Putting a speech training program into practice

Its implementation and effects on elementary school children's public speaking skills and levels of speech anxiety

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

Published in:
Contemporary Educational Psychology

DOI:
10.1016/j.cedpsych.2018.09.003

Publication date:
2018

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

Contemporary Educational Psychology 55 (2018) 176–188

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

Putting a speech training program into practice: Its implementation and effects on elementary school children's public speaking skills and levels of speech anxiety

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

[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:
Implementation fidelity
Public speaking
Randomized controlled trials
Speech anxiety
Video ratings

ABSTRACT

Successfully putting an intervention into widespread practice requires—in addition to an effective intervention—an effective implementation. On the basis of the repeated demand for public speaking promotion programs for elementary school children, a corresponding multicomponent training program was recently developed as part of an extracurricular enrichment program for talented children in Germany. The instructional goals were to promote the children's ability to appropriately use public speaking skills and to decrease their speech anxiety. The program was previously evaluated in an efficacy study with positive treatment effects. Herein, we conducted an effectiveness study in which the effects of the speech training program were replicated after it was put into practice (i.e., offered by several instructors). Specifically, 61 children and eight trained course instructors participated in the study, which used a randomized wait-list control group design with repeated measures. We assessed the effects of the training program on students' public speaking performance (i.e., their ability to appropriately apply nonverbal–visual, nonverbal–auditory, organizational, and language use skills) with video ratings. We evaluated speech anxiety with self-reports. We found positive treatment effects for both organizational public speaking skills and speech anxiety. Thus, the findings revealed that the intervention was partly successful when put into practice. The results of this effectiveness study are compared and discussed with those from the previous efficacy study.

1. Introduction

In an educational context, interventions are developed to answer a defined need to support the behavior, conditions, or development of a certain target group (e.g., Blase, van Dyke, Fixsen, & Bailey, 2012; Humphrey et al., 2016). The intervention that is at the center of the present study addresses a very important general competence of school children: public speaking competence. There is an increasing awareness that competence in giving public speeches is a core requirement for well-educated students (van Ginkel, Gulikers, Biemans, & Mulder, 2015). Whether a speaker is perceived as competent depends on the effectiveness and appropriateness of the shown public speaking skills. This is influenced by the person's knowledge and repertoire of nonverbal, organizational, and language use skills as well as by the person's motivation (e.g., speech anxiety; see Backlund & Morreale, 2015; De Grez & Valcke, 2010; van Ginkel et al., 2015). Even for the young target group of elementary school children, repeated calls for interventions that are aimed at promoting public speaking competence have been made (Hunt, Wright, & Simonds, 2014). However, only a few interventions have been developed for this age group. Furthermore, research on how to promote elementary school children's public speaking competence is similarly sparse (Hunt et al., 2014) and also often suffers from rather severe methodological limitations (De Grez & Valcke, 2010). To close this gap, a corresponding speech training program was developed and evaluated (Herbein et al., 2018)—herein, we test this speech training program when offered by several course instructors.

In general, an intervention is considered successful if its implementation results in an improvement in the intended outcomes (Fixsen, Blase, Metz, & van Dyke, 2013). In order to provide a successful intervention for dissemination in practice, different stages are necessary to ensure effectiveness and practicability (see Fig. 1 based on, e.g., Humphrey et al., 2016; Lendrum & Wigelsworth, 2013). First, the

* 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), ingo.zettler@psy.ku.dk (I. Zettler), ulrich.trautwein@uni-tuebingen.de (U. Trautwein).

https://doi.org/10.1016/j.cedpsych.2018.09.003

Available online 07 September 2018

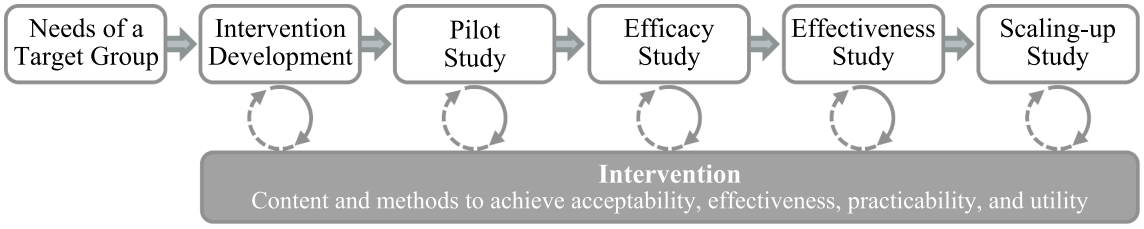

Fig. 1. Putting an intervention into practice: The six stages from identifying the needs of a target group to scaling-up (based on e.g., Humphrey et al., 2016; Lendrum & Wigelsworth, 2013).

instructional goals need to be deduced from the needs of the target group (Humphrey et al., 2016). Second, the intervention needs to be developed on the basis of a change model (Nelson, Cordray, Hulleman, Darrow, & Sommer, 2012) that is grounded, for example, in psychological theories (see Graham, McKeown, Kiuhara, & Harris, 2012; Lendrum & Wigelsworth, 2013). Next, pilot, efficacy, and effectiveness studies should be conducted to investigate the intervention's acceptability, practicability, and utility as well as effectiveness. Whereas an efficacy study is applied to examine whether an intervention is successful under optimal and controlled implementation conditions, an effectiveness study is conducted to examine the effects of an intervention when it is implemented in the real world (Gottfredson et al., 2015; Greene, 2015). Finally, scaling-up studies follow. These can be aimed at increasing the impact of the intervention by broadening the intended target group and training context (Gottfredson et al., 2015).

The current study focused on the stepwise development of an intervention targeting the improvement of elementary school children's public speaking skills. Based on a change model grounded in psychological theories, a multicomponent intervention for elementary school children was previously conceptualized, piloted, and evaluated in an efficacy study (Herbein et al., 2018). The results indicated that public speaking skills can already be fostered in elementary-school-aged children. Whereas this efficacy study showed a high level of internal validity, conclusions regarding whether the treatment effects would hold if the training were offered under less strict conditions (e.g., by using nonresearch course instructors) remained unclear. Consequently, the current study tested the intervention's effectiveness in a real-world implementation (i.e., with several course instructors) for the same outcomes that Herbein et al. (2018) used.

1.1. Background on the present study

1.1.1. Fostering elementary school children's public speaking skills: Need for an intervention

In comparison with reading and writing skills, the promotion of speaking (and listening) skills is often neglected in education (EDK, 2010; Wilson, 1997). This is problematic because public speaking competence is required not only for a single subject (specifically language arts) in school but also for the whole curriculum as it serves as a tool for learning and knowledge assessment (Lee, Quinn, & Valdes, 2013). For instance, in school subjects such as mathematics and science, students need to be able to present their knowledge and ideas in informative speeches. Thereby, students' abilities to competently transmit their knowledge influence assessments of their knowledge and skills (Min, 2014). The ability to give an informative speech is also relevant outside the curriculum when students are involved in extracurricular activities (e.g., STEM competitions such as the European Union Contest for Young Scientists) and, of course, later in life in the workplace (Morreale, Valenzano, & Bauer, 2016; van Ginkel et al., 2015).

Even as early as elementary school, children face the task of needing to give informative public speeches (Common Core State Standards Initiative, 2010; EDK, 2010; Kultusministerkonferenz, 2005). However, research on public speaking and its promotion in elementary and secondary schools is still scarce when compared with higher education (Hunt et al., 2014). Further, the majority of studies examining the effectiveness of public speaking interventions for younger students have been characterized by a number of methodological limitations (De Grez & Valcke, 2010). For instance, some studies failed to randomize students into intervention and control groups, did not employ a control group at all, or used only one source for assessing effectiveness measures (e.g., only students' self-reports), potentially resulting in less generalizable conclusions. Based on the need for and the lack of public speaking interventions, a new speech training program for the target group of elementary school children was developed and aimed at promoting their public speaking skills (see Herbein et al., 2018).

1.1.2. Conceptualization of the public speaking training program

The speech training program was designed for third and fourth graders participating in an extracurricular enrichment program and a group size of six to 10 students. The training consisted of 11 course units, each lasting 90 min. The training program was based on a conceptual framework for public speaking competence. The framework summarized theoretical and practical approaches from different fields (e.g., communication studies, psychology, rhetoric, and speech science) and took into consideration theoretical frameworks, empirical studies, educational standards, and assessment tools. It suggests that the ability to show competent public speaking behavior is influenced by the manifestation and combination of three underlying dimensions of public speaking competence (also referred to as components; see e.g., Morreale, Spitzberg, & Barge, 2013): knowledge, motivation, and skills (Backlund & Morreale, 2015; De Grez & Valcke, 2010; van Ginkel et al., 2015). *Knowledge* incorporates the theoretical background of how to prepare and deliver a speech (Morreale et al., 2013). *Motivation* comprises a person's willingness to communicate and is determined by different components (e.g., by communication apprehension or speech anxiety; Croucher, 2013). Finally, a repertoire of *skills* is required for a person to be able to speak competently. The skills can be grouped into four dimensions: (a) nonverbal behavior–visual impression (e.g., eye contact, gestures, posture), (b) nonverbal behavior–auditory impression (e.g., articulation, fluency, pitch, voice, volume), (c) language use (e.g., use of rhetorical devices or technical terms), and (d) organization (e.g., structure of the speech, reasonable selection of information).

Overall, a speaker's public speaking competence depends on the effectiveness (i.e., the achievement of the communicational goal) and the appropriateness of the speech (Morreale, Moore, Surges-Tatum, & Webster, 2007). Thus, the perception of public speaking competence is based on a speaker's actual implementation of public speaking behavior in a specific context. In order to exhibit contextually appropriate behavior (i.e., appropriate for the specific situation, the target audience, the communicational intention, the topic, and for certain norms and standards), specific skills need to be selected, adapted, and used. This process is guided by the speaker's knowledge and motivation (Backlund & Morreale, 2015). Whether or not a speaker is perceived as competent depends on the extent to which he or she demonstrates these skills and whether or not he or she appropriately addresses the context. However, inferences about motivation and knowledge can be made as a function of observed public speaking behavior to some extent.

The content of the training program was derived on the basis of the

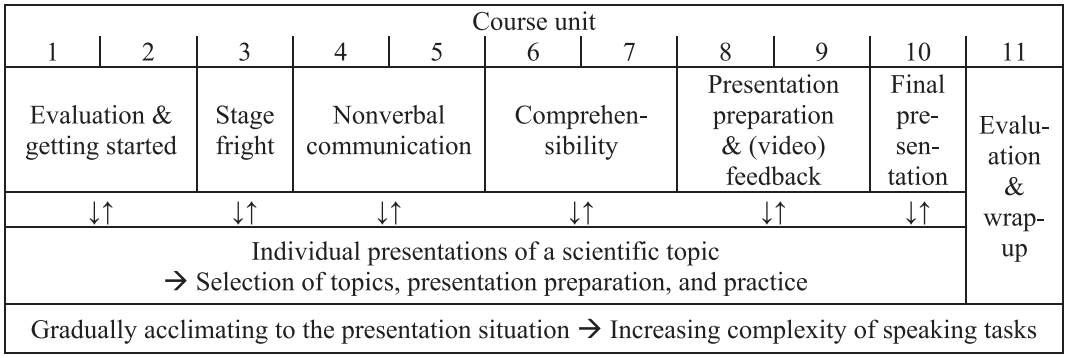

Fig. 2. Contents and structure of the speech training program (based on Herbein et al., 2018).

framework for public speaking competence. Thus, the final program included—in order of thematization—stage fright (Course Unit 3; see Fig. 2 for the course structure and contents), nonverbal communication (visual and auditory impression; Course Units 4 and 5), and comprehensibility (language use and organization of speech; Course Units 6 and 7). These topics were framed by the introductory and closing parts of the course (Units 1 and 2, and 8 to 11). The closing parts focused on the preparation and delivery of a final talk on a self-chosen subject. For all topics, the focus was on the public speaking task of informing others.

Across the course units, the course gradually enabled a deepening of knowledge, an extension of the student's repertoire of skills, an increase in confidence when speaking (i.e., motivation), and a reflection on the adequate application of behavior (contextual appropriateness). Thereby, the training methods were derived from the underlying core components that were defined on the basis of research on the promotion of public speaking competence in higher education (van Ginkel et al., 2015): behavior modeling, opportunity to practice, different forms of feedback (peer and trainer feedback), self-assessment via video feedback, and phases of transfer by giving the children the opportunity to work on their own scientific speech topics. More specifically, each course unit began with an introductory game and, depending on the instructional goal, was followed by theoretical instruction provided by the instructor, practical exercises, conversations in class, demonstrations by the instructor, observation tasks, phases in which students provided or received feedback, or work on the student's own topic. Finally, each course unit ended with a closing ritual. Across the course units, the difficulty and complexity of practical speaking tasks increased steadily, that is, from short impromptu speeches sitting in the group to 3- to 5-min talks in front of the others at the end of the training. For more information about the speech training program and its theoretical background, see Herbein et al. (2018).

1.1.3. The challenge of successfully implementing an intervention

Before a psychological intervention is put into practice, its efficacy and practicability need to be demonstrated (Fixsen et al., 2013). These qualities can be supported by ensuring that the intervention passes through six different stages, whereby the single stages can be repeated if necessary (see Fig. 1 based on, e.g., Humphrey et al., 2016; Lendrum & Wigelsworth, 2013). Transferring this procedure to the public speaking training program, Stages 1 to 5 were made concrete as follows.

After defining the need for speech training for elementary school children (Stage 1) and choosing the contents and teaching methods (Stage 2), a pilot study followed (Stage 3). In general, the aim of the piloting stage is to examine the intervention's practicability, acceptability, appropriateness, and utility (Humphrey et al., 2016). In the piloting stage of the speech training program, the intervention was offered by the course developer to two groups of eight and nine children, respectively. After examining and incorporating feedback from the pilot phase into the training concept, the efficacy of the training program was evaluated (Stage 4).

The aim of an efficacy study is to examine the improvements in predefined outcome variables (Bywater, 2012). To avoid confounding influences, the efficacy study should be conducted in a controlled setting (Hulleman & Cordray, 2009). To do so for the speech program, the course was offered as an extracurricular enrichment program for elementary school children. Only one course instructor, namely, the first author and developer (a psychologist and speech scientist) offered the course, whereby she followed a scripted manual in implementing the course, and the same teaching materials were used in all groups. A randomized controlled trial with pretest and posttest measures and a treated control group were employed to assess the efficacy of the trainings. Treatment effects on public speaking skills were assessed with external ratings of videotaped presentations, and the treatment effect on speech anxiety was assessed via self-reports. The findings of the efficacy study revealed that the intervention successfully improved most of the public speaking skills it was designed to improve. No treatment effects were found on speech anxiety (Herbein et al., 2018).

When an efficacy study of an intervention shows good results, the intervention can subsequently be implemented in the field. Accordingly, an investigation of the intervention's effectiveness under less controlled conditions is usually suggested (Gottfredson et al., 2015). Owing to the positive findings in the study in which the efficacy of the speech training was evaluated, the aim of the present study was to conduct an effectiveness study.

1.2. The present study

On the basis of the repeated demand for public speaking interventions for elementary school children (Hunt et al., 2014) and the first promising results of a corresponding intervention (Herbein et al., 2018), with the present study, we took the next step toward the practical implementation of the intervention. While keeping the basic conditions identical to those used in the efficacy study (e.g., a training program offered as an enrichment program for elementary school children with the same course structure and contents), we relaxed the standardized delivery of the intervention: Instead of having the course developer offer the training program, a group of nonresearch course instructors offered it in a natural setting. When an intervention is put into practice, the extent to which the intervention is implemented as intended by the developer—called implementation fidelity—needs to be considered (Humphrey et al., 2016). To maximize implementation fidelity in the present study, course instructors participated in a workshop and received a course manual and all of the teaching materials. The effectiveness of the program with respect to public speaking skills and speech anxiety was assessed in a randomized wait-list control group design. The same measures as used by Herbein et al. (2018) were applied: Public speaking skills were assessed with video ratings, and speech anxiety was assessed with self-reports. Implementation fidelity was assessed by administering self-reports to the course instructors. On the basis of the findings of the efficacy study by Herbein et al. (2018),

we had the following hypotheses:

Hypothesis One: We expected that the training program would foster students' ability to adapt their public speaking skills to the context. Consequently, we hypothesized that children in the training group would show more appropriate public speaking skills than children in the control group as assessed via external ratings.[1]

Hypothesis Two: We predicted that the training program would reduce speech anxiety. Specifically, we expected that the children in the training group would report fewer cognitive and bodily speech anxiety symptoms compared to the children in the control group.

2. Method

We first describe the participants taking part in the study, and then present the study procedure. This includes a delineation of the steps that were undertaken to prepare the implementation, the research design used, and information on rater training. This is followed by a description of the measures used, a section on missing data, and the analyses.

2.1. Participants

The speech training program was offered in the second half of the 2014-2015 school year. It was part of the Hector Children's Academy Program (HCAP), an extracurricular enrichment program for talented elementary school children. The HCAP is a statewide promotion program in the German state of Baden-Württemberg. This program addresses children who have been nominated by their teacher. At 65 different local HCAP sites (as of June 2017), accepted children can be enrolled in a variety of extracurricular afternoon/weekend courses. The courses are offered not only by teachers but also by a large number of external course instructors who have different kinds of professional backgrounds (e.g., architects, computer scientists, engineers, graphic designers, or interpreters).

In order to recruit course instructors for the speech training program, we sent informational material about the course contents and the implementation study to all local HCAP sites 5 months before the study began. Eleven course instructors registered for the course, agreed to participate in the study, and provided written informed consent. However, two course instructors could not offer the course because personal issues came up, and one instructor needed to cancel the course because only a small number of students enrolled. The remaining eight participating course instructors were all women and had a mean age of 46.50 (SD = 14.21) years. They had worked between 0 and 13 years at a local HCAP site (M = 4.71, SD = 4.15 years) and had offered a mean number of 10 courses in the past (SD = 11.85). For more details on the course instructors, see Table 1.

Before the study began, written parental consent was received for the children's participation. The study participants consisted of 61 (46% girls) third- and fourth-grade students. They were enrolled at one of the eight local HCAP sites that offered the public speaking training program. The mean age of the children was 9.10 (SD = 0.84) years. They showed a mean verbal IQ of 110.90 (SD = 15.05; see Section 2.3).

2.2. Procedure

To achieve high implementation fidelity and to examine the effectiveness of the speech training program, the study included several steps. To increase implementation fidelity, a workshop for course instructors was scheduled before the pretest. To assess the treatment effects, a randomized controlled trial with pre- and posttest measures was used. The dependent variables (i.e., appropriateness of the public speaking skills, assessed via external ratings of videotaped presentations, and speech anxiety, assessed via self-reports) were assessed at both measurement points. To this end, the children gave a short presentation and filled out questionnaires before the training began and after it ended. The control variables and the characteristics of the instructors were assessed at pretest only. After the children had completed the posttest, external raters were trained to rate the children's videotaped presentations to examine the appropriateness of their public speaking skills. Interrater reliability was examined afterwards. All steps of the study are summarized in chronological order in Fig. 3 and are described in more detail next.

