

RESEARCH

Open Access



# Methodological challenges of evaluating the effects of an early language education programme in Germany

Yvonne Anders<sup>1\*</sup>, Hans-Günther Rossbach<sup>2</sup> and Wolfgang Tietze<sup>3</sup>

\*Correspondence:

yvonne.anders@fu-berlin.de

<sup>1</sup> Early Childhood Education,  
Department of Education  
and Psychology, Freie  
Universität Berlin,,  
Habelschwerdter Allee 94,  
14195 Berlin, Germany  
Full list of author information  
is available at the end of the  
article

## Abstract

Many early language support programmes have been implemented in ECEC settings over the last years in Germany. Most of them are targeted programmes which have not shown to have longer term beneficial effects. The concept of early language education embedded into daily preschool routines is a child-oriented approach for all children and all age groups. The national programme “core daycare centres language & integration” aimed at promoting early language education embedded into daily routines at 4000 participating centres. In this paper, the longitudinal evaluation study of the programme is described, and the specific methodological challenges and potentials of the evaluation are discussed. The lack of instruments in ECEC research, the nature of pedagogy in ECEC, the autonomy given to the settings and professionals, the lack of control and the interpretation of effect sizes are highlighted as the main challenges. At the same time, the urgent need of such evaluations and the benefits for policy-makers, research, practitioners and the public are explored.

**Keywords:** Implementation of policy, Evaluation, Early language education, Methods

## Background

It is widely acknowledged that the development of language skills plays a key role in child development (e.g. Cohen 2006; Stevenson et al. 1985; Tomblin et al. 2000; Weinert 2006). International school achievement studies documented that students in Germany lagged behind compared to students of other countries, especially in their reading skills (Baumert and Schümer 2001; Bos et al. 2003). Furthermore, the influence of the socio-economic background on students’ skills was found to be stronger in Germany compared to other countries. In comparison, the average skills of German students have improved over the last years, but strong disadvantages of children who grow up in educationally disadvantaged families are still evident (Müller and Ehmke 2013).

These findings have stimulated a debate about prevention and intervention strategies. One result was a rising attention towards potential benefits of high-quality preschool education programmes. International studies had accumulated evidence that high-quality preschool education can foster children’s early cognitive and academic development (e.g. Anders 2013; Barnett and Hustedt 2005; Camilli et al. 2010; Reynolds et al. 2010; Sylva et al. 2011). To enhance the capacity of preschools in promoting children’s

language skills in Germany, various targeted language support approaches and projects have been implemented, mainly aiming at children age three and older. However, the few existing evaluation studies point to disappointing results; most regional model projects did not yield any positive effects on children's development (Gasteiger-Klicpera et al. 2010; Kammermeyer et al. 2011; Sachse et al. 2012; Wolf et al. 2011). Programmes aiming at children's phonological awareness have proven to be effective, but the effects seem to fade out quickly. Furthermore, no positive transfer on other language-related skills, the development of grammatical knowledge or reading abilities could be established (Wolf et al. 2016). To overcome the weaknesses of existing approaches, the Federal Ministry for Family Affairs, Senior Citizens, Women and Youth (Bundesministerium für Familie, Senioren, Frauen und Jugend—BMFSFJ) launched the initiative “Early Chances” and the programme “core daycare centres language & integration (Schwerpunkt-Kitas: Sprache and Integration)” in 2011 which was set up to raise the quality of early language education in the centres.

In this paper, we describe the evaluation study of the programme and discuss the specific methodological challenges and potentials of the evaluation for different programme partners. To start, we give some important background information on the early childhood education and care (ECEC) system in Germany, and then we describe the concepts of early language education embedded into daily routines and language-specific ECEC quality. Against this background, the programme “core daycare centres language & integration” and the design of the evaluation study are presented. This will be followed up by the discussion of challenges and possibilities of evaluating ECEC policy; a conclusion sums up the implications. The presentation of the study and the discussion of the challenges and possibilities of evaluations in this area are the core topics of this paper. It will use examples of findings, but it will not give a comprehensive overview of the main findings of the evaluation. These will be presented and discussed in the final report and other forthcoming publications.

### **The German ECEC System**

In Germany, participation in preschool programmes or daycare is voluntary, compulsory schooling starts at the age of 6 years. Some federal states allow enrolling children into primary school at the age of 5 years. In 2015, 94.9% of all children between three and 5 years and 32.9% of all children up to 2 years attended non-familial ECEC (Statistische Ämter des Bundes und der Länder 2016). Although the attendance rates of children younger than three years are still relatively low compared to other countries, this number has been consistently rising over the last years. In 2013, the legal entitlement for a half-day place in daycare for all children aged 1 year and older came into effect. In Germany, the ECEC services are not located within the public education system, but within the child and youth welfare sector. Traditionally, a lot of autonomy is given to the settings and pedagogical staff, especially with regard to the interpretation and implementation of pedagogical approaches. However, between 2003 and 2007, official curricular guidelines were introduced in all 16 federal states of Germany. All curricular frameworks define learning areas, but no learning goals. The frameworks differ greatly between federal states, and varying implementation strategies have been developed. They provide evidence of a holistic perspective on ECEC (Prött and Preissing 2006). Most of the

ECEC settings in Germany work according to the situation-oriented approach (Oertel 1984). This approach is a child-centred approach which emphasizes that any learning in early childhood takes place in social situations, mainly based on children's play (Preissing 2007). Thus, daily experiences and interests of children form the basis for any pedagogical strategy. The pedagogic principles stress the objective of promoting children to become an active and responsible member of society. Preschool groups in Germany are traditionally mixed-age groups sometimes covering the age span from one year of age until school enrolment.

In most of the federal states, the childcare providers are responsible for ensuring the quality of provision, and they are free to choose how this is achieved. A number of quality management systems are run, including internal and external quality management systems. The vast majority of early childhood professionals have completed a 3-year post-secondary vocational training programme. Since 2003, a fast-growing number of higher education degree-level courses in early childhood pedagogy have become available. But the number of professionals with college or university degree in the settings is still low (2014: 5.3%, Bock-Famulla et al. 2015) compared to other countries.

### **Early language education embedded into daily preschool routines**

Many early language support programmes that have been implemented in ECEC settings over the last years in Germany are targeted programmes for children approaching enrolment into primary school or children with an immigration background or children with special need of language support. They take place in small groups, in addition, and outside the daily routines. Furthermore, they use mainly teacher-directed pedagogy (Lisker 2011). Despite of the fact that these programmes have not proven to be effective (see above), the teacher-directed nature of these programmes seems to be unsuitable for the German ECEC context. Additionally, these programmes ask children to acquire skills in unnatural situations, so that the ability to transfer and train these skills in other situations is difficult. Alternatives have been discussed and the concept of early language education embedded into daily preschool routines has emerged (Fried 2013; Jampert et al. 2011). This concept is a child-oriented approach for all children and all age groups. The promotion of language skills takes place systematically as part of the daily routines, e.g. when changing diapers, during undressing and dressing, at the lunch table or during free play time. Children's questions and interests are taken up actively and the communication with the child takes place either with words or with gestures. Intense dialogues with the children are a key component of this way of stimulating children's development of language skills. The expected interactions are quite similar to the concept of sustained shared thinking of Siraj-Blatchford and colleagues (2002). According to this concept, in high-quality interactions, adults are genuinely interested in what the child is doing, and adults are listening and extending children's thoughts and knowledge. Further strategies, especially to promote children's acquisition of new knowledge, are open-ended questions or comments, giving the child time to respond, and using knowledge of the child to extend the interaction. Child-centred approaches require the ability to react spontaneously to children's interests and ideas, whereas teