicting early reading achievement. *Reading Research Quarterly*, 33, 294-318.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks: Sage.
- Mattingly, D. J., Prislun, R., McKenzie, T. L., Rodriguez, J. L., & Kazary, B. (2002). Evaluating evaluations. The case of parent involvement programs. *Review of Educational Research*, 72(4), 549-576.
- Maxwell, S. E., & Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*, (2nd ed.). Mahwah, NJ: Erlbaum.
- McElvany, N. (2008). *Förderung von Lesekompetenz im Kontext der Familie [Promotion of reading competence in the family context]*. Münster, Germany: Waxmann.
- McElvany, N., & Artelt, C. (2007). Das Berliner Eltern-Kind Leseprogramm: Konzeption und Effekte [The Berlin Parent-Child Reading Program: Design and effects]. *Psychologie in Erziehung und Unterricht*, 4, 312-330.
- Miller, G. M., & Chapman, J. W. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, 110(1), 40-48.
- Morrow, L. M., & Young, J. (1997). A family literacy program connecting school and home: Effects on attitude, motivation, and literacy achievement. *Journal of Educational Psychology*, 89(4), 736-742.
- Mullis, I. V. S., Martin, M. O., Gonzales, E. J., & Kennedy, A. M. (2003). *PIRLS 2001 International Report: IEA's Study of Reading Literacy Achievement in Primary Schools in 35 Countries*. Chestnut Hill, MA: Boston College.
- Muthén, L. K., & Muthén, B. O. (2001). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: "Teaching children to read. An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction"*. Washington, DC: U.S. Government Printing Office.
- Neuman, S. B., & Gallagher, P. (1994). Joining together in literacy learning: Teenage mothers and children. *Reading Research Quarterly*, 29(4), 383-401.
- Organization for Economic Cooperation and Development. (2001). *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD.
- Olds, D. L., & Kitzman, H. (1993). Review of research on home visiting. *The Future of Children*, 3, 51-92.
- Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117-175.
- Plewis, I. (1985). *Analysing change: Measurement and explanation using longitudinal data*. Chichester, UK: Wiley.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hillsdale, NJ: Erlbaum.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review*, 14, 245-302.
- Schafer, J. L. (2000). *NORM: Multiple imputation of multivariate continuous data under a normal model (Version 2.03, Computer-software for Windows 95/98/NT ed.)*. Retrieved 8.01.08 from the World Wide Web. Pennsylvania: State College. <http://www.stat.psu.edu/~jls/misoftwa.html>.

- Schlagmüller, M., Visé, M., & Schneider, W. (2001). Zur Erfassung des Gedächtniswissens bei Grundschulkindern: Konstruktionsprinzipien und empirische Bewährung der Würzburger Testbatterie zum deklarativen Metagedächtnis [Assessing metamemory in elementary school children: Construction and evaluation of the Würzburg Metamemory Test]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 33(2), 91-102.
- Schneider, W. (1989). *Zur Entwicklung des Metagedächtnisses bei Kindern [The development of metamemory in children]*. Bern: Huber.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty*. Mahwah, NJ: Erlbaum.
- Snow, C. E. (1993). Families as social contexts for literacy development. In C. Daiute (Ed.), *The development of literacy through social interaction* (pp. 11-24). San Francisco: Jossey-Bass.
- Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-407.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wasik, B. (2000). *Family literacy: An annotated bibliography*. Chapel Hill, NC: National Center for Early Development & Learning.
- Wagner, M. M., & Clayton, S. L. (1999). The parents as teachers program: Results from two demonstrations. *Future of Children*, 9, 91-115.
- Weinberger, J. (1996). A longitudinal study of children's early literacy experiences at home and later literacy development at home and school. *Journal of Research in Reading*, 19(1), 14-24.
- Weiß, R. H. (1987). *Grundintelligenztest Skala 2 CFT 20 [Basic intelligence test, Scale 2, CFT 20]* (3rd ed.). Göttingen, Germany: Hogrefe.
- White, K. R., Taylor, M. J., & Moss, V. D. (1992). Does research support claims about the benefits of involving parents in early intervention programs? *Review of Educational Research*, 62(1), 91-125.
- Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development*, 69(3), 848-872.
- Wimmer, H., Landerl, K., & Frith, U. (1999). Learning to read German: Normal and impaired acquisition. In M. Harris & G. Hatano (Eds.), *Learning to read and write* (pp. 34-50). Cambridge, UK: Cambridge University Press.
- Wood, C. (2002). Parent-child pre-school activities can affect the development of literacy skills. *Journal of Research in Reading*, 25(3), 241-258.
- Wood, D., Bruner, J. S., & Ross, G. (1976). *The role of tutoring in problem solving*. *Journal of Child Psychology and Psychiatry*, 17, 89-100.
- Wu, M., Adams, R., & Wilson, M. (1998). *ACER ConQuest: Generalised item response modelling software manual*. Camberwell, Australia: ACER