Table 1
Descriptive statistics for the sample of instructors.

	M	*SD*
Age	46.50	14.21
Communication competence	82.24	12.31
Communication apprehension–public speaking	2.19	0.65
Theoretical knowledge (self-reports)		
Emergence of stage fright	3.43	0.54
Symptoms of stage fright	3.71	0.49
Handling stage fright	3.57	0.54
Nonverbal–visual	3.57	0.54
Nonverbal–auditory	3.57	0.54
Importance of nonverbal communication	3.71	0.49
Comprehensibility	3.86	0.38
Structure	3.86	0.38
Feedback rules	3.71	0.49
Video feedback	3.29	0.76
Notes	3.71	0.49
Visualization	3.86	0.38
	Yes	*No*
Pedagogical qualifications	5	3

Note. N = 8.

2.2.1. Preparing the implementation

The process of putting an intervention into practice is called implementation (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005). Across different frameworks, eight dimensions of implementation are typically distinguished (e.g., Durlak & DuPre, 2008; Humphrey et al., 2016; Nelson et al., 2012). These can be allocated to the behavior of the instructor (adherence, dosage, quality, adaption), the behavior of the participants (responsiveness, reach), and the characteristics of the intervention and control conditions (program differentiation, monitoring of control conditions; Durlak & DuPre, 2008; Humphrey et al., 2016). Based on these dimensions, implementation fidelity can be assessed (Humphrey et al., 2016). High implementation fidelity is the core of a successful intervention in practice. Generally, the smaller the gaps between the intentions of the developer and the actual behavior of the instructors and participants as well as the characteristics of the conditions, the higher the effects of the intervention on the intended outcomes (see Durlak & DuPre, 2008; O'Donnell, 2008). In addition, implementation fidelity needs to be assessed to examine whether (a lack of) improvements in the intended outcome are due to the (lack of) effectiveness of the intervention (Munter, Wilhelm, Cobb, & Cordray, 2014) or due to other factors (e.g., implementation or evaluation failure; Greene, 2015; Humphrey et al., 2016).

The shift from controlled efficacy studies to real-world implementations is often associated with a change in instructors. This change is often associated with low implementation fidelity, and thus, the effects on the intended outcomes are typically reduced (Greene, 2015; Hulleman & Cordray, 2009; Lendrum & Wigelsworth, 2013). One

[1] Given that Herbein et al. (2018) found high correlations between the extent and appropriateness of skills (with a *Mdn* correlation = .83), they considered only appropriateness in the analyses, on the basis of the assumption that the ability to behave appropriately reflects a higher level of public speaking competence (Rubin & Morreale, 1996; Staton & Tomlinson, 2001). We followed the same approach.

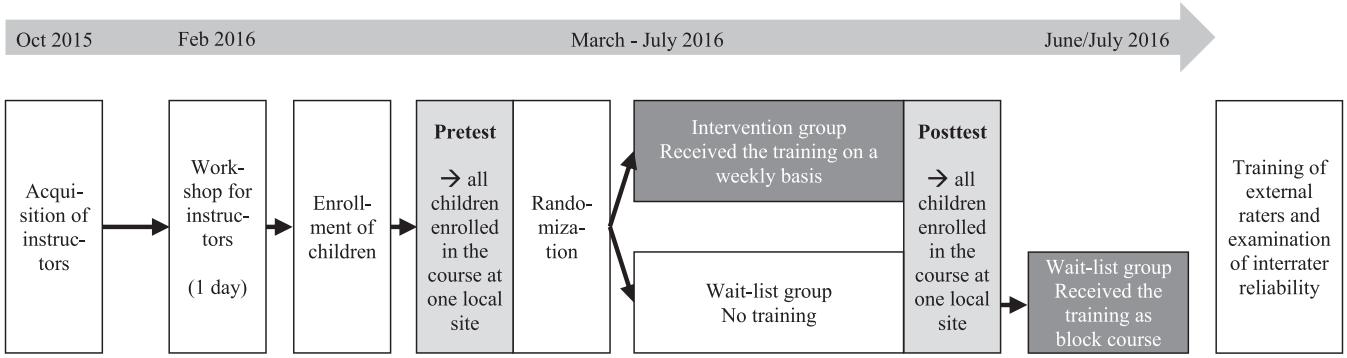

Fig. 3. Procedure used in the effectiveness study.

reason for this low implementation fidelity is that, in practice, instructors might not completely adhere to the intervention as outlined but might modify it (Greene, 2015; Humphrey et al., 2016; Rimm-Kaufman et al., 2014). Thus, to promote high implementation fidelity, specific factors concerning the instructors and their behavior need to be addressed. For example, this includes developing support systems (e.g., workshops for instructors) and supplying instructors with information and teaching materials that will help them implement the intervention (Bywater, 2012; Humphrey et al., 2016; Souvignier & Mokhlesgerami, 2006).

In the present effectiveness study, several steps were taken to enhance implementation fidelity. The aim was to increase instructors' knowledge and acceptance of the intervention while simultaneously reducing their financial, personal, and temporal costs. Specifically, the course instructors were given a scripted course manual[2] and were asked to participate in a 1-day workshop prior to the course. The course manual contained a description of the general theoretical background of public speaking competence, detailed explanations of each of the 11 course units, as well as information about the goal, introduction, implementation, conclusion, and transferring of each exercise. Furthermore, time frames were given for all exercises (for an extract from the manual, see Appendix A).

The 1-day workshop was offered by the first author, who had developed the course. Two groups of five and six instructors participated, respectively. The workshop consisted of four parts: First, the theoretical background of the course was presented. Second, the course instructors were familiarized with the course manual and its construction and contents. Third, each course unit was introduced in more detail, including a short theoretical introduction, the goals, the exercises, and the related materials. Fourth, organizational matters were clarified with respect to the evaluation of the implementation of the course. The teaching methods included lectures, practical exercises, and (small group) discussions. To achieve a high level of implementation fidelity, the course instructors were given all of the teaching materials (e.g., all materials needed to conduct the exercises) in addition to the course manual.

2.2.2. Research design

A randomized controlled trial with repeated measures (pretest and posttest measures; Humphrey et al., 2016) was conducted to assess the effectiveness of the training program with respect to public speaking skills and speech anxiety. To be able to draw valid causal inferences about the effectiveness of a treatment, the standard practice is to employ a control group (Gottfredson et al., 2015). Furthermore, assigning the participants randomly to either the intervention or the control group is considered the best method for generating unbiased estimates of an intervention's treatment effects (Gottfredson et al., 2015; Humphrey et al., 2016). In the present study, blocked randomization was conducted on the basis of computer-generated random numbers. Owing to registration numbers smaller than seven, cluster randomization had to be carried out for three local HCAP sites. On the basis of the randomization, 28 children (61% girls, age: $M = 9.17$, $SD = 0.86$) participated in the intervention group. They received the intervention on a weekly basis. The other 33 children (33% girls, age: $M = 9.05$, $SD = 0.84$) joined the wait-list control group, receiving the public speaking training program in a block course after the posttest. Before participants found out whether they were assigned to the intervention or the control group, all children took the pretest. The pretest was administered during the intervention group's first course unit. All children who were enrolled in the speech training program (i.e., in either the intervention or the control group) at one local HCAP site participated. The course instructors (at pretest) as well as the test leaders (at pre- and posttest) were blind to the children's group assignment. The posttest took place during the intervention group's last course unit, in which, again, all children from both groups participated. Afterwards, the control group participated in the intervention, offered as a block course. In contrast to the weekly courses that the intervention group took, the block course was organized into fewer but longer course lessons. For example, all course units were offered directly after one another in two consecutive Friday/Saturday sessions.

2.2.3. External rater training

The appropriateness of public speaking skills was assessed by external raters. To ensure high interrater reliability, all raters participated in a training session before examining the videotaped presentations. In addition, they were given a manual. For more information on the external video ratings, see Section 2.3.

2.3. Measures

Implementation fidelity was assessed continuously as the instructors taught the course. The dependent variables (appropriateness of the public speaking skills and speech anxiety) were assessed at pre- and posttest using the same measures. The control variables (i.e., general fluid abilities and verbal intelligence) and instructor characteristics were assessed at pretest only.

2.3.1. Implementation fidelity

How implementation fidelity is assessed is always unique to some extent because it strongly depends on the particular intervention (Abry, Hulleman, & Rimm-Kaufman, 2015). When examining the effectiveness of an intervention on students' outcomes, it is reasonable to conclude that the quantity as well as the quality of implemented exercises affects children's development (Greene, 2015; Odom et al., 2010). Thus, in this study, we assessed implementation fidelity along two dimensions: *adherence* (i.e., compliance with the exercises in the course manual) and *quality of delivery* (i.e., the manner of implementation and thus how

[2] The course manual was prepared and then pretested by three speech scientists offering the course after the pilot study.

Table 2
Implementation fidelity: adherence and quality of delivery.

	Course instructors												Total	
	A		B		C		D		E		F			
Course unit	Adh. %	Quality *M* (*SD*)	Adh. %	Quality *M* (*SD*)	Adh. %	Quality *M* (*SD*)	Adh. %	Quality *M* (*SD*)	Adh. %	Quality *M* (*SD*)	Adh. %	Quality *M* (*SD*)	Adh. % *M* (*SD*)	Quality *M* (*SD*)
2	100	3.00 (0)	100	3.00 (0)	100	2.50 (0.55)	92	2.80 (0.45)	100	2.33 (0.82)	100	3.00 (0)	98.67 (3.27)	2.77 (0.49)
3	100	2.86 (0.38)	80	n.a.	100	2.00 (0.82)	93	3.00 (0)	87	2.50 (0.55)	93	3.00 (0)	92.17 (7.73)	2.67 (0.59)
4	79	2.83 (0.41)	86	n.a.	79	3.00 (0)	86	2.83 (0.41)	79	2.40 (0.89)	100	2.71 (0.49)	84.83 (8.18)	2.77 (0.50)
5	85	2.80 (0.45)	100	2.57 (0.53)	92	2.33 (0.52)	77	3.00 (0)	92	2.60 (0.55)	100	2.86 (0.38)	91.00 (8.90)	2.69 (0.47)
6	100	2.43 (0.53)	100	n.a.	100	2.29 (0.49)	100	n.a.	63	2.00 (1.00)	100	2.71 (0.49)	93.83 (15.11)	2.42 (0.62)
7	57	2.00 (0)	100	2.86 (0.38)	86	2.20 (0.45)	86	2.50 (0.55)	57	2.00 (0)	100	3.00 (0)	81.00 (19.62)	2.55 (0.51)
8	37	2.67 (0.58)	95	2.86 (0.38)	68	2.50 (0.55)	89	2.86 (0.38)	68	2.00 (0.82)	100	3.00 (0)	76.17 (23.47)	2.71 (0.52)
9	88	3.00 (0)	58	1.75 (0.50)	100	1.89 (0.33)	100	3.00 (0)	58	n.a.	100	2.67 (0.50)	84.00 (20.67)	2.55 (0.60)
10	50	3.00 (0)	64	n.a.	100	2.14 (0.38)	71	3.00 (0)	50	n.a.	64	2.00 (0.82)	66.50 (18.44)	2.44 (0.62)
Total	77	2.73 (0.42)	87	2.63 (0.53)	92	2.32 (0.56)	88	2.84 (0.36)	73	2.35 (0.72)	95	2.77 (0.44)	85.33 (8.59)	2.63 (0.55)

Note. Course Units 1 and 11 are missing because the pretest and posttest were conducted during these units. Two course instructors are missing because they offered the block course only for the wait-list control group due to cluster randomization. Adh. = adherence (i.e., the percentage of exercises conducted by each course instructor and in each course unit). Quality = quality of delivery. n.a. = no data were available because the instructor reported the quality for less than 75% of the conducted exercises in this unit, and thus, no mean was calculated.

well the exercise was delivered; see Humphrey et al., 2016; Nelson et al., 2012; Schoenwald et al., 2011). For this purpose, course instructors filled out a questionnaire after each course unit. To assess adherence, they first identified whether each exercise was conducted or not (item: "Was the exercise conducted?"; dummy-coded: 0 = *no*, 1 = *yes*). As the exercises differed in the extent to which they were important for reaching the instructional goals, weights were assigned to each exercise ranging from 1 = *little importance* to 3 = *great importance*. Overall, the course consisted of 68 exercises. Weighing the exercises by importance, i.e., from 1 to 3, led to a maximal total score of 181 for adherence. To describe adherence, the percentage of exercises that were conducted was calculated for each course instructor. To assess quality of delivery, for each exercise, the instructors reported how well its implementation had worked (item: "How did the implementation go?"; 3-point Likert-type scale: 1 = *not well*, 2 = *okay*, 3 = *well*). Because one of the instructors had a great deal of missing data, quality could be ascertained for only five out of the nine course units for this instructor (Instructor B; see Table 2).

2.3.2. Public speaking skills

To measure public speaking skills, we used a standardized *public speaking task* and videotaped the resulting presentations the children gave. The public speaking context was the same for the pretest and posttest, that is, the children faced the same task, preparation condition, material, and speaking situation at each measurement point. Namely, the children's task was to draw a scientist (Chambers, 1983) on a poster and to consider what the scientist might be doing throughout an entire day. They were told that they would need to present information about their poster and ideas in a speech after they were done with the poster. They prepared their speeches in a group setting for 15 min by creating the poster on a prepared template. After creating their speeches, the children were randomly assigned to one of three different test leaders. Next, each test leader accompanied one child to a separate room where each child subsequently presented his or her speech. This speaking situation—presenting separately in front of one test leader instead of in front of the class—was chosen to prevent the children from imitating each other and to rule out effects of learning from observation (for a discussion of the advantages of this approach vs. having students present in front of the class, see Herbein et al., 2018). The order in which the children presented was randomly assigned.

On the basis of the short videotaped presentations, the students' *public speaking skills* were rated. A previously developed observation sheet was used to capture the appropriateness of the students' public speaking skills (i.e., the ability to adapt one's public speaking skills to a particular context). The appropriateness of the public speaking skills was assessed via 34 items (e.g., "Eye contact is situationally appropriate"). Each public speaking skill was assigned to one of the four dimensions: nonverbal behavior–visual, nonverbal behavior–auditory, language use, and organization of speech. For more information about the instrument, see Table 3. Ratings were made with a 4-point Likert-type scale (ranging from 1 = *strongly disagree* to 4 = *strongly agree*). Three external raters were trained to use the observation sheet.[3] All three raters rated all videotaped presentations. The videos were presented in a random order, and the raters were blind to measurement point and group membership. To achieve satisfactory precision in the interrater reliability estimates (Bonett, 2002), all available video data on comparable speeches by elementary school children were used. These videos stemmed from this study ($N = 114$) and the previous efficacy study ($N = 110$). Thus, all three raters rated a total of 224 videos. With this design, interrater reliability was examined by using two-way, absolute, average-measure ICCs for each item across all 224 videos. We included only items with good (between .60 and .74) or excellent ($\geq$ .75) interrater reliabilities (Cicchetti, 1994), resulting in 11 items for

[3] They received a manual and three repeated practical training sessions offered by a speech and rating expert. In the training program, videos of short presentations of children performing the same task were used. Afterwards, the ratings of the presentations conducted in this study began.

Table 3
Observation sheet for assessing the situational appropriateness of public speaking skills.

Dimension		Assessed public speaking skills
Nonverbal– visual	6 items	Posture, gestures, mimic, eye contact, proxemics, use of notes
Nonverbal– auditory	15 items	Speech respiration, pitch of voice, variation of voice, articulation, accentuation, variation of accentuation, volume, variation of volume, speech rate, variation in speech rate, breaks, variation in breaks, intonation, variation in intonation, speech fluency
Language use	4 items	Linguistic expression, use of rhetorical devices, activation, personal address
Organization	8 items	Structure, amount of information, length of speech, length of introduction, length of conclusion, visualization, intention of communication, reference to listener
	1 item	Global performance

Note. The exact wording of each item followed the structure "*Public speaking skill* is situationally appropriate" (based on Herbein et al., 2018).

Table 4
Means and standard deviations of all variables for each measurement point and group.

	Pretest					Posttest						α	ICC
	Speech training		Control group		MIS	Speech training		Control group		MIS	Items	t1/t2	
	M	*SD*	*M*	*SD*		*M*	*SD*	*M*	*SD*				
Dependent variables (video ratings)													
Nonverbal–visual													
Posture	3.13	0.67	3.08	0.49	0	3.42	0.54	3.10	0.58	13	1		.80
Gestures	2.50	0.89	2.57	0.72	0	2.85	0.61	2.71	0.63	13	1		.74
Eye contact	2.95	0.59	2.92	0.58	0	3.06	0.49	2.93	0.60	13	1		.63
Nonverbal–auditory													
Speech respiration	3.44	0.78	3.39	0.54	0	3.44	0.46	3.24	0.69	13	1		.75
Breaks	3.29	0.62	2.98	0.70	0	3.13	0.59	3.11	0.46	13	1		.64
Speech fluency	3.47	0.55	3.18	0.55	0	3.49	0.48	3.28	0.43	13	1		.79
Organization													
Length of speech	2.96	0.62	3.00	0.71	0	3.10	0.56	2.86	0.56	13	1		.66
Length of introduction	2.13	0.47	2.14	0.59	0	2.59	0.96	2.02	0.84	13	1		.61
Length of conclusion	1.33	0.29	1.41	0.53	0	1.78	0.71	1.33	0.45	13	1		.69
Reference to listener	2.80	0.44	2.84	0.40	0	3.40	0.40	3.17	0.43	13	1		.61
Global performance	2.88	0.50	2.87	0.50	0	3.06	0.38	2.78	0.54	13	1		.77
Dependent variables (self-reports)													
Speech anxiety cognitive	2.21	0.85	1.86	0.71	0	1.94	0.94	1.99	0.85	13	9	.88/.93	
Speech anxiety bodily	1.96	0.72	1.70	0.51	0	1.79	0.70	1.77	0.63	13	9	.80/.82	
Control variables[a]													
General fluid abilities	10.08	2.84	9.96	2.28	16						16	.57	
Verbal intelligence	110.11	15.60	111.58	14.77	0						30	.85	

Note. [a]Gender was used as another control variable (speech training: $N = 28$, 61% female; control group: $N = 33$, 33% female). α t1/t2 = Cronbach's alpha at pretest/posttest. ICC = two-way, absolute, average-measure ICCs for the total sample of 224 videos (including pretest and posttest videos) based on three raters. Conducting the interrater reliability of these 11 items separately for the 114 videotaped presentations in this study, the ICCs ranged from .50 to .78 with six items having an ICC > .60. MIS = Missing data in %. The variation in sample size is due to occasional missing data except for the variable general fluid abilities: There, missing values trace back to nonimplementation of the test because of time problems at one local HCAP site.

the analyses (see Table 4). Ten of the 11 items could be assigned to the public speaking skill dimensions as follows: nonverbal behavior–visual: *posture*, *gestures*, and *eye contact*; nonverbal behavior–auditory: *speech respiration*, *breaks*, and *speech fluency*; organization of speech: *length of speech*, *length of introduction*, *length of conclusion*, and *reference to listener*. For language use, no item had an ICC ≥ .60. The 11th item referred to the overall appropriateness of the given presentation. Note that in the previous study, which used the same observation sheet (Herbein et al., 2018), three additional items showed an ICC > .60: *mimics* and *proxemics* (nonverbal behavior–visual) and *variation of breaks* (nonverbal behavior–auditory). Although excluded from the analyses in the current study, the effects of the training on these three items are reported for comparison in Appendix B. After the videos were rated, the effectiveness of the training was examined on the 11 items and based on the subsample of videotaped presentations belonging to this study (114 videos; 61 pretest and 53 posttest; mean duration: 56 s, $SD = 25.27$ s).

2.3.3. Speech anxiety

The German version of the Performance Anxiety Questionnaire (Cox & Kenardy, 1993; German: Fehm & Hille, 2005) was used to assess speech anxiety. The questionnaire assesses cognitive (10 items; e.g., "I am worried about my presentation") and bodily (10 items; e.g., "I have sweaty palms") symptoms. Two items were excluded because of poor corrected item-total correlations. These items came from the cognitive symptoms subscale (i.e., "It is very important for me to give a good presentation"; $r_{it} = .08$ for t1 and $r_{it} = .10$ for t2) and the bodily symptoms subscale (i.e., "I could faint"; $r_{it} = -.08$ for t1 and $r_{it} = .18$ for t2). Each item was rated on a 5-point Likert scale ranging from 1 = *never* to 5 = *always*. The children were asked to think about a situation in which they had to give a speech when they rated the items, and the items were administered before the children had to give their speeches in the current study.

2.3.4. Control variables

In order to enhance the precision of the regression coefficients and

to control for any bias that could result from a lack of baseline equivalence between the two groups (Cohen, Cohen, West, & Aiken, 2003), we included control variables. General fluid abilities and verbal intelligence were selected because the training program was offered as an enrichment program for talented students, and previous studies on the HCAP have shown that one factor for being nominated for participation is students' intelligence (see Rothenbusch, Zettler, Voss, Lösch, & Trautwein, 2016). *General fluid abilities* were measured via the figural abilities subscale from the BEFKI-short (Schroeders, Schipolowski, Zettler, Golle, & Wilhelm, 2016). The test consists of 16 items. *Verbal intelligence* was assessed with the 30 items of the vocabulary subtest from the German version of the Culture Fair Intelligence Test (CFT 20-R; Weiß, 2006).

2.3.5. Instructor characteristics

Communication competence was assessed with the Self-Perceived Communication Competence Scale (McCroskey & McCroskey, 1988). Instructors were asked to indicate in an open space next to each of the 12 items how competent they believed they were in the specific communication situation that was described (0 = *completely incompetent* to 100 = *competent*). Communication apprehension–public speaking was assessed with six items from the Personal Report of Communication Apprehension (PRCA24; McCroskey, Beatty, Kearney, & Plax, 2009; 5-point Likert-type scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*). Twelve items were used to assess the instructors' theoretical knowledge of the contents of the speech training at pretest (i.e., after the workshop). The items were answered on a 4-point Likert-type scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. In addition, the course instructors were asked whether they had a pedagogical qualification (0 = *no*, 1 = *yes*).

2.4. Missing data

Missing data resulted from nonresponses to single items or the absence of students at posttest. Overall, missing data for all variables, both groups, and measurement times ranged from 0% to 16% (Table 4). A closer look at the missing data at posttest revealed no differential rate of missing data between the treatment and control groups because four children were missing from each group. In addition, there were no differential missing patterns for the groups on the pretest variables (all p-values $> .05$). However, a comparison of the means of the dependent variables at pretest for the children missing at posttest with the children who were not missing at posttest showed significant differences on the *global performance* variable in both groups. Lower scores were found for children missing at posttest in the control group, $U(4, 29) = 16$, $p = .016$, and in the treatment group, $U(4, 24) = 15$, $p = .023$. Furthermore, lower scores were found for children missing at posttest in speech fluency in the treatment group, $U(4, 24) = 16$, $p = .029$. For all other variables, no significant differences were found (all p-values $> .05$). Because we assumed that the data were missing at random, we used full-information maximum likelihood (FIML) estimation (Enders, 2010) for our analyses.

2.5. Analyses

We used multiple linear regression analyses to analyze the effectiveness of the training program. Analyses were computed in Mplus Version 7 (Muthén & Muthén, 1998–2012) with maximum likelihood robust estimation (MLR).[4] One-tailed tests of significance ($\alpha = .05$) were used to assess the treatment effects because we formulated directed hypotheses. We investigated treatment effects on 13 dependent variables. They covered public speaking skills and speech anxiety.

A model was run for each dependent variable, using its posttest value. For any dependent variable that came from the video ratings, the mean of all three raters was used in the analysis. In addition, the pretest score on the respective dependent variable was included as a predictor variable to eliminate any bias that may have been due to baseline differences between the two groups. To assess differential effects for children with low versus high pretest scores on the dependent variable, the interaction between the treatment and the pretest score was included as an additional predictor variable (Cohen et al., 2003). Furthermore, three control variables were included as predictor variables: general fluid abilities,[5] verbal intelligence, and gender.

Prior to the analyses, all continuous variables (dependent and control variables) were standardized. Treatment was dummy-coded (1 = *speech training* and 0 = *control group*), and so was gender (1 = *boys* and 0 = *girls*). Because the dependent variable was standardized, the effect size was the same as the regression coefficient of the treatment indicator. In line with Herbein et al. (2018), and because there are no other comparable studies, we used the classification of effect sizes provided by Cohen (1992) as the reference standard: small: $d = 0.20$, medium: $d = 0.50$, large: $d = 0.80$.

3. Results

Implementation fidelity must be established before an intervention's effectiveness can be explored any further. Thus, the results on implementation fidelity (adherence and quality) are reported first. Next, treatment effects on public speaking skills are reported, focusing on global performance as well as organizational, nonverbal–visual, and nonverbal–auditory public speaking skills. Finally, the treatment effects on speech anxiety are presented.

Descriptive statistics for the implementation fidelity variables are displayed in Table 2. Descriptive statistics for the dependent and independent variables are reported in Table 4. The correlations between the dependent variables at pretest and posttest are reported in Table 5. The treatment effects are displayed in Tables 6–8.

3.1. Implementation fidelity

High implementation fidelity was expected because the instructors had attended the workshop and had been provided with materials and support. Implementation fidelity was assessed along the two dimensions adherence and quality of delivery using instructors' self-reports. For adherence, the percentage of exercises conducted by each instructor was calculated and ranged from 73% to 95% ($M = 85.33$, $SD = 8.59$; see Table 2). No remarkable omissions were observed except for one course instructor (A) who conducted only 37% of the exercises in Course Unit 8 due to technical problems. The mean reported quality of delivery across all instructors and course units was 2.63 ($SD = 0.55$). Across the instructors, the mean reported quality (i.e., how well the implementation of the exercise went) ranged from 2.32 ($SD = 0.56$; Instructor C) to 2.84 ($SD = 0.36$; Instructor D). Thus, for the exercises they conducted, the instructors rated the quality of delivery as medium to high.

3.2. Treatment effects on public speaking skills

For Hypothesis 1, we expected that the training would positively affect the students' ability to use public speaking skills. Concerning the

[4] Although we intended to do so, we did not account for the nesting of the data because we encountered errors when trying to run corresponding analyses due to the relatively small sample size and the relatively small numbers of clusters.

[5] When interpreting the results of this study, it has to be noted that for general fluid abilities, the reliability coefficient was only .57, probably because the assessment took place in a group setting. Note that the low reliability may have resulted in an underestimation of the effects.

Table 5
Correlations between the dependent variables at pretest (below diagonal) and posttest (above diagonal).

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	Posture		.17	.36*	.01	.15	.12	−.12	.00	.17	.10	.25	.04	−.09
(2)	Gestures	.33*		.22	−.14	.21	.36*	.36*	.31*	.48*	.35*	.47*	−.20	−.09
(3)	Eye contact	−.16	.08		.25	.43*	.27*	.15	.13	.20	.37*	.54*	.08	.05
(4)	Speech respiration	.26*	−.04	−.13		.18	.04	−.03	−.15	−.05	−.04	.16	.14	.12
(5)	Breaks	.36*	.35*	.06	−.02		.67*	.38*	.04	.22	.24	.41*	−.01	−.06
(6)	Speech fluency	.34*	.40*	−.06	−.11	.82*		.53*	.21	.22	.29*	.61*	−.21	−.23
(7)	Length of speech	.22	.31*	.26*	−.15	.40*	.50*		.26	.22	.25	.58*	.08	.05
(8)	Length of introduction	.01	.29*	.24	−.26*	.26*	.26*	.37*		.53*	.57*	.39*	−.17	−.10
(9)	Length of conclusion	−.06	.35*	.12	−.19	.16	.11	.27*	.34*		.33*	.40*	−.08	−.16
(10)	Reference to listener	.02	.23	.30*	−.14	.32*	.33*	.22	.48*	.17		.57*	−.14	−.07
(11)	Global performance	.37*	.46*	.41*	−.05	.52*	.51*	.63*	.55*	.37*	.52*		−.12	−.08
(12)	Speech anxiety cognitive	−.20	−.25*	−.24	.02	.08	.00	−.15	.03	−.12	−.10	−.26*		.73*
(13)	Speech anxiety bodily	−.07	−.04	−.24	.09	.04	.00	−.09	.16	−.06	−.05	−.10	.62*	

Note. Variables 1 to 11 were assessed with external ratings of the videotaped presentations. Variables 12 and 13 were self-reported.
* $p < .05$.

Table 6
Treatment effects on global performance and organizational skills (posttest).

	Global performance			Organization												
				Length of speech			Length of introduction			Length of conclusion			Reference to listener			
	B	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*
Treatment	0.56	0.23	.008	0.61	0.23	.004	0.81	0.26	.001	0.89	0.24	< .001	0.64	0.27	.010
Pretest score	0.56	0.20	.006	0.12	0.16	.461	0.29	0.13	.031	0.03	0.14	.839	0.51	0.14	< .001
Treatment x Pretest score	−0.19	0.27	.482	0.10	0.22	.631	−0.09	0.29	.765	0.33	0.26	.200	−0.52	0.22	.017
Gender	−0.11	0.24	.658	0.62	0.19	.001	0.53	0.27	.053	0.53	0.24	.027	0.11	0.28	.707
General fluid abilities	−0.05	0.15	.753	−0.01	0.12	.965	−0.04	0.18	.806	0.23	0.11	.037	−0.18	0.15	.246
Verbal intelligence	0.32	0.11	.003	0.24	0.16	.138	0.17	0.12	.137	−0.01	0.11	.966	0.22	0.12	.074
Explained variance (R^2)	.39			.27			.26			.34			.27		

Note. All continuous variables were standardized prior to the analysis. Treatment was dummy-coded (1 = intervention, 0 = control group). Gender was dummy-coded (1 = boys, 0 = girls). One-tailed significance levels are reported for the treatment.

Table 7
Treatment effects on nonverbal public speaking skills (posttest).

	Nonverbal–visual									Nonverbal–auditory								
	Posture			Gestures			Eye contact			Speech respiration			Breaks			Speech fluency		
	B	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*	*B*	*(SE)*	*p*
Treatment	0.50	0.31	.056	0.24	0.25	.170	0.26	0.29	.185	0.10	0.22	.324	0.03	0.29	.457	0.42	0.27	.064
Pretest score	0.14	0.23	.525	0.65	0.20	.001	0.35	0.21	.103	0.79	0.22	< .001	0.18	0.17	.302	0.06	0.19	.749
Treatment x Pretest score	0.44	0.30	.147	−0.46	0.25	.068	0.08	0.31	.792	−0.46	0.27	.087	−0.53	0.30	.079	0.23	0.30	.435
Gender	0.37	0.31	.243	0.08	0.33	.801	0.13	0.29	.668	−0.37	0.25	.144	−0.31	0.33	.338	0.27	0.29	.353
General fluid abilities	0.01	0.18	.948	0.21	0.13	.109	−0.05	0.22	.828	0.30	0.16	.052	0.21	0.13	.117	−0.09	0.14	.527
Verbal intelligence	0.04	0.11	.723	0.06	0.09	.532	0.12	0.10	.249	−0.28	0.14	.048	0.33	0.09	.001	0.39	0.10	< .001
Explained variance (R^2)	.23			.28			.18			.36			.22			.23		

Note. All continuous variables were standardized prior to the analysis. Treatment was dummy-coded (1 = intervention, 0 = control group). Gender was dummy-coded (1 = boys, 0 = girls). One-tailed significance levels are reported for the treatment.

appropriateness of *global performance*, in support of Hypothesis 1, we found a statistically significant positive effect of the training ($B = 0.56$, $p = .008$; Table 6). For the dimensions of public speaking competence, we found the following results: For the organization dimension, significant positive effects of the treatment were found for all four public speaking skills: *length of speech*: $B = 0.61$, $p = .004$; *length of introduction*: $B = 0.81$, $p = .001$; *length of conclusion*: $B = 0.89$, $p < .001$; and *reference to listener*: $B = 0.64$, $p = .010$ (Table 6). For the nonverbal–visual dimension, no significant positive effects of the treatment were found for any of the three public speaking skills *posture*, *gestures*, or *eye contact* (all *p*-values > .05; Table 7). For the nonverbal–auditory dimension, no positive treatment effects were found for any of the three skills *speech respiration*, *breaks,* or *speech fluency* (all *p*-values > .05; Table 7). Overall, children who participated in the speech training program, compared with the control group, showed a more appropriate *global performance* and were better able to adapt their organizational skills (*length of speech*, *length of introduction*, *length of conclusion*, *reference to listener*) to the specific context. Concerning the appropriateness of nonverbal public speaking skills (i.e., visual and auditory), the children who participated in the speech training program did not show more appropriate behavior than those in the control group.

Concerning the differential effects that depended on the initial manifestation of the dependent variables, no significant interaction between the treatment and pretest score was found except for *reference*

Table 8
Treatment effects on the reduction of speech anxiety (posttest).

	Speech anxiety					
	Cognitive			Bodily		
	B	*(SE)*	*p*	*B*	*(SE)*	*p*
Treatment	−0.68	0.21	< .001	−0.48	0.22	.017
Pretest score	0.81	0.08	< .001	0.87	0.14	< .001
Treatment x Pretest score	−0.05	0.16	.739	−0.23	0.20	.243
Gender	−0.72	0.25	.004	−0.53	0.21	.014
General fluid abilities	0.24	0.12	.034	0.14	0.07	.054
Verbal intelligence	0.10	0.08	.204	−0.07	0.08	.385
Explained variance (R^2)	.60			.58		

Note. Speech Anxiety = Self-reported frequency of speech anxiety symptoms. All continuous variables were standardized prior to the analysis. Treatment was dummy-coded (1 = intervention, 0 = control group). Gender was dummy-coded (1 = boys, 0 = girls). One-tailed significance levels are reported for the treatment.

to listener ($B = -0.52$, $p = .017$; Table 6). Thus, the children who used references to the listener in a less appropriate fashion at pretest benefitted more from the training.

3.3. Treatment effects on speech anxiety

For Hypothesis 2, we expected that the children in the intervention group would report less speech anxiety after participating in the program than those in the control group. In line with this hypothesis, significant negative effects of the treatment were found for speech anxiety–cognitive ($B = -0.68$, $p < .001$) and speech anxiety–bodily ($B = -0.48$, $p = .017$; Table 8). Children who participated in the program reported a lower intensity of speech anxiety symptoms compared with the control group. With regard to the differential effects that depended on the initial manifestation of speech anxiety, no significant interaction between the treatment and pretest score was found.

4. Discussion

This study assessed the effectiveness of a speech training program after it was implemented by nonresearch course instructors in the field. Given that moving from controlled to real-world instructional settings often results in smaller treatment effects (e.g., owing to low implementation fidelity), the course instructors were thoroughly prepared and supported before they administered the training. Treatment effects were assessed by applying a randomized wait-list control group design with pretest and posttest measures. The target outcomes were public speaking skills assessed by external ratings of videotaped presentations and self-reported speech anxiety.

4.1. Implementation fidelity

In order to analyze whether the effects of the training program on specific outcomes were due to the intervention as intended, it was necessary to demonstrate implementation fidelity (Munter et al., 2014). A one-to-one match between intended and implemented interventions is often not realistic. In the review by Durlak and DuPre (2008), no study reported an implementation of 100% for all instructors. Improvement in the outcome had often been observed with implementation levels of around 60%, and only a few studies achieved a level above 80%. Thus, the mean adherence of 85.33% ($SD = 8.59$) in the current study can be considered very high. Furthermore, medium to high quality of delivery was reported. The results indicate that across the units and instructors, the implementation of the exercises worked quite well.

The high level of implementation fidelity in this study indicates that the implementation support system we provided (workshop, manual, all teaching materials) was sufficiently detailed and adequately prepared so that it allowed the instructors to implement the speech training as intended. For example, because the instructors were given all of the teaching materials, they were able to implement each exercise without high temporal costs for preparation. In addition, the materials seemed to be detailed enough to allow for an effective implementation. After demonstrating high fidelity in the implementation of the speech training, we could then analyze and interpret the effects of the training.

4.2. From the efficacy study to the effectiveness study

4.2.1. Differences in treatment effects on public speaking skills

Because moving from controlled to instructional settings with different nonresearch instructors often results in smaller treatment effects, we compare the results of the previous efficacy study (Herbein et al., 2018) with the results of the present effectiveness study. However, as the studies used different control groups (treated and nontreated), we discuss the differences in the effect sizes only on a descriptive level. In the previous efficacy study (with only one course instructor and a treated control group), positive treatment effects were found for public speaking skills on the following dimensions: *global performance*, organizational public speaking skills (*length of speech*, *length of conclusion*, *reference to listener*), nonverbal–visual skills (*posture*, *gestures*, *eye contact*), and nonverbal–auditory skills (*breaks*, *speech fluency*). In line with the efficacy study, we herein found positive effects that were medium to large in size for *global performance* and organizational public speaking skills (*length of speech*, *length of conclusion*, *reference to listener*). Furthermore, in comparison with the efficacy study, there were no significant effects of the training on the nonverbal–visual skills *posture, gestures,* and *eye contact* and the nonverbal–auditory skills *speech respiration, breaks*, and *speech fluency* in the effectiveness study. For an overview of the effect sizes in both studies, see Table 9.

The varying effects between the efficacy and effectiveness studies might be explained by differences between the studies concerning the characteristics of the course instructors and the control groups that were employed (Greene, 2015). Thus, one step may be to take a closer look at the instructors' professional competence, including, for example, content knowledge, pedagogical content knowledge, and motivational aspects such as communication apprehension. More specifically, in order to teach public speaking, instructors need to be able to speak competently in front of the class (Book, 1989) because they are role models not only within the practical exercises but also across the entire course (Staton & Tomlinson, 2001). Having a good role model is important when acquiring new skills. In addition, teaching public speaking requires knowledge about public speaking dimensions, their application, and their impact when used in a certain way in a specific context (Wilson, 1997). Furthermore, an awareness of the range of

Table 9
Overview of the treatment effects for the efficacy and effectiveness studies.

	Efficacy study	Effectiveness study
Global performance	0.62	0.56
Organization		
Length of speech	0.54	0.61
Length of conclusion	0.59	0.89
Reference to listener	0.54	0.64
Nonverbal–visual		
Posture	0.69	0.50
Gestures	0.77	0.24
Eye contact	0.45	0.26
Nonverbal–auditory		
Speech respiration	−0.70	0.10
Breaks	0.64	0.03
Speech fluency	0.54	0.42
Speech anxiety	−0.15	−0.48 and −0.68

Note. The treatment effects are reported as Cohen's *d*.

possible skills is required to be able to provide detailed (video) feedback. Thus, future studies need to consider specific characteristics of course instructors (e.g., their public speaking competence) when analyzing treatment effects. Therefore, not only should instructors' self-reported communication competence be measured, but external ratings of their public speaking skills should also be taken into account.

4.2.2. Differences in treatment effects on speech anxiety

The effects on speech anxiety differed decisively between the efficacy and the effectiveness study in favor of the effectiveness study. There were statistically significant treatment effects on both scales of speech anxiety (i.e., cognitive and bodily) in the effectiveness study. By contrast, no significant reduction in speech anxiety was found in the efficacy study (see Table 9).

The two studies differed in the characteristics of the control group: treated versus wait-listed. In the previous efficacy study, the control group was treated because they received a course on understanding science. This course also included exercises in which the children had to present the results of an experiment in front of the other students. Thus, the confidence of the children who participated in the treated control group may have increased with respect to speaking in front of others as well. This might explain why there were no differences in speech anxiety between groups in the efficacy study, but the intervention had a significant effect on both speech anxiety scales in the current effectiveness study (which used a wait-list control group design).

4.2.3. Summary and classification of the treatment effects

Overall, positive treatment effects were found on most of the public speaking skills across both studies. Thus, the intervention fosters public speaking skills not only under standardized conditions (only one instructor) but also when administered in the real world (different course instructors). Children participating in the speech training program, compared with a control group, showed a more appropriate *global performance* and were able to more adequately adapt their organizational skills (*length of speech*, *length of conclusion*, *reference to listener*) to the specific context. Furthermore, they also showed substantive improvement in the appropriateness of their *posture* and *eye contact* (nonverbal–visual) and *speech fluency* (nonverbal–auditory) compared with the control group.

However, the slightly less pronounced effects for nonverbal–visual as well as nonverbal–auditory public speaking skills in the effectiveness study are not surprising when the findings of previous studies are taken into consideration. For example, a similar pattern of treatment effects on public speaking skills was observed in higher education. De Grez, Valcke, and Roozen (2009a, 2009b), for instance, found statistically significant positive effects on organizational skills (introduction, structure, conclusion, contact audience), but they found no treatment effects for nonverbal skills (eye contact, vocal delivery). The authors discussed two possible reasons for their findings: Either the training methods that had been employed were inefficient or fostering nonverbal public speaking skills is very difficult in general. Additional studies on nonverbal public speaking skills thus appear to be necessary, especially when considering that, under controlled conditions (efficacy study), Herbein et al. (2018) observed significant medium-sized effects on nonverbal–visual and nonverbal–auditory public speaking skills.

4.3. Limitations and strengths

When an intervention is put into practice by different nonresearch course instructors, the developers need to consider how to help the instructors implement the intervention as intended. High implementation fidelity can be supported, for example, through coaching, performance feedback, prompts, scripts, or workshops. For the present target group of course instructors, a scripted manual was used in addition to a 1-day workshop because this procedure seemed to be a practicable and acceptable way to achieve high implementation fidelity. However, a script has advantages as well as disadvantages. A possible negative effect of a detailed script is that purposeful adaptations might not be made (e.g., with regard to language, examples, metaphors used, or feasibility; Souvignier & Mokhlesgerami, 2006). In addition, creativity and teaching quality might be negatively affected (see Carl, 2014; Timberlake, Thomas, & Barrett, 2017). However, with regard to the target group of course instructors from the HCAP, the script offered several advantages. First, for most of the instructors, the course topic was quite new. The detailed script, which included theoretical background information as well as the contents and teaching activities for each course unit, reduced the effort they needed to put forth to prepare the course. Second, some of the course instructors at the HCAP who participated in this study did not have a formal pedagogical qualification. Thus, the quality of their teaching was most likely supported by the clear structure and instructional goals for each unit that were provided.

Several approaches are available for measuring the dimensions of implementation fidelity (e.g., instructors' behavior, participants' behavior; see Humphrey et al., 2016), and measurements may vary in accordance with the characteristics of the intervention (Abry et al., 2015). A combination of different fidelity measures is often suggested (Munter et al., 2014). In order to consider structural and process variables with respect to implementation fidelity, and on the basis of the approaches used to assess implementation fidelity in different psychological interventions in an educational setting (see Abry et al., 2015; Odom et al., 2010), we combined the assessments of adherence and quality of delivery in the current study. By asking instructors to rate whether each exercise was conducted or not and how well the implementation worked, we were able to provide a very detailed analysis of the structure of the intervention that was implemented and identify any elements that were (systematically) left out when the intervention was conducted. The assessment of adherence also enabled us to examine whether the structure was feasible in the specific teaching context (e.g., in terms of time-related and organizational factors). However, the assessment was based solely on the course instructors' self-reports, potentially resulting in some limitations. Self-reports might be biased, and this might result, for example, in an overestimation of adherence and quality because of social desirability. Furthermore, the accuracy of the reported data may have been reduced because the estimations of fidelity were retrospective (James Bell Associates, 2009). In addition to self-reports, observational measures (i.e., ratings made by someone observing the course or ratings of videotapes) can be used to assess implementation fidelity (Chambless & Hollon, 2012; Walton, Spector, Tombor, & Michie, 2017). However, the use of external raters in class or the generation of videotapes might also increase bias because course instructors might adhere to the manual more intensely when observed or videotaped. These approaches might also reduce instructors' acceptance of and commitment to the intervention and the study. Furthermore, the procedures of observing or videotaping and rating each lesson come with higher financial and temporal costs than those associated with self-reports (James Bell Associates, 2009). On the basis of these pros and cons, we decided to use self-reports in the present study. In order to mitigate potential limitations due to social desirability bias, we informed the instructors that we planned to use their feedback on the course to further develop and adapt the speech training, and thus, we asked them to help us by answering the questionnaire honestly. Giving the instructors another practical reason to answer the fidelity questionnaire honestly might have reduced socially desirable responding. However, in future studies, external ratings of the lessons (in-class observers or videotape ratings) should also be used (Chambless & Hollon, 2012; Walton et al., 2017).

The effectiveness study provided one step in the process of putting the speech training into real-world practice and achieving the scaling-up of the training. To design the effectiveness study with adequate power, we had to refer to the results of the efficacy study to obtain an estimate of the expected effect sizes because—to the best of our

knowledge—there are no other comparable studies on the promotion of public speaking skills in elementary school children. However, because only small numbers of course instructors and children signed up, the study had a somewhat smaller sample size than originally planned. The treatment effects for some dimensions of public speaking skills were so small that—to keep the treatment effects steady—a larger sample size would have been necessary to achieve a statistically significant effect. But although the sample size of the participating children did not exceed the sample size from the previous efficacy study, it could be judged as reasonable.

The speech training program was designed for third and fourth graders nominated for an extracurricular enrichment program for gifted students. Therefore, the findings should not automatically be generalized to the whole population of third and fourth graders. However, students' nomination for the enrichment program was based on teachers' recommendation, and not on intelligence test scores. Correspondingly, the children participating in our training program still varied in their verbal intelligence ($M = 110.90$, $SD = 15.05$) and general fluid abilities ($M = 10.02$, $SD = 2.53$), with a mean clearly below recommended thresholds for giftedness based on intelligence testing. Further, initially concentrating on a specific target group is the approach that is recommended in the literature for implementing and examining the effectiveness of an intervention, before broadening the target group and the context in which the training is implemented in a potential next step (see Gottfredson et al., 2015).

Because only a small number of instructors participated in the current study, systematic influences of their characteristics on the treatment effects could not be tested for significance. Given the fact that cluster randomization was used at three local HCAP sites, only five out of eight course instructors offered the intervention to both groups, and thus, differences in the treatment effects could be explored only for these five instructors. The first initial analysis showed variation in the treatment effects (see Appendix C). These results highlight the necessity of looking at the characteristics of the instructors in more detail and of assessing whether they led to differential effects on students' outcomes. However, any variability that might depend on the course instructor was confounded by the local site and—given the small size of the sample of instructors—provided only a first impression of the potential variation in effects across instructors. Here, sound studies that can clearly investigate the effects of this variable offer a promising next step. Nevertheless, teaching other course instructors to conduct the training is the only way to estimate whether positive treatment effects will also occur in a natural setting.

4.4. Conclusion

Across all stages of intervention development and evaluation and by applying a design-based research approach, researchers aim to effectively put empirical findings and theory into practice while at the same time creating a starting point for deriving new theories and design principles (Anderson & Shattuck, 2012; Greene, 2015). Despite this approach, whether a conceptualized intervention can be effectively applied in the real world has not been examined very often (see Fixsen et al., 2013), and no (see Star et al., 2015) or reduced effects on the outcomes have often been found (e.g., Durlak & DuPre, 2008; Hulleman & Cordray, 2009). These findings highlight the importance of conducting systematic investigations of effectiveness to provide reasons for applying an intervention in the field (Humphrey et al., 2016) and to refine theory (Greene, 2015; Hughes, 2015).

The current effectiveness study integrated the effects of a previously conducted efficacy study with an implementation approach that included a workshop for instructors and provided all of the course materials. The findings revealed that the intervention was successful when put into practice, even though some treatment effects were descriptively smaller in size compared to the efficacy study. Thus, the results suggest that the core components derived from research on promoting public speaking competence in higher education can successfully be transferred to the context of communication education in elementary schools. Future studies might take these results up, investigating which of the training's core components are most effective for promoting elementary-school-aged children's public speaking skills. In addition, by teaching and supporting the course instructors—in particular, by offering a supplemental workshop and assessing the fidelity of the implementation—we provided a procedure that future studies can apply to obtain further insights into the feasibility of the training program. This will enable further revisions of the structure and processes of the intervention in order to improve its practicability in real-world settings as well as its effectiveness. Crucial decisions about a scaling-up of the intervention can then be made on the basis of its level of effectiveness and an evaluation of its practicability.

Acknowledgments

This research project was funded by a grant from the Hector Foundation II to the last author. When this study was conducted, Evelin Herbein was a doctoral student at the LEAD Graduate School & Research Network [GSC 1028], which is funded by the Excellence Initiative of the German federal and state governments.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cedpsych.2018.09.003.

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.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25. https://doi.org/10.3102/0013189X11428813.

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.

Blase, K. A., van Dyke, M., Fixsen, D. L., & Bailey, F. W. (2012). Key concepts, themes, and evidence for practitioners in educational psychology. In B. Kelly, & D. F. Perkins (Eds.). *Handbook of implementation science for psychology in education* (pp. 13–34). Cambridge, New York, NY: Cambridge University Press.

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

Book, C. L. (1989). Communication education: Pedagogical content knowledge needed. *Communication Education, 38*(4), 315–321. https://doi.org/10.1080/03634528909378770.

Bywater, T. (2012). Developing rigorous programme evaluation. In B. Kelly, & D. F. Perkins (Eds.). *Handbook of implementation science for psychology in education* (pp. 37–53). Cambridge, New York, NY: Cambridge University Press.

Carl, N. M. (2014). Reacting to the script: Teach for America teachers' experiences with scripted curricula. *Teacher Education Quarterly, 41*(2), 29–50. http://www.jstor.org/stable/teaceducquar.41.2.29.

Chambers, D. W. (1983). Stereotypic images of the scientist: The Draw-A-Scientist Test. *Science Education, 67*(2), 255–265. https://doi.org/10.1002/sce.3730670213.

Chambless, D. L., & Hollon, S. D. (2012). Treatment validity for intervention studies. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.). *APA handbook of research methods in psychology. Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological* (pp. 529–552). Washington: American Psychological Association. https://doi.org/10.1037/13620-028.

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. https://doi.org/10.1037/1040-3590.6.4.284.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159. https://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, N.J.: L. Erlbaum Associates.

Common Core State Standards Initiative. (2010). Common core state standards for English language arts & literacy in history/social studies, science, and technical subjects. Washington D.C.

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. https://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. https://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. https://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. https://doi.org/10.1007/BF03174762.
Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0.
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.
Enders, C. K. (2010). *Applied missing data analysis. Methodology in the social sciences.* New York, NY: Guilford Press.
Fehm, L., & Hille, C. (2005). Bühnenangst bei Musikstudierenden [Stage fright amongst students of music]. *Verhaltenstherapie und Verhaltensmedizin, 2*, 199–212.
Fixsen, D. L., Blase, K., Metz, A., & van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children, 79*(2), 213–230. https://doi.org/10.1177/001440291307900206.
Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). Implementation research: A synthesis of the literature. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
Gottfredson, D. C., Cook, T. D., Gardner, F. E. M., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (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. https://doi.org/10.1007/s11121-015-0555-x.
Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(4), 879–896. https://doi.org/10.1037/a0029185.
Greene, J. A. (2015). Serious challenges require serious scholarship: Integrating implementation science into the scholarly discourse. *Contemporary Educational Psychology, 40*, 112–120. https://doi.org/10.1016/j.cedpsych.2014.10.007.
Herbein, E., Golle, J., Tibus, M., Schiefer, J., Trautwein, U., & Zettler, I. (2018). Fostering elementary school children's public speaking skills: A randomized controlled trial. *Learning and Instruction, 55*, 158–168. https://doi.org/10.1016/j.learninstruc.2017.10.008.
Hughes, J. N. (2015). Integrating theory and empirical science in school psychology: Progress and remaining challenges. *School Psychology Review, 44*(3), 262–270.
Hulleman, C. S., & Cordray, D. S. (2009). Moving from the lab to the field: The role of fidelity and achieved relative intervention strength. *Journal of Research on Educational Effectiveness, 2*(1), 88–110. https://doi.org/10.1080/19345740802539325.
Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. (2016). Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature. Retrieved from Education Endowment Foundation website: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Review_Final.pdf.
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. https://doi.org/10.1080/03634523.2014.926016.
James Bell Associates (2009). Evaluation brief: Measuring implementation fidelity. Arlington, VA. Retrieved from http://www.jbassoc.com/ReportsPublications/Evaluation%20Brief%20-%20Measuring%20Implementation%20Fidelity_Octob%E2%80%A6.pdf.
Kultusministerkonferenz (2005). Beschlüsse der Kultusministerkonferenz: Bildungsstandards im Fach Deutsch für den Primarbereich (Jahrgangsstufe 4) [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)]. München.
Lee, O., Quinn, H., & Valdes, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for Common Core State Standards for English language arts and mathematics. *Educational Researcher, 42*(4), 223–233. https://doi.org/10.3102/0013189X13480524.
Lendrum, A., & Wigelsworth, M. (2013). The evaluation of school-based social and emotional learning interventions: Current issues and future directions. *The Psychology of Education Review, 37*(2), 70–76.
McCroskey, J. C., Beatty, M. J., Kearney, P., & Plax, T. G. (2009). The content validity of the PRCA-24 as a measure of communication apprehension across communication contexts. *Communication Quarterly, 33*(3), 165–173. https://doi.org/10.1080/01463378509369595.
McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports, 5*(2), 108–113. https://doi.org/10.1080/08824098809359810.
Min, S. J. (2014). The basic course isn't really all that basic. The Official Journal of the Georgia Communication Association, Inc. (83), 48–52.
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., Spitzberg, B. H., & Barge, J. K. (2013). *Communication: Motivation, knowledge, skills* (3rd ed.). New York, NY: Peter Lang Publishing Inc.
Morreale, S. P., Valenzano, J. M., & Bauer, J. A. (2016). Why communication education is important: A third study on the centrality of the discipline's content and pedagogy. *Communication Education, 3*(41), 1–21. https://doi.org/10.1080/03634523.2016.1265136.
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. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research, 39*(4), 374–396. https://doi.org/10.1007/s11414-012-9295-x.
O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research, 78*(1), 33–84. https://doi.org/10.3102/0034654307313793.
Odom, S. L., Fleming, K., Diamond, K., Lieber, J., Hanson, M., Butera, G., ... Marquis, J. (2010). Examining different forms of implementation and in early childhood curriculum research. *Early Childhood Research Quarterly, 25*(3), 314–328. https://doi.org/10.1016/j.ecresq.2010.03.001.
Rimm-Kaufman, S. E., Larsen, R. A. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., ... DeCoster, J. (2014). Efficacy of the responsive classroom approach: Results from a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal, 51*(3), 567–603. https://doi.org/10.3102/0002831214523821.
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. https://doi.org/10.1037/edu0000085.
Rubin, R. B., & Morreale, S. P. (1996). Setting expectations for speech communication and listening. *New Directions for Higher Education, 1996*(96), 19–29. https://doi.org/10.1002/he.36919969604.
Schoenwald, S. K., Garland, A. F., Chapman, J. E., Frazier, S. L., Sheidow, A. J., & Southam-Gerow, M. A. (2011). Toward the effective and efficient measurement of implementation fidelity. *Administration and Policy in Mental Health, 38*(1), 32–43. https://doi.org/10.1007/s10488-010-0321-0.
Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence, 59*, 84–95. https://doi.org/10.1016/j.intell.2016.08.003.
Souvignier, E., & Mokhlesgerami, J. (2006). Using self-regulation as a framework for implementing strategy instruction to foster reading comprehension. *Learning and Instruction, 16*(1), 57–71. https://doi.org/10.1016/j.learninstruc.2005.12.006.
Star, J. R., Pollack, C., Durkin, K., Rittle-Johnson, B., Lynch, K., Newton, K., & Gogolen, C. (2015). Learning from comparison in algebra. *Contemporary Educational Psychology, 40*, 41–54. https://doi.org/10.1016/j.cedpsych.2014.05.005.
Staton, A. Q., & Tomlinson, S. D. (2001). Communication education outreach in elementary school classrooms. *Southern Communication Journal, 66*(3), 211–223. https://doi.org/10.1080/10417940109373200.
Timberlake, M. T., Thomas, A. B., & Barrett, B. (2017). The allure of simplicity: Scripted curricula and equity. *Teaching and Teacher Education, 67*, 46–52. https://doi.org/10.1016/j.tate.2017.05.017.
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. https://doi.org/10.1016/j.edurev.2015.02.002.
Walton, H., Spector, A., Tombor, I., & Michie, S. (2017). Measures of fidelity of delivery of, and engagement with, complex, face-to-face health behaviour change interventions: A systematic review of measure quality. *British Journal of Health Psychology, 22*(4), 872–903. https://doi.org/10.1111/bjhp.12260.
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.
Wilson, J. A. (1997). *A program to develop the listening and speaking skills of children in a first grade classroom.* University of Virginia.