Fostering elementary school children's public speaking skills

A randomized controlled trial

Herbein, Evelin; Golle, Jessika; Tibus, Maike; Schiefer, Julia; Trautwein, Ulrich; Zettler, Ingo

Published in:
Learning and Instruction

DOI:
10.1016/j.learninstruc.2017.10.008

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Learning and Instruction 55 (2018) 158–168

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Fostering elementary school children's public speaking skills: A randomized controlled trial

Evelin Herbein[a,*], Jessika Golle[a], Maike Tibus[a], Julia Schiefer[a], Ulrich Trautwein[a], Ingo Zettler[b]

[a] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Europastr. 6, 72072 Tübingen, Germany
[b] Department of Psychology, University of Copenhagen, Øster Farimagsgade 2a, 1353 Copenhagen K, Denmark

ARTICLE INFO

Keywords:
Enrichment program
Elementary school children
Public speaking
Randomized controlled trial
Video ratings

ABSTRACT

Mastering public speaking is a competence which is already required in elementary school. Surprisingly, however, systematic research on the promotion of public speaking competence among elementary school children is scarce. In order to address this gap, we developed and evaluated a public speaking training for elementary school children. The training, an extracurricular enrichment program, consisted of 12 units covering speech anxiety, nonverbal communication, and comprehensibility. A randomized controlled trial with repeated measures (N = 65 elementary school children) was used to investigate the training effects on public speaking skills and speech anxiety. The dependent variables were assessed via self-ratings (extent of public speaking skills, speech anxiety) and video ratings of a public speech (appropriateness of public speaking skills). Findings revealed positive training effects on public speaking skills overall: Participating in the training elicited more appropriate speeches in terms of nonverbal and organizational skills but did not influence speech anxiety.

1. Introduction

The capability to communicate competently is essential for personal contentment, academic achievement, and professional career success (Morreale & Pearson, 2008). In the context of school and work, competent speakers are more successful in conveying their knowledge, ideas, and opinions. Further, being able to communicate competently can enhance relationships with peers, parents, and teachers (e.g., Hunt, Wright, & Simonds, 2014; Morreale & Pearson, 2008).

One important communicational task is giving an informative public speech, i.e., presenting content to an audience (Schreiber, 2011). Mastering public speeches is acknowledged as a core competence for well-educated students (van Ginkel, Gulikers, Biemans, & Mulder, 2015), and it has been integrated into the educational standards of several countries such as Germany (Kultusministerkonferenz, 2005) or the United States (Common Core State Standards Initiative, 2010). Correspondingly, researchers have called for an early promotion of public speaking skills. Even elementary school children need to give informative public speeches, but the promotion of public speaking skills of this age group has been neglected in research and practice (Hunt et al., 2014). Specifically, the development of a competence model for public speaking in elementary schools is in the beginning stages (Morreale, Cooper, & Perry, 2000; The Swiss Conference of Cantonal Ministers of Education [EDK], 2010), and only very few teaching materials or interventions to foster public speaking skills in this age group are available. Further, the effectiveness of the existing interventions has been investigated only sporadically, and the conducted investigations often chose study designs which might make it challenging to draw causal inferences on their effectiveness.

In the light of the importance of informative public speaking skills even among elementary school children and the lack of available, evaluated interventions, we developed and evaluated a pertinent enrichment program. Specifically, we developed a program that aimed at fostering elementary school children's informative public speaking skills and we evaluated the program via both self-reports and video ratings using a randomized controlled trial with repeated measurements.

1.1. Defining public speaking competence

Scholars in fields such as communication studies, psychology, rhetoric, and speech science have researched public speaking competence (Backlund & Morreale, 2015). Across different approaches, a speaker is perceived as competent if her or his public speech is effective and appropriate. A speech is considered to be effective when the communicational intention, for example, informing an audience, is

* Corresponding author.
E-mail addresses: evelin.herbein@uni-tuebingen.de (E. Herbein), jessika.golle@uni-tuebingen.de (J. Golle), maike.tibus@uni-tuebingen.de (M. Tibus), julia.schiefer@uni-tuebingen.de (J. Schiefer), ulrich.trautwein@uni-tuebingen.de (U. Trautwein), ingo.zettler@psy.ku.dk (I. Zettler).

https://doi.org/10.1016/j.learninstruc.2017.10.008
Received 24 June 2016; Received in revised form 12 October 2017; Accepted 29 October 2017
Available online 22 November 2017

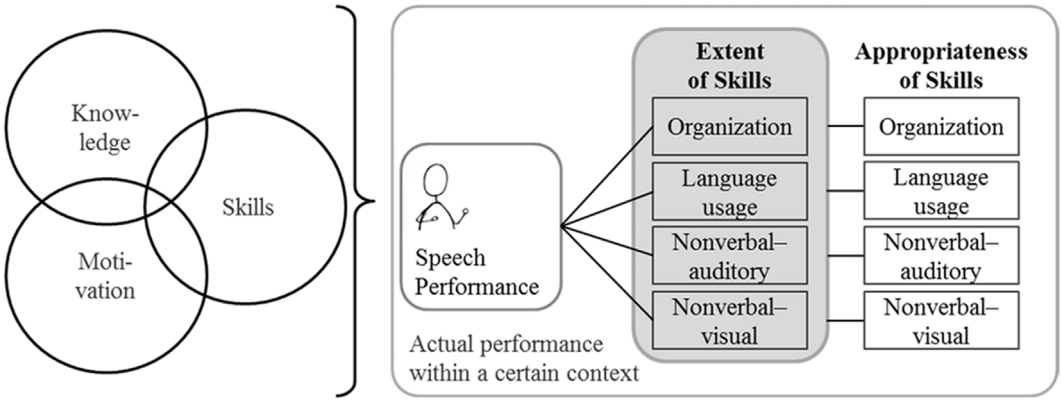

Fig. 1. The three dimensions knowledge, motivation, and skills form the basis for the public speaking competence. Each speech is realized within a certain context. The context is defined by the specific situation, the target audience, the communicational intention, and the topic, and implies certain norms and standards. The speech performance is composed of micro-level skills which can be subdivided into four dimensions: nonverbal behavior–visual, nonverbal behavior–auditory, language usage, and organization. The finally perceived competence of a speaker derives of the shown extent of skills and whether or not a speaker appropriately addresses the context. The figure is based on, e.g., De Grez and Valcke (2010), EDK (2010), Morreale et al. (2013), and Schreiber et al. (2012).

reached. A speech is considered to be appropriate when the speaking behavior is adequate to the specific context (Morreale, Moore, Surges-Tatum, & Webster, 2007). Thus, the appraisal of public speaking competence depends on the speaker's actual demonstrated performance within a specific speaking context.

Although there are several definitions of public speaking competence (Backlund & Morreale, 2015), the construct is often considered to encompass three underlying dimensions: knowledge, motivation, and skills (Morreale et al., 2007). It is suggested that whether or not a person is able to give a public speech competently does not depend on a single dimension, but rather on their combination (Backlund & Morreale, 2015; De Grez & Valcke, 2010). In order to summarize and integrate theoretical and practical approaches from different fields (e.g., De Grez & Valcke, 2010; EDK, 2010; Morreale, Spitzberg, & Barge, 2013; Schreiber, Paul, & Shibley, 2012), a conceptual framework of the underlying mechanisms that influence public speaking competence was derived (Fig. 1). The individual dimensions and their interplay are described in the following text.

Knowledge represents the theoretical background deployed in planning and delivering a speech. It includes knowledge about public speaking skills and their specific utilization (Morreale et al., 2013), e.g., knowing how to structure a speech, using nonverbal communication, and taking the characteristics of a specific audience into account. However, knowledge on how to plan and deliver a speech does not inevitably result in competent behavior (Backlund & Morreale, 2015). This is why public speaking competence is seldom assessed by tests of content (Morreale & Backlund, 2007). *Motivation* represents the willingness to communicate, i.e., the readiness to speak in a specific situation. Several variables underlie and influence motivation, in particular self-perceived communication competence, communication apprehension, and speech anxiety. Whereas self-perceived communication competence is positively correlated with the willingness to communicate, communication apprehension and speech anxiety can result in avoiding certain communicational situations or can generally limit a speaker's ability to show competent public speaking (Croucher, 2013).

Finally, a repertoire of public speaking *skills* is necessary. Descriptions of public speaking skills encompass both macro- and micro-behaviors (Morreale et al., 2007). These behaviors are hierarchically organized. On the micro-level, the particular public speaking skills can be categorized into four dimensions, namely (i) nonverbal behavior–visual impression (including eye contact, gestures, mimics, posture, proxemics i.e., spatial behavior, usage of notes), (ii) nonverbal behavior–auditory impression (including accentuation, articulation, breaks, intonation, volume, pitch, speech fluency, speech rate, speech respiration i.e., respiration used when speaking, voice), (iii) language usage (including activation of the listener, linguistic expression, personal address, usage of rhetorical devices), and (iv) organization (including amount of information, intention of communication, length of speech, length of introduction, length of conclusion, reference to listener, structure of the speech, visualization; based e.g., on De Grez & Valcke, 2010; EDK, 2010; Pabst-Weinschenk, 2005; Schreiber et al., 2012). Macro-level skills are formed by combinations of micro-behaviors such as eye contact, gestures, and speech rate (Morreale et al., 2013). Examples of macro-level skills are the perceived empathy or credibility of a speaker.

In summary, the perceived public speaking competence of a speaker depends on the effectiveness and appropriateness of her/his speech, which arises from actual demonstrated public speaking behavior within a certain context. The context is defined by the specific situation, the target audience, the communicational intention, the topic, norms, and standards. The manifestation of the behavior and whether or not effectiveness and appropriateness are reached depends on the knowledge, motivation, and skills of the speaker (Backlund & Morreale, 2015; van Ginkel et al., 2015). To exhibit competent behavior, a speaker needs a broad repertoire of public speaking skills to choose from. Furthermore, a speaker needs to be willing and able to actually apply and adapt these skills in a way that is guided by the speaker's knowledge and motivation (Backlund & Morreale, 2015). Regarding the development of public speaking competence, a distinction is made between basic and advanced levels. Having gained a certain skills repertoire and being able to apply these skills in a specific way when giving a public speech reflects the basic level of public speaking competence (Rubin & Morreale, 1996). This is mirrored by the extent of skills demonstrated when speaking. Being further able to choose and adapt one's skills in order to tailor one's public speaking performance to the specific speaking context reflects the more advanced level (Staton & Tomlinson, 2001). Thus, the promotion of a skills repertoire and the application of certain skills needs to be followed by the promotion of adequate skills usage to reach an advanced level of competence.

1.2. Public speaking in elementary school

The most frequent public speaking tasks in elementary school are narrating and informing (Common Core State Standards Initiative, 2010; Pabst-Weinschenk, 2005). The task of giving a narration is already quite familiar to elementary school children (Schick & Melzi, 2010), and corresponding interventions have been established (Pesco & Gagné, 2015). By contrast, the task of giving informative speeches is rather new for elementary school children (EDK, 2010). For this age

group, the main goal is to inform peers about new topics. This is not only required in language lessons but also in subjects such as biology, geography, or history. In order to master this challenge, knowledge about the topic, verbal abilities, and public speaking competence are necessary.

The development of a competence model for public speaking at elementary school is vital. Differences across countries or curricula notwithstanding, the following public speaking skills can be considered as key to high public speaking competence of third to sixth graders (Backlund, 1985; Common Core State Standards Initiative, 2010; EDK, 2010; Günther, 2012; Morreale et al., 2000): In a 1-min presentation, students should be able to manage eye contact, gestures, mimics, and posture to some degree (dimension of nonverbal behavior–visual), to speak clearly, fluently, and at an understandable speed (nonverbal behavior–auditory), to use appropriate and proper grammar and vocabulary (language), to use basic structural elements, and to consider audience characteristics (organization).

Though the question of how to promote public speaking competence in K–12 and higher education has gained increasing attention over the last 25 years, research on public speaking competence among elementary and secondary school students—as compared to students in higher education—is still scarce (Hunt et al., 2014). In higher education, the implementation of different design principles has shown positive effects on students' public speaking competence (for an overview see van Ginkel et al., 2015). For example, De Grez, Valcke, and Roozen (2009a, 2009b) found positive effects on body language (nonverbal–visual) as well as length of conclusion, reference to listener, and length of introduction (organization) when the learning objectives were clearly stated, and when behavior modelling, authentic learning tasks, opportunity to practice, and self-assessment were used. In comparison, studies investigating the effectiveness of interventions for younger students are rare.

In terms of elementary school children, a review of the literature elicited, to the best of our knowledge, only one study assessing the effects of a curricular intervention on speaking skills (Wilson, 1997). Concerning the promotion of secondary school students' public speaking skills, a small number of pilot (e.g., Brann-Barrett, 2005; Thielemann, 2014) or evaluation studies (e.g., Fellenberg, 2008; Gärtner, 2011; Parr & Cartwright-Hatton, 2009) were found. However, the majority of these studies chose study designs which might make it challenging to draw causal inferences on the effectiveness of the intervention (see Friedman, Furberg, & DeMets, 2010). For instance, the studies did not randomize students to treatment (intervention) and control (no intervention) groups (Fellenberg, 2008), did not have any control group at all (Brann-Barrett, 2005; Gärtner, 2011; Thielemann, 2014; Wilson, 1997), or used one source of assessing effectiveness measures only (e.g., only students' self-reports; Brann-Barrett, 2005), potentially resulting in less generalizable conclusions. Importantly, however, the studies give first evidence on the feasibility and acceptance of specific intervention concepts, i.e., the implemented methods and content, in K–12 communication education.

1.3. The present study

In order to address this gap in research, we developed a coherent speech training specifically targeting the abilities of elementary school children and the challenges they face, and evaluated its effectiveness. The content and didactics of the training were based on theory as well as effectiveness studies conducted in higher education (e.g., van Ginkel et al., 2015), covering the multiple dimensions of public speaking skills relevant for elementary school children (Backlund, 1985; Common Core State Standards Initiative, 2010; EDK, 2010; Günther, 2012; Morreale et al., 2000). Specifically, the aim of the training was to foster children's informative public speaking skills. As speech anxiety influences the way in which speaking tasks are approached, its reduction was a further instructional goal. The training consisted of 12 course units covering the topics of speech anxiety, nonverbal behavior, and comprehensibility. To investigate the effectiveness of the intervention, we conducted a randomized controlled trial (RCT; Friedman et al., 2010) in a natural setting. The control group also received a training covering 12 course units but the topic was different. Public speaking skills were assessed on the micro-level by means of self-reports (extent of public speaking skills) and video ratings (extent and appropriate use of public speaking skills).[1] Speech anxiety was assessed via self-reports. Furthermore, a number of different control variables were measured. As the training was part of an enrichment program for talented children and children's intelligence formed one factor affecting nomination for the program (see Rothenbusch, Zettler, Voss, Lösch, & Trautwein, 2016), fluid and verbal intelligence were considered as control variables. Social and emotional skills were included due to the fact that speaking in front of others is a special form of social interaction, especially for young children. Additionally, the reason for enrollment in the course-tandem was assessed to know which children had only wanted to attend the course in the course-tandem they had not been randomly assigned to. To assess differential effects for children with low versus high pretest scores on the dependent variable, the interaction term of course and pretest score was included as additional predictor variable.

We first expected that children participating in the speech training would improve their public speaking skills, reflected by a higher extent of self-perceived public speaking skills (Hypothesis 1). Second, we assumed that children participating in the speech training would show more appropriate behavior in their public speech (Hypothesis 2). Third, we expected a reduction of speech anxiety in the training group (Hypothesis 3).

2. Method

2.1. Participants and procedure

The speech training was conducted as a course offering within the Hector Children's Academy Program (HCAP) in the first term of the school year 2013/2014. The HCAP is an extracurricular enrichment program for talented elementary school children in the German state of Baden-Württemberg. Teachers nominate children for the HCAP, but not for specific courses. After acceptance by one of the approximately 60 local sites of the HCAP, children may attend extracurricular courses (for more information about the HCAP, see Rothenbusch et al., 2016).

Before the study started, written parental consent was received. Participants were 65 third and fourth grade students (42% female) who were enrolled in one of four local sites of the HCAP which offered the training in public speaking. The children had a mean age of 8.74 years (SD = 0.58). The mean fluid intelligence (see Measures section) was 118.71 (SD = 15.10). Therewith, the sample forms a specific group of talented children having an IQ slightly more than one standard deviation above the average IQ in the norm population.

To evaluate the effectiveness of the speech training an RCT with repeated measures (pre-post measurement) was used (Friedman et al., 2010). The pretest measurements took place before the intervention started (embedded in the first two course units). The posttest measurements took place in the last two course units. At each local site, one control group participated in a different enrichment course, a science-focused STEM training. Both, the intervention and the control course took place at the same time. The speech training was offered by the first author, a psychologist and speech scientist who developed the training,

[1] Although public speaking competence is based on the effectiveness and appropriateness of a speaker, the present study solely focused on the latter. To examine whether a speaker effectively transmitted the information to the audience, the listeners' knowledge of the topic also needs to be assessed. However, this is seldom done in either research or practice (see Morreale & Backlund, 2007). Thus, by concentrating on the appropriateness of the speech, we followed a common procedure applied across fields.

to all groups. The science-focused STEM training was offered by three further researchers who developed the training. Two of them were psychologists, one was an educational researcher and content specialist in science education.

Randomization was enabled by offering the two courses as a course-tandem titled *Talking about Science–With and to Others*. Enrollment was possible for the course-tandem only. After enrollment, blocked randomization based on computer generated random numbers was conducted. The blocked randomization resulted in equal numbers of students in control and experimental groups at each participating academy. In total, 33 children were randomly assigned to the speech training (age: $M = 8.75$, $SD = 0.58$; 45.5% female) and 32 to the alternative course (age: $M = 8.74$, $SD = 0.58$; 37.5% female). Subsequent to the study, all children were invited to participate in the respective other course.

2.2. Intervention: the speech training

The speech training aimed at fostering third and fourth graders to give informative public speeches competently. It was conceptualized for groups of six to ten children and consisted of 12 course units of 90 min each. The content and teaching methods of the training were derived from research, theoretical and practical work in communication education, psychology, rhetoric, and speech science (e.g., EDK, 2010; Morreale et al., 2000; Pabst-Weinschenk, 2005; van Ginkel et al., 2015).

The basis for the specification of the content was the conceptual framework of public speaking competence (Fig. 1). To foster children's informative public speaking skills, all four dimensions were considered: nonverbal–visual, nonverbal–auditory, language usage, and organization. Motivational aspects influence the way in which public speaking tasks are approached, and thus the degree of active participation in the training. Consequently, the reduction of stage fright was addressed prior to all other topics in the training program and remained relevant across all units. More precisely, the content covered within the training program, listed in the order of their inclusion, were stage fright, nonverbal communication visual and auditory, and comprehensibility, which comprises language usage and organization of speech. In addition, the training program was framed by introductory and conclusive elements within the first and last course units. Throughout the training, each child worked on an individual speech on an individually chosen scientific topic. The training ended with the presentation of this speech (for a detailed overview of the content allocated to each main topic, see Fig. 2).

As competent public speaking performance depends on the combination of knowledge, motivation, and skills, all three were addressed within the training. In addition to considering stage fright as one variable which influences a speaker's motivation, students' knowledge (i.e., deepening of knowledge and reflection on effective and appropriate behavior) and skills level (i.e., increasing the extension of the micro-level skills repertoire) were targeted within every unit and with regard to each topic.

The training's teaching methods were selected in a two-step approach. First, as studies on the effectiveness of different design principles on elementary-school-aged children's public speaking skills are scarce, principles that had shown positive effects on public speaking skills in higher education were considered (van Ginkel et al., 2015). Thus, behavior modelling, opportunity to practice, different types of feedback, and transfer were applied. The application of these principles was based on research on motivation (e.g., Niemiec & Ryan, 2009) and Bandura's (1977) social learning theory, among other concepts. In applying these concepts, behavior modelling formed one key instructional strategy, as it has been used in many studies on the promotion of public speaking skills (see De Grez & Valcke, 2010). Second, the explicit teaching activities were derived from these principles. In order to address the target group of elementary-school-aged children, methods evaluated in higher education were adapted to be more in line with practically tested, but unevaluated teaching materials for elementary school children (e.g., Pabst-Weinschenk, 2005). Final teaching activities included the following: First, demonstrations by the instructor and video and audio examples were used (referring to behavior modelling). Second, practical exercises which increased in complexity over the course of the training were applied (referring to opportunity to practice). The training started with short impromptu and associated speeches while sitting in the group, and finished with the students' task to prepare and deliver an informative public speech in front of the group. Third, different types of feedback were included in the training (peer and trainer feedback), ending with individual video feedback. Finally, transfer of newly learned knowledge and skills was facilitated within every unit by giving the children the opportunity to work on their own speech topic. Overall, these teaching methods are parallel to those applied in courses conducted within higher education. However, the explicit activities were adapted to serve the needs of the target group. Thus, age-appropriate instructions, images, work sheets, and time frames were used.

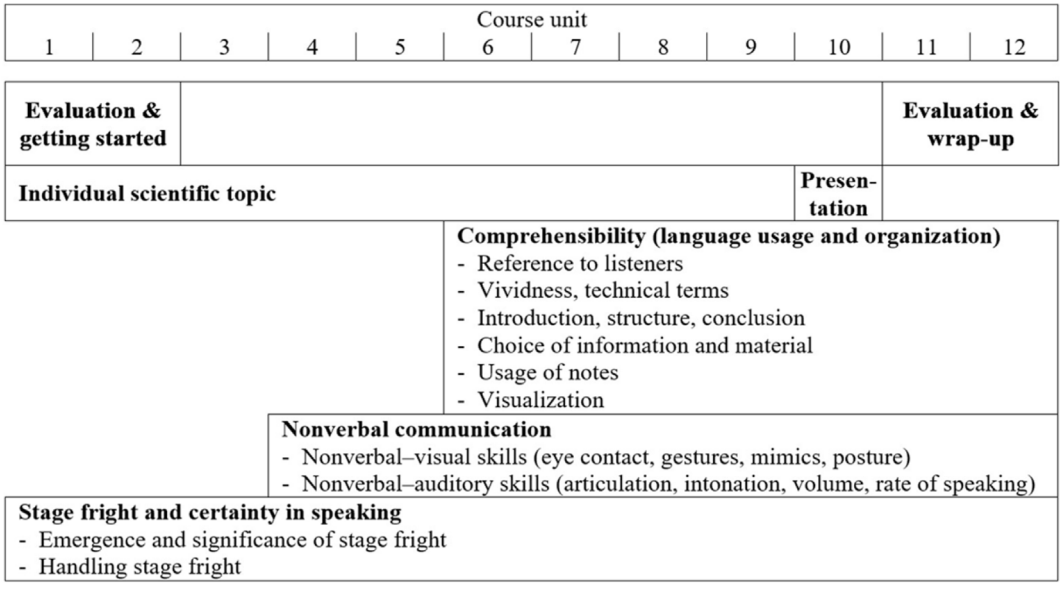

Fig. 2. Content of the speech training.

Overall, when designing the learning environment, decisions concerning instructional strategies were made for each course unit. In order to find an appropriate balance between high and low scaffolding, i.e., supplantive and generative instructional strategies, we considered students' characteristics (e.g., motivation, speech anxiety, prior knowledge), the specific learning tasks of each course unit, and the context of the training (see Smith & Ragan, 2005). As students' prior knowledge, public speaking skills, and confidence in speaking were expected to increase as the training went on and the learning tasks became more complex, the instructional strategies shifted from more supplantive instructional strategies at the beginning of the course to more generative strategies at the end. To illustrate how the training program was conceptualized, the structure of Course Unit 3 is depicted in Appendix A. It includes the specific goals of the unit, the main principles considered, teaching activities, and content.

2.3. Control condition: a science-focused STEM training

The training for the control condition focused on fostering children's understanding of science, more specifically inquiry-based methodological competencies and epistemic beliefs (Elder, 2002; Lederman, 2007; Osborne, 2013). It was conceptualized for groups of six to ten children and consisted, like the speech training, of 12 course units of 90 min each. The content and teaching methods of the STEM training were derived from research, education plans, and theoretical and practical work in science education (e.g., Akerson & Hanuscin, 2007; European Commission, 2007; Wendt et al., 2016). The content of the training covered topics such as experiments about the human senses, experiments in a student neuroscience lab, examination of an unknown object—a so-called "black box" (Frank, 2005)—, and simple physical experiments, e.g., about swimming and sinking.

The training was conducted using an inquiry-based learning approach (Blanchard et al., 2010; Colburn, 2000). Thereby, the intervention was based on the principle of a step-by-step unfolding of the inquiry process (Colburn, 2000). Scientific work according to what is known as the *scientific inquiry cycle* (SIC; see Kuhn, 2002) was another basic design principle of the training. The SIC was implemented in a step-by-step fashion and applied in all of the course sessions of the intervention. The sessions were arranged in such a way that the children experienced and applied the cumulative and cyclical process of scientific research within their research topics. Based on theories and derived hypotheses about their fields of research, the children conducted experiments, analyzed data, evaluated evidence, presented their results, and drew inferences with the goal of generating or revising the theories (Kuhn, 2002). Finally, the transition from hands-on activities to reflection and thinking about relevant aspects of the epistemology of science by means of discussions, science communication, and critical scrutiny was the third basic design principle of the STEM training (Akerson & Hanuscin, 2007). With regard to instructional strategies, the STEM training also gradually reduced the scaffolding provided by the instructor as time went on. Thus, more supplantive instructional strategies were used at the beginning of the course, while more generative strategies were applied at the end (see Smith & Ragan, 2005). Overall, the STEM training (control condition) was comparable to the treatment condition in terms of length, group size, and target group.

2.4. Treatment fidelity

To enable valid inferences on the effectiveness of the training, treatment fidelity needs to be ensured over all groups (Graham, Harris, & Zito, 2005). Hence, a course manual was prepared and pretested by offering the training to two groups of eight and nine children before the actual study was conducted. The final manual was scripted, including the introduction, explanation, conclusion, and transfer of each exercise. For an excerpt from the course manual see Appendix B; the complete set of instructions (in German) is available from the first author (Herbein, 2017). Within each treatment group in this study, all teaching materials were equal and the course content was presented in the same order. Furthermore, time frames were given for all exercises within one unit. The pretest of the training as well as the trainings at all four academies within the presented study were implemented by the first author. Doing this ensured that the training was conducted as intended (Graham & Harris, 2014).

2.5. Measures

All dependent variables were measured twice, at the beginning and at the end of the course. The variables can be summarized in three categories: 1) extent of public speaking skills, 2) appropriateness of public speaking skills, and 3) speech anxiety. The extent of public speaking skills was measured by one variable (self-reported). The appropriateness of public speaking skills was measured by 13 variables—including one global assessment—based on video ratings. Individual speech anxiety was measured as a single dependent variable via self-reports. In sum, 15 variables were assessed each at the pre- and posttest.

2.5.1. The public speaking task

To measure the extent and appropriateness of public speaking skills, the children gave short presentations on a scientific topic that were videotaped. To establish a standardized public speaking situation with equal preparation conditions in both groups, children of both groups performed the same task at the pre- and posttest: They were asked to draw a scientist (Chambers, 1983), to consider what the scientist does the whole day long, and afterwards to inform listeners about their picture and ideas via a speech. After a preparation time of 20 min in a group setting, children were randomly assigned to different test leaders who were blinded to the study aims. Each test leader took one child with her/him to a separate room in which the child presented her/his scientist to the test leader. This situation—presenting separately in front of one test leader instead of the class—was chosen in order to rule out imitation and learning effects by observation.[2] The order in which the children presented was randomly assigned. Based on the presentation, both the extent of students' skills (self-assessment) and the appropriateness of the shown behavior (external assessment) were assessed.

2.5.2. Public speaking skills—extent

The self-perceived extent of public speaking skills was assessed via the German version of the Performance Questionnaire (child version; PQ-C; Cartwright-Hatton, Tschernitz, & Gomersall, 2005). The PQ-C was translated from English to German using established scientific standards of translation-retranslation (e.g., Brislin, 1980). One item "How loud and clear was your voice?" was split to assess volume and clarity separately. Directly after presenting their drawings the children filled out the 11 items of the questionnaire. Responses were given on a 4-point Likert-type scale ranging from 1 = *not very much* to 4 = *very much*. Internal consistency as measured by Cronbach's α was .64 at pretest and .70 at posttest.

2.5.3. Public speaking skills—appropriateness

The assessment of the appropriateness of public speaking skills was based on the videotaped presentations, using a newly developed observation sheet. The observation sheet was developed on the basis of the conceptual framework of public speaking competence (Fig. 1). Consequently, micro-level public speaking skills from all four

[2] In the pilot phase of the training, these speeches took place in front of the class. However, we then observed that many children borrowed ideas of the previous speakers. Therefore, we thoroughly discussed different presentation options among the research group and with other speech scientists, resulting in the described approach.

dimensions, i.e., nonverbal–visual, nonverbal–auditory, language usage, and organizational skills, were included. In order to assess both the basic and advanced levels of public speaking competence (Rubin & Morreale, 1996; Staton & Tomlinson, 2001), public speaking skills were rated on two levels: First, the extent of skills was rated, allowing for the examination of the speaker's skills repertoire. Within this, the intensity of the public speaking skills shown, e.g., of eye contact, was rated. Subsequently, the appropriateness of the skills was assessed, e.g., whether the exhibited extent of eye contact was suitable for the context. This made up the advanced level of public speaking competence. Herein, however, we focus on the appropriateness of skills, because ratings of the extent and appropriateness of skills were highly correlated with each other in our study (88% of the item pairs had a correlation of $r \geq .50$). Comparable results for the effectiveness of the training were found for both types of skills assessment, and we deemed the appropriateness of skills as more important for the quality of an informative public speech because it reflects a higher level of public speaking competence (Rubin & Morreale, 1996).[3]

We developed 34 items capturing the appropriateness of public speaking skills (see Appendix C). Three external raters assessed the videotaped presentations on a 4-point Likert-type scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*; a sample items is "Eye contact is situationally appropriate". Prior to the assessment of the presentations of our study, the raters were trained using videos from different children performing the same task (i.e., talking about their drawn scientist). All raters received a rater manual and participated in three repeated training sessions with feedback from a speech and rating expert. After training, all three raters rated all 110 videos (61 pretest and 49 posttest; mean video duration: 1.08 min, $SD = 0.49$). The order of video presentation was randomized and the raters were blinded to group membership. Based on this design, interrater reliability was examined by using two-way, absolute, average-measure ICCs. To achieve a satisfactory precision of the ICC estimates (Bonett, 2002) interrater reliability was assessed for each item across all 110 videos. For our study, we only included items with good (between .60 and .74) or excellent interrater reliabilities (≥ .75; Cicchetti, 1994), resulting in 13 items for the analyses (see Appendix D).

Of these 13 items, twelve can be assigned to three of the four dimensions of public speaking skills. Nonverbal behavior–visual: *posture*, *gestures*, *mimics*, *eye contact*, and *proxemics;* nonverbal behavior–-auditory: *speech respiration*, *breaks*, *variation of breaks*, and *speech fluency;* organization of speech: *length of speech*, *length of conclusion*, and *reference to listener*. Note that no item of language usage had an ICC ≥ .60. The thirteenth item represents an overall rating of the appropriateness of the given speech.

2.5.4. Speech anxiety

Speech anxiety was measured using the German version of the Performance Anxiety Questionnaire (Cox & Kenardy, 1993; German: Fehm & Hille, 2005), assessing cognitive and bodily symptoms. This questionnaire includes 20 items, of which three items were slightly reworded to make them more appropriate for the age of the target group (e.g., original item: "I feel that I lack confidence."; reworded: "I feel uncertain."). For the analyses on the effectiveness of the training, one item was excluded because of a poor corrected item-total correlation ($r_{it} = .02$ for the first and $r_{it} = -.05$ for the second measurement occasion). Each item was answered on a 5-point Likert-type scale ranging from 1 = *never* to 5 = *always*. Cronbach's α was = .91 for the pretest and .94 for the posttest. Prior to the completion of the questionnaire, the children were instructed to think about the public speaking situation in general.

[3] Results regarding the treatment effects on the extent of public speaking skills can be obtained by contacting the first author.

2.5.5. Control variables

To increase the precision of the regression coefficients and to eliminate any bias that may be caused by differences between the two groups at the beginning of the study (Cohen, Cohen, West, & Aiken, 2003), control variables were included. General fluid intelligence was measured via the German version of the Culture Fair Intelligence Test (CFT 20-R; Weiß, 2006). The test consists of four subtests, namely series, classifications, matrices, and typologies. Cronbach's α was = .77. Verbal intelligence was assessed using two subtests, vocabulary and word classification, of the German Munich High Ability Test Battery for primary school level (Heller & Perleth, 2007). Cronbach's α for grade level 3 was .84 and .80 for grade level 4.

Social and emotional skills were measured using the parental version of the Devereux Student Strengths Assessment (DESSA; LeBuffe, Shapiro, & Naglieri, 2009). Specifically, we administered the subscales self-awareness (Cronbach's α = .69), social-awareness (α = .79), self-management (α = .80), relationship skills (α = .86), and personal responsibility (α = .80). The parents assessed their child on a 5-point Likert-type scale ranging from 1 = *never* to 5 = *very often*. Again, established scientific standards of translation-retranslation were used to translate the DESSA from English to German.

Parents assessed the reason for enrollment by means of four items ("My child was enrolled to the course because of his/her interest in mathematics and natural sciences/technical topics/verbal topics/creative topics."). The extent of agreement was indicated on a 4-point Likert-type scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*.

2.6. Analyses

Although participants were randomly assigned to the training conditions, baseline equivalence on key characteristics was examined to ensure the groups were similar enough at the beginning of the study. Therefore, two-tailed t-tests were conducted for all dependent and control variables (pretest measures). To estimate the effectiveness of the intervention multiple linear regressions were used. Analyses were conducted via Mplus Version 7 (Muthén & Muthén, 1998–2012) using maximum likelihood robust estimation (MLR). Over both groups the amount of missing data ranged between 2% and 25% (see Appendix D). The higher missing rate resulted from absence of students at posttest. However, there was no differential drop-out between treatment and control group ($\chi^2(1,65) = .26$; $p = .614$). Furthermore, when comparing the means of the dependent variables at pretest for the children missing at posttest with children not missing at posttest, no significant differences were found (all p-values > .05). This is compatible with the assumption that the missing data were missing at random (Enders, 2010), and thus the full-information maximum likelihood (FIML) estimator was used. FIML analysis represents a direct estimation approach to handling missing data that uses all available information in the data to calculate parameter estimators and standard errors (Buhi, Goodson, & Neilands, 2008). One-tailed tests of significance were used with significance level (α) set at .05, because directed hypotheses were formulated for the effects of the training.

The dependent variables were the posttest measurements for each of the 15 variables (two self-reports, 13 observer ratings). For the video ratings, the mean of all three raters was used. In order to assess the effects of pretest differences and differential effects for children with low versus high pretest scores on the dependent variable, the pretest score and interaction term of course and pretest score were included as additional predictor variables. In case of a significant interaction term, the effect of course participation differed for children depending on their initial score on the dependent variable (Cohen et al., 2003). All continuous variables were standardized prior to the analyses. Each course was binary coded, with speech training = 1 and control group = 0. The size of the course or treatment effect is indicated by the standardized mean differences between the two groups (Hedges, 2007).

Due to standardization, the regression coefficient of the course variable is the same as the effect size of the course effect. Given that there are no similar studies, we applied the commonly used classification of effect sizes as a reference standard: small: $d = 0.20$, medium: $d = 0.50$, large: $d = 0.80$ (Cohen, 1992). As treatment effects were tested on 15 dependent variables, we corrected for multiple testing using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) and adjusted *p*-values are reported.

3. Results

The descriptive statistics for both groups and measurement points are reported in Appendix D. With regard to baseline equivalence, the t-tests revealed that the intervention and control groups did not significantly differ on the dependent and control variables at pretest, with the exception of two variables: Significant differences between the groups were found for social-awareness ($t(62) = 2.22$, $p = .030$) and personal responsibility ($t(62) = 2.12$, $p = .033$). Thus, both variables (but not the other DESSA subscales) were further included as control variables. However, correcting for multiple testing (Benjamini & Hochberg, 1995), no differences between the groups were found for any variable.[4] The correlations among the dependent variables at pre- and posttest are shown in Appendix E. The correlations among the dependent variables at the posttest and the control variables are displayed in Appendix F. The training effects are shown in Tables 1–3.

First, we hypothesized that children participating in the speech training would improve the extent of their self-perceived public speaking skills. In line with this hypothesis, the training had a positive effect on the development of self-perceived public speaking skills ($B = 0.70$, $p = .008$; Table 1): Children who participated in the speech training reported significantly higher scores in self-perceived public speaking skills compared to children in the control group, controlling for the pretest, the interaction term between course and pretest score, fluid and verbal intelligence, course registration verbal interest, social-awareness, and personal responsibility.

Second, we hypothesized that children participating in the speech training were better able to adapt and use their skills in light of a public speaking context. Based on the external ratings of the students' presentations, positive training effects were found for four out of five nonverbal–visual behavior items: *posture* ($B = 0.69$, $p = .009$), *gestures* ($B = 0.77$, $p = .008$), *eye contact* ($B = 0.45$, $p = .029$), and *proxemics* ($B = 0.71$, $p = .008$); no significant effect was found for appropriate usage of *mimics* ($B = 0.03$, $p = .448$; Table 1). For three out of four nonverbal–auditory behavior items, positive effects were found: *breaks* ($B = 0.64$, $p = .013$), *variation of breaks* ($B = 0.76$, $p = .009$), and *speech fluency* ($B = 0.54$, $p = .016$); by contrast, children of the training group showed worse *speech respiration* after the training than the children of the control group ($B = -0.70$, $p = .004$; Table 2). Concerning organizational skills, positive effects of the training were found for all three items (*length of speech*: $B = 0.54$, $p = .030$; *length of conclusion*: $B = 0.59$, $p = .039$; *reference to listener*: $B = 0.54$, $p = .008$; Table 3). Overall, children participating in the intervention used *eye contact*, *posture*, *gestures*, *proxemics* (nonverbal–visual), *breaks*, *variation of breaks*, *speech fluency* (nonverbal–auditory), *length of speech*, *length of conclusion*, and *reference to listeners* (organization) more appropriately, and *speech respiration* less appropriately as compared to children of the control condition. Finally, the course had a positive effect on the appropriateness of *global performance* ($B = 0.62$, $p = .005$; Table 3): Children of the intervention group showed a significantly more appropriate *global performance* when talking than did children of the control group.

With regard to differential effects depending on children's pretest scores, a significant interaction term of pretest score and course was only found for the variable *speech respiration* ($B = 0.70$, $p = .004$). Accordingly, children with higher appropriateness of speech respiration at the pretest benefitted more from the course.

Finally, although we expected the training to produce a reduction in speech anxiety, no such effect was found ($p = .242$; Table 3). Children attending the course reported no different intensity in speech anxiety symptoms (i.e., perceived cognitive and bodily symptoms) than did the control group. The results of all analyses remained stable when we excluded fluid and verbal intelligence from the analyses.

4. Discussion

Even young school children need to show public speaking competence. However, corresponding training offers are rare and have hardly been evaluated. In order to address this imbalance, we developed and evaluated a training which aimed at fostering third and fourth graders' public speaking skills. We applied a robust research design, including a treated control group, randomization, and repeated measures, to test the effectiveness of the training.

Generally, the findings revealed positive effects of medium size for both the extent of self-perceived public speaking skills and the appropriateness of public speaking skills assessed by external raters. In their presentations after the course, children participating in the training showed more appropriate public speaking skills for the dimensions nonverbal–visual, nonverbal–auditory, and organization, as well as a better global performance than children of the control group. However, the training did not affect *mimics* (nonverbal–visual) and speech anxiety of the children attending the course, who even showed worse *speech respiration* (nonverbal–auditory) after the course as compared to their counterparts.

4.1. Classification and implication of the training effects

The results of our study indicate that the training—which adapted instructional designs from secondary and higher education to the context of elementary school—was successful in promoting public speaking competence of elementary school children. Including teaching methods such as the use of role models, the opportunity to practice, and different feedback methods, the training showed medium effects on a broad set of public speaking skills. Training effects were found on all considered dimensions of public speaking skills and were not isolated to single dimensions. Furthermore, children participating in the training not only showed a higher extent of public speaking skills (improvement at the basic level of public speaking competence), but they were also better

[4] To check the robustness of the effects, we also ran five additional models which differed in (a) how we adjusted for multiple testing when predicting the outcomes and (b) whether or not we controlled for pre-treatment differences. In Model 2 and 3, we used the same predictor variables as in Model 1 but applied two further approaches to adjust for multiple testing when predicting the outcomes, using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). In comparison to Model 1 in which we adjusted for multiple testing across all 15 outcome variables (most conservative approach), we adjusted alpha within each set of variables belonging to the same dimension of public speaking skills (i.e., nonverbal–visual, nonverbal–auditory, and organizational public speaking skills) in Model 2 (less conservative). In Model 3, we did not adjust alpha for multiple testing at all. In Models 4, 5, and 6, we excluded the DESSA variables *social-awareness* and *personal responsibility* from the analyses because no significant baseline differences were found after adjusting alpha for multiple testing at pretest. Using the same predictor variables across the Models 4, 5, and 6, we again adjusted alpha for multiple testing in three different ways. In Model 4, we adjusted alpha across all 15 outcome variables, in Model 5, we adjusted alpha within each set of variables belonging to the same dimension of public speaking skills, and in Model 6, we did not adjust alpha. In Models 1 to 3, the regression coefficients (i.e., treatment effects) are identical, only the alpha level varies between these models (the same applies to Models 4 to 6). A comparison of the effect sizes of Models 1 to 3 and 4 to 6 revealed that the treatment effects were quite similar. More specifically, the differences in effect sizes had a median value of $d = 0.12$ and a range between 0.00 and 0.36. Controlling for pre-treatment differences (in Model 1, 2, and 3) was associated with slightly higher effect sizes. However, as expected, the level of significance differed across the models, with the largest number of significant effects found in Models 1 (presented in this paper), 2, and 3 (13 significant treatment effects), 12 significant effects in Model 6, 11 significant effects in Model 5, and the lowest number in Model 4 (eight significant treatment effects). These additional tables can be obtained from the first author.

Table 1
Training effects on self-perceived public speaking skills and the appropriateness of nonverbal–visual public speaking skills (posttest).

	Speech Performance			Nonverbal–Visual														
				Posture			Gestures			Mimics			Eye Contact			Proxemics		
	B	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*
Course	0.70	0.27	.008	0.69	0.26	.009	0.77	0.24	.008	0.03	0.25	.448	0.45	0.22	.029	0.71	0.25	.008
Pretest Score	0.54	0.19	.009	0.41	0.21	.114	0.51	0.16	.015	0.07	0.20	.734	0.69	0.20	.001	0.32	0.16	.141
Course × Pretest Score	−0.32	0.27	.428	−0.21	0.23	.806	−0.14	0.24	1	0.21	0.30	.483	−0.44	0.24	.086	0.07	0.23	1
Fluid Intelligence	0.17	0.10	.165	0.14	0.11	.399	0.30	0.11	.090	−0.22	0.21	.296	0.14	0.11	.290	−0.14	0.13	.939
Verbal Intelligence	−0.12	0.15	.797	0.07	0.16	1	0.07	0.13	1	0.32	0.16	.051	0.15	0.15	.447	0.19	0.12	.366
Social-Awareness	0.14	0.13	.559	0.14	0.15	.701	−0.17	0.12	1	−0.03	0.19	.892	0.09	0.14	.665	0.40	0.11	< .001
Personal Responsibility	0.05	0.18	1	0.14	0.16	.840	0.31	0.11	.105	0.25	0.17	.142	−0.13	0.16	.588	−0.14	0.12	.663
Course Registration Verbal Interest	−0.25	0.13	.094	−0.12	0.13	.767	−0.24	0.12	.675	−0.22	0.14	.125	−0.14	0.14	.421	−0.16	0.12	.480
Explained variance (R^2)	.25			.25			.49			.19			.28			.41		

Note. Speech Performance = Self-perceived extent of public speaking skills. All continuous variables were standardized prior to the analysis. Course was binary-coded (1 = intervention, 0 = control group). One-tailed significance levels are reported for Course, set to an overall level of α = .05 and adjusted with the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). *p* = the adjusted *p*-value.

Table 2
Training effects on the appropriateness of nonverbal–auditory public speaking skills (posttest).

	Nonverbal–Auditory											
	Speech Respiration			Breaks			Variation of Breaks			Speech Fluency		
	B	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*
Course	−0.70	0.21	.004	0.64	0.27	.013	0.76	0.28	.009	0.54	0.23	.016
Pretest Score	0.08	0.15	1	0.33	0.18	.102	0.07	0.20	1	0.59	0.16	.001
Course × Pretest Score	0.70	0.17	.004	−0.14	0.22	.882	0.26	0.24	.693	−0.10	0.20	.938
Fluid Intelligence	−0.05	0.10	1	−0.01	0.11	1	0.08	0.14	1	0.05	0.11	.947
Verbal Intelligence	−0.08	0.10	1	0.49	0.14	.002	0.42	0.14	.008	0.30	0.13	.030
Social-Awareness	0.10	0.12	1	0.03	0.15	1	0.06	0.16	1	−0.13	0.16	.612
Personal Responsibility	−0.06	0.09	1	0.29	0.17	.135	0.26	0.19	.420	0.35	0.16	.045
Course Registration Verbal Interest	0.05	0.09	1	−0.13	0.12	.522	−0.16	0.13	.603	−0.20	0.10	.069
Explained variance (R^2)	.57			.30			.29			.39		

Note. All continuous variables were standardized prior to the analysis. Course was binary-coded (1 = intervention, 0 = control group). One-tailed significance levels are reported for Course, set to an overall level of α = .05 and adjusted with the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). *p* = the adjusted *p*-value.

Table 3
Training effects on the appropriateness of organizational skills, on global performance, as well as on the reduction of speech anxiety (posttest).

	Organization									Global Performance			Speech Anxiety		
	Length of Speech			Length of Conclusion			Reference to Listener								
	B	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*
Course	0.54	0.28	.030	0.59	0.32	.039	0.54	0.19	.008	0.62	0.20	.005	−0.15	0.20	.242
Pretest Score	0.58	0.20	.004	0.17	0.25	.577	0.41	0.16	.038	0.79	0.19	.003	0.77	0.12	< .001
Course × Pretest Score	−0.28	0.24	.305	0.53	0.27	.054	0.07	0.29	1	−0.17	0.22	1	−0.06	0.20	.837
Fluid Intelligence	−0.06	0.08	.565	0.13	0.13	.363	−0.10	0.13	1	0.12	0.13	1	−0.02	0.12	.918
Verbal Intelligence	0.38	0.12	.003	0.05	0.12	.808	0.31	0.16	.229	0.32	0.13	.060	−0.01	0.09	.950
Social-Awareness	0.15	0.14	.393	0.00	0.14	1	0.02	0.18	1	−0.06	0.12	1	0.05	0.10	.638
Personal Responsibility	0.45	0.20	.025	0.07	0.14	.707	0.16	0.22	1	0.35	0.13	.030	−0.06	0.08	.474
Course Registration Verbal Interest	−0.27	0.12	.025	−0.14	0.11	.207	−0.21	0.13	.423	−0.21	0.11	.255	0.13	0.10	.179
Explained variance (R^2)	.38			.38			.36			.62			.58		

Note. Speech Anxiety = Self-reported frequency of speech anxiety symptoms. All continuous variables were standardized prior to the analysis. Course was binary-coded (1 = intervention, 0 = control group). One-tailed significance levels are reported for Course, set to an overall level of α = .05 and adjusted with the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995). *p* = the adjusted *p*-value.

able to adapt their skills in order to behave appropriately (advanced level of public speaking competence; Rubin & Morreale, 1996).

Due to the lack of similar studies assessing the effects of interventions on public speaking skills for elementary school children, a rather rough interpretation of our results can be done by pointing at the results of studies conducted in higher education. For instance, De Grez et al. (2009a, 2009b) assessed the effectiveness of two interventions on public speaking skills, combining several design principles. They found small effects on body language (nonverbal–visual) and medium to strong effects on length of conclusion, reference to listener, and length of introduction (organization). They did not observe any effects for eye contact (nonverbal–visual) and vocal delivery (nonverbal–auditory). Therefore, in general, the findings of these studies conducted in higher education are quite similar to our results.

Unexpectedly, worse *speech respiration* (nonverbal–auditory) was found for children participating in the speech training compared to children in the control group. A possible explanation might be that within the course unit on speech anxiety, children thought about their

personal tip to reduce nervousness. One tip, derived from speech science, is to consciously exhale deeply before speaking. This allows a change from thoracic to abdominal breathing. Consequently, children in the intervention group might have used this tip in an intensive way, leading to a perceived inappropriateness of respiration. Future implementations of the speech training may additionally provide training on how to use this tip in a less obvious way.

Concerning speech anxiety, only small, non-significant effects were found. One reason for this result could be that the children participating in the course-tandem already showed low manifestations of speech anxiety before the training started ($M = 2.17$, $SD = 0.76$, on a 5-point Likert-type scale). This could have made it difficult to detect changes on speech anxiety based on the training. Another explanation could be that although the training enhanced certainty and self-perceived public speaking competence, it might also have heightened self-awareness concerning public speaking behavior. Knowing what is expected of a competent speaker might have increased students' expectations with regard to their own performances. This could have resulted in persistence in the speech anxiety level as students tend to view their own behavior most critically. Further studies are needed to test these or other explanations.

4.2. Putting public speaking competence on the agenda of research on learning and instruction

A main strength of the present study is the development of a training in informative public speaking competence specifically for the target group of elementary school children as well as the evaluation of this training using a robust study design. In elementary school, the ability to give informative public speeches competently is required across the entire curriculum, and the student's performance influences the final grade in the subject in which the speech is delivered (Rubin & Morreale, 1996). Further, the importance of public speaking competence increases continuously from elementary to secondary up to higher education. In secondary school, the frequency and complexity of speeches across the entire curriculum increase steadily; in higher education, giving an informative public speech is often a necessary requirement for participating in a course; in both secondary and higher education, oral exams are part of many final exams; finally, public speaking competence is required in many professions (see Common Core State Standards Initiative, 2010; Girard, Pinar, & Trapp, 2011; Hristova, 2014).

It is often erroneously assumed that after learning to talk, children learn to communicate competently without further instruction or specific educational support (Morreale et al., 2000). However, elementary school children vary in their public speaking abilities from the outset, and many students ultimately start their professional education with deficits in this regard (Hunt et al., 2014). Consequently, there have been calls for the early promotion of public speaking skills beginning at the elementary school level (Morreale et al., 2000). The successful development of such skills among third and fourth graders may predict future public speaking competence. We thus hope that our study can serve as a starting point for more extensive research on public speaking skills among younger students.

4.3. Limitations and strengths

The promotion and assessment of public speaking competence entail various challenges. Similarly to previous research on this topic it is difficult to generalize our findings to other public speaking situations. The appraisal as to whether a speaker is competent depends on the specific situation, the target audience, the communicational intention, the topic, and whether certain norms and standards are considered. Within this study, the children solved one specific public speaking task on a predetermined topic within one defined situation. However, the used situation is one of the most popular situations in elementary school. Children had to inform an audience about a topic that they prepared by themselves.

In addition, the public speaking task was the same at the pre- and posttest. If students perform the same test twice, at the pre- and the posttest, unintended practice effects may occur which threaten internal validity. In the present study, this may have led to an improvement in skills required for this specific task. In addition, children might have gotten used to the task, influencing speech anxiety. Consequently, the treatment effects found in this study might be rather conservative. However, we tried to keep the negative effect of testing on internal validity to a minimum, as the intervention and the control groups did not differ in the inclusion of the task in the subsequent training (see D'Agostino, 2005).

To assess public speaking competence, we used self-reports and external ratings to examine the extent and the appropriateness of the exhibited behavior. The advantage of this procedure is to get a more comprehensive view of actual public speaking competence by considering both the speaker's and audience's perspectives. However, the rater only assessed videotaped presentations and the person to whom the speech was actually directed did not rate the child's public speaking competence. It might have been interesting to include this in the study. By using external, trained, independent observers who did not know the children or the research question, we aimed at having an effective way to assess public speaking competence without strong biases. Furthermore, the children's knowledge of the topic they spoke about was not assessed. This was due to the fact that children presented their ideas on what a scientist is and does, which does not require specific knowledge. However, future studies may additionally consider knowledge of the topic of presentation.

Concerning the assessment of public speaking competence, future research is necessary. We developed an observation sheet addressing relevant public speaking skills to a highly differentiated degree on the basis of research from different areas. However, interrater reliabilities were not high enough to analyze all developed items. Twenty-one items had to be excluded because acceptable interrater reliability ($> .60$; Cicchetti, 1994) was not reached, even though the raters were trained and received a manual. One reason why rating the presentations might have been difficult might be the fact that the videotaped presentations were very short (mean duration 1.08, $SD = 0.49$ min), which might have impeded the analysis of organizational elements such as the amount of information and the communication intention. Furthermore, items related to nonverbal–auditory skills and language usage seemed to be particularly difficult to assess. Difficulties in rating nonverbal–auditory skills might have resulted from the raters' lack of experience. Although trained over the course of three training sessions, they were not speech science or communication professionals. Subtle acoustical nuances need to be perceived in order to rate the appropriateness of nonverbal–auditory skills, such as nuances of dynamics, hoarseness, or intonation. This might be more difficult than rating nuances of visual impressions. By comparison, 11 out of 15 items for nonverbal–auditory skills had to be excluded, but only one out of six for nonverbal–visual skills. Difficulties in rating language usage skills might have arisen from raters' difficulty distinguishing between high and low language proficiency (Cheng & Warren, 2005). In line with our findings, studies conducted in higher education contexts have reported comparably low interrater reliability on items related to language usage when assessing public speaking competence (e.g., Joe, Kitchen, Chen, & Feng, 2015). Thus, further research is needed to address these difficulties, including, for example, adjustments to the assessment tool, rater training, and manual (Cheng & Warren, 2005; Joe et al., 2015). Nevertheless, to the best of our knowledge this is the first study to apply a detailed observation sheet to assess the extent and appropriateness of elementary school children's public speaking skills using several external raters and taking interrater reliability into account.

Within the present study, the speech training was offered as an enrichment program targeting talented elementary-school-aged

children. To participate in the program, children were nominated by their teachers. Compared to their non-nominated peers, they have higher fluid and crystallized intelligence scores (see Rothenbusch et al., 2016). Thus, the restricted range needs to be considered in the present study. However, as nomination by teachers was used as a precondition for enrolment instead of test scores, the children still varied in their cognitive and verbal abilities. This could be confirmed in the present study (fluid intelligence $M = 118.71$, $SD = 15.10$; verbal intelligence $M = 25.75$, $SD = 7.09$).

Finally, the speech training was conducted by one trainer in all four academies. This approach was chosen to ensure high fidelity and training integrity as well as a high degree of internal validity (Graham et al., 2005). However, having only one trainer offer the intervention forms a confounding factor and needs to be considered as a possible reason for the differences in the outcomes (What Works Clearinghouse, 2014). Despite this limitation, having only one trainer offer the training forms a reasonable approach at this stage of intervention implementation and evaluation (Gottfredson et al., 2015). The present study represents an efficacy study, which aims to prove the success of the training program under rather controlled conditions. Thereby, the aim is to examine whether the training is at all efficacious for the intended target group. Therefore, the emphasis is set on internal rather than external validity. Accordingly, the framework conditions of the training implementation and the characteristics of the trainer are highly controlled. If positive treatment effects on the intended outcomes are found in an efficacy study, considerations can be made with regard to further implementing the training program (Gottfredson et al., 2015). This means that, based on the results of this study, a next step might be to evaluate the effectiveness of the training implemented by different trainers.

In addition to further studies on *how* the training program affects public speaking skills and speech anxiety when put into practice, research on *what* specifically about the training works needs to be conducted. The training program represented a multi-component intervention including design principles suggested by van Ginkel et al. (2015). However, future studies are necessary to disentangle the effectiveness of each individual design principle (behavior modelling, opportunity to practice, different types of feedback, and transfer). Identifying the active ingredients of the training program, i.e., the components responsible for the observed effects (Abry, Hulleman, & Rimm-Kaufman, 2015; Munter, Wilhelm, Cobb, & Cordray, 2014), will allow conclusions to be drawn on how to train elementary school-aged children's public speaking skills more effectively and efficiently (Abry et al., 2015). Overall, given the importance of the topic, further studies on the promotion of young students' public speaking skills are needed. As our investigation shows, it is possible to improve diverse public speaking skills among elementary school students. We thus hope that more effective learning interventions will be made available for this group of students.

Acknowledgements

Funding: This research project was funded by a grant of the Hector Foundation II to the fifth author. Evelin Herbein and Julia Schiefer were doctoral students at the LEAD Graduate School & Research Network [GSC 1028], funded by the Excellence Initiative of the German federal and state governments.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.learninstruc.2017.10.008.

References

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation, 36*(3), 320–338. https://doi.org/10.1177/1098214014557009.

Akerson, V. L., & Hanuscin, D. L. (2007). Teaching nature of science through inquiry: Results of a 3-year professional development program. *Journal of Research in Science Teaching, 44*(5), 653–680. http://dx.doi.org/10.1002/tea.20159.

Backlund, P. (1985). Essential speaking and listening skills for elementary school students. *Communication Education, 34*(3), 185–195. http://dx.doi.org/10.1080/03634528509378606.

Backlund, P. M., & Morreale, S. P. (2015). Communication competence: Historical synopsis, definitions, applications, and looking to the future. In A. F. Hannawa, & B. H. Spitzberg (Eds.). *Handbooks of communication Science: Vol. 22. Communication competence*(pp. 11–38). Berlin: De Gruyter.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289–300. Retrieved from http://www.jstor.org/stable/2346101.

Blanchard, M. R., Southerland, S. A., Osborne, J. W., Sampson, V. D., Annetta, L. A., & Granger, E. M. (2010). Is inquiry possible in light of accountability? A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education, 94*(4), 577–616. http://dx.doi.org/10.1002/sce.20390.

Bonett, D. G. (2002). Sample size requirements for estimating intraclass correlations with desired precision. *Statistics in Medicine, 21*(9), 1331–1335. http://dx.doi.org/10.1002/sim.1108.

Brann-Barrett, M. T. (2005). Youth speaks up: Perceived communication changes experienced by grade 6 participants in a personal development program. *Alberta Journal of Educational Research, 51*(2).

Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In H. C. Triandis, & J. W. Berry (Eds.). *Handbook of cross-cultural psychology. Methodology* (pp. 389–444). Boston, MA: Allyn and Bacon.

Buhi, E. R., Goodson, P., & Neilands, T. B. (2008). Out of sight, not out of mind: Strategies for handling missing data. *American Journal of Health Behavior, 32*(1), 83–92. https://doi.org/10.5993/AJHB.32.1.8.

Cartwright-Hatton, S., Tschernitz, N., & Gomersall, H. (2005). Social anxiety in children: Social skills deficit, or cognitive distortion? *Behaviour Research and Therapy, 43*(1), 131–141. http://dx.doi.org/10.1016/j.brat.2003.12.003.

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science Education, 67*(2), 255–265. http://dx.doi.org/10.1002/sce.3730670213.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing, 22*(1), 93–121. http://dx.doi.org/10.1191/0265532205lt298oa.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. http://dx.doi.org/10.1037/1040-3590.6.4.284.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: L. Erlbaum Associates.

Colburn, A. (2000). An inquiry primer. *Science Scope, 23*(6), 42–44.

Common Core State Standards Initiative (2010). *Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects.* Retrieved from http://www.corestandards.org/wp-content/uploads/ELA_Standards.pdf.

Cox, W. J., & Kenardy, J. (1993). Performance anxiety, social phobia, and setting effects in instrumental music students. *Journal of Anxiety Disorders, 7*(1), 49–60. http://dx.doi.org/10.1016/0887-6185(93)90020-L.

Croucher, S. M. (2013). Communication apprehension, self-perceived communication competence, and willingness to communicate: A French analysis. *Journal of International and Intercultural Communication, 6*(4), 298–316. http://dx.doi.org/10.1080/17513057.2013.769615.

De Grez, L., & Valcke, M. (2010). Learning and instruction of oral presentation skills. In L. E. Kattington (Ed.). *Education in a competitive and globalizing world. Handbook of curriculum development* (pp. 177–208). New York, NY: Nova Science.

De Grez, L., Valcke, M., & Roozen, I. (2009a). The impact of an innovative instructional intervention on the acquisition of oral presentation skills in higher education. *Computers & Education, 53*(1), 112–120. http://dx.doi.org/10.1016/j.compedu.2009.01.005.

De Grez, L., Valcke, M., & Roozen, I. (2009b). The impact of goal orientation, self-reflection and personal characteristics on the acquisition of oral presentation skills. *European Journal of Psychology of Education, 24*(3), 293–306. http://dx.doi.org/10.1007/BF03174762.

D'Agostino, J. (2005). Measuring learning outcomes: Reliability and validity issues. In G. D. Phye, D. H. Robinson, & J. R. Levin (Eds.). *Empirical methods for evaluating educational interventions* (pp. 113–146). San Diego: Elsevier; Elsevier Academic Press.

EDK (2010). *SCHULSPRACHE: Wissenschaftlicher Kurzbericht und Kompetenzmodell: Provisorische Fassung (vor Verabschiedung der Standards)*. [School language: Scientific short report and competence model]. Retrieved from http://www.edudoc.ch/static/web/arbeiten/harmos/L1_wissB_25_1_10_d.pdf.

Elder, A. (2002). Characterizing fifth grade students' epistemological beliefs in science. In B. K. Hofer, & P. R. Pintrich (Eds.). *The psychology of beliefs about knowledge and knowing* (pp. 347–363). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences.* New York, NY: Guilford Press.

European Commission (2007). *Science education now: A renewed pedagogy for the future of Europe. Gemeinschaftsforschung Expertengruppe, Vol 22845*. Luxembourg: Office for Official Publications of the European Communities.

Fehm, L., & Hille, C. (2005). Verhaltenstherapie und VerhaltensmedizinBühnenangst bei Musikstudierenden. [Stage fright amongst students of music], Vol 2, 199–212.
Fellenberg, M. (2008). [Practical rhetoric at school: An empirical study on different teaching methods for preparing student presentations]Praktische Rhetorik in der Schule: Eine empirische Untersuchung verschiedener Unterrichtsmethoden zur Vorbereitung von Schülervorträgen, Vol. 6 (Augsburg: Wißner).
Frank, A. (2005). Naturwissenschaftliches Arbeiten mit der Blackbox. [Scientific work with a blackbox]. Unterricht Biologie, 307/308, Beilage.
Friedman, L. M., Furberg, C., & DeMets, D. L. (2010). Fundamentals of clinical trials (4th ed.). New York, NY: Springer.
Gärtner, C. (2011). Die Effektivität eines Rhetoriktrainings: Eine Einzelfallstudie anhand von Schülergruppen[Effectiveness of a rhetorical training: A case study with groups of students]. (Master-Thesis). Regensburg: Universität Regensburg.
van Ginkel, S., Gulikers, J., Biemans, H., & Mulder, M. (2015). Towards a set of design principles for developing oral presentation competence: A synthesis of research in higher education. Educational Research Review, 14, 62–80. http://dx.doi.org/10.1016/j.edurev.2015.02.002.
Girard, T., Pinar, M., & Trapp, P. (2011). An exploratory study of class presentations and peer evaluations: Do students perceive the benefits? Academy of Educational Leadership Journal, 15(1), 77–94.
Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., et al. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. Prevention Science: The Official Journal of the Society for Prevention Research, 16(7), 893–926. http://dx.doi.org/10.1007/s11121-015-0555-x.
Graham, S., & Harris, K. H. (2014). Conducting high quality writing intervention research: Twelve recommendations. Journal of Writing Research, 6(2), 89–123. http://dx.doi.org/10.17239/jowr-2014.06.02.1.
Graham, S., Harris, K. H., & Zito, J. (2005). Promoting internal and external validity. In G. D. Phye, D. H. Robinson, & J. R. Levin (Eds.). Empirical methods for evaluating educational interventions (pp. 235–265). San Diego, CA: Elsevier Academic Press.
Günther, H. (2012). Sprechen und Zuhören: Wie Lehrerinnen und Lehrer Sprachunterricht ökonomisch und effektiv planen und durchführen. [Speaking and listening: How teachers plan and implement language instruction economically and effectively] Baltmannsweiler: Schneider Hohengehren.
Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. Journal of Educational and Behavioral Statistics, 32(4), 341–370. http://dx.doi.org/10.3102/1076998606298043.
Heller, K. A., & Perleth, C. (2007). MHBT-P. Münchner Hochbegabungstestbatterie für die Primarstufe. [The Munich High Ability Test Battery, Primary Level]Göttingen: Hogrefe.
Herbein, E. (2017). Public speaking training as an enrichment program for elementary school children: Conceptualization, evaluation, and implementation (Doctoral dissertation) Tübingen, Germany: Eberhard Karls Universität Tübingen; Universitätsbibliothek Tübingen. Retrieved from https://doi.org/10.15496/publikation-17711.
Hristova, Z. (2014). Using peers to assess oral presentations to foster learning. International Journal of Innovation in Science and Mathematics Education, 22(3), 74–80.
Hunt, S., Wright, A., & Simonds, C. (2014). Securing the future of communication education: Advancing an advocacy and research agenda for the 21st century. Communication Education, 63(4), 449–461. http://dx.doi.org/10.1080/03634523.2014.926016.
Joe, J., Kitchen, C., Chen, L., & Feng, G. (2015). A prototype public speaking skills assessment: An evaluation of human-scoring quality. ETS Research Report Series, (2), 1–21. http://dx.doi.org/10.1002/ets2.12083 2015.
Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.). Handbook of childhood cognitive development. Oxford: Blackwell.
[Decisions of the Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany. Scholastic standards for German for elementary schools (grade 4)]. MünchenIn Kultusministerkonferenz (Ed.). (2005). Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4).
LeBuffe, P., Shapiro, V., & Naglieri, J. (2009). The Devereux student strengths assessment (DESSA). Lewisville, NC: Kaplan Press.
Lederman, N. G. (2007). Nature of science: Past, present, and future. In S. K. Abell, & N. G. Lederman (Eds.). Handbook of research on science education (pp. 831–879). New York, NY: Routledge.
Morreale, S. P., & Backlund, P. M. (2007). Large scale assessment in oral communication: P–12 and higher education (3rd ed.). Washington, D.C.: National Communication Association.
Morreale, S. P., Cooper, P., & Perry, C. (2000). Guidelines for developing oral communication curricula in kindergarten through twelfth grade. Annandale, VA: National Communication Association.
Morreale, S. P., Moore, M., Surges-Tatum, D., & Webster, L. (2007). The competent speaker speech evaluation form (2nd ed.). Washington, D.C: National Communication Association.
Morreale, S. P., & Pearson, J. C. (2008). Why communication education is important: The centrality of the discipline in the 21st century. Communication Education, 57(2), 224–240. http://dx.doi.org/10.1080/03634520701861713.
Morreale, S. P., Spitzberg, B. H., & Barge, J. K. (2013). Communication: Motivation, knowledge, skills (3rd ed.). New York, NY: Peter Lang Publishing, Inc.
Munter, C., Wilhelm, A. G., Cobb, P., & Cordray, D. S. (2014). Assessing fidelity of implementation of an unprescribed, diagnostic mathematics intervention. Journal of Research on Educational Effectiveness, 7(1), 83–113. https://doi.org/10.1080/19345747.2013.809177.
Muthén, L. K., & Muthén, B. O. (1998-2012). Mplus User's guide (7th ed.). Los Angeles, CA: Muthén & Muthén.
Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. Theory and Research in Education, 7(2), 133–144. https://doi.org/10.1177/1477878509104318.
Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. Thinking Skills and Creativity, 10, 265–279. http://dx.doi.org/10.1016/j.tsc.2013.07.006.
Pabst-Weinschenk, M. (2005). Freies Sprechen in der Grundschule: Grundlagen; praktische Übungen. [Monologic speaking in primary school: Principles; exercises]Berlin: Cornelsen Scriptor.
Parr, C. J., & Cartwright-Hatton, S. (2009). Social anxiety in adolescents: The effect of video feedback on anxiety and the self-evaluation of performance. Clinical Psychology & Psychotherapy, 16(1), 46–54. http://dx.doi.org/10.1002/cpp.599.
Pesco, D., & Gagné, A. (2015). Scaffolding narrative skills: A meta-analysis of instruction in early childhood settings. Early Education and Developmenthttp://dx.doi.org/10.1080/10409289.2015.1060800.
Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. Journal of Educational Psychology, 108(6), 883–897. http://dx.doi.org/10.1037/edu0000085.
Rubin, R. B., & Morreale, S. P. (1996). Setting expectations for speech communication and listening. New Directions for Higher Education, (96), 19–29. http://dx.doi.org/10.1002/he.36919969604 1996.
Schick, A., & Melzi, G. (2010). The development of children's oral narratives across contexts. Early Education & Development, 21(3), 293–317. http://dx.doi.org/10.1080/10409281003680578.
Schreiber, L. (2011). Informative speaking. In The Public Speaking Project (Ed.). Public speaking. The virtual text. (pp. 15–1–15-16). Retrieved from http://www.publicspeakingproject.org/PDF%20Files/Informative%20web%201%20gs.pdf.
Schreiber, L. M., Paul, G. D., & Shibley, L. R. (2012). The development and test of the public speaking competence rubric. Communication Education, 61(3), 205–233. http://dx.doi.org/10.1080/03634523.2012.670709.
Smith, P. L., & Ragan, T. J. (2005). Instructional design (3rd ed.). Hoboken, NJ: Wiley.
Staton, A. Q., & Tomlinson, S. D. (2001). Communication education outreach in elementary school classrooms. Southern Communication Journal, 66(3), 211–223. http://dx.doi.org/10.1080/10417940109373200.
Thielemann, S. (2014). Kompetenzbereich Sprechen und Zuhören - Konzept für das Vortragen in der Schule [Competence area speaking and listening - Concept for presenting at school]. DGSS @ktuell, (1), 9–15.
Weiß, R. H. (2006). CFT 20-R mit WS/ZF-R: Grundintelligenztest Skala 2 – revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – revision (WS/ZF-R). [Culture Fair Intelligence Test-20 – Revision (CFT 20-R) with tests for vocabulary and numerical order – Revision (WS/ZF-R)]Göttingen: Hogrefe.
Wendt, H., Bos, W., Selter, C., Köller, O., Schwippert, K., & Kasper, D. (Eds.). (2016). [International comparison of elementary school children's mathematical and science competencies]. Münster, New York: Waxmann.
What Works Clearinghouse (2014). What Works Clearinghouse procedures and standards handbook, version 3.0. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.
Wilson, J. A. (1997). A program to develop the listening and speaking skills of children in a first grade classroom. University of Virginia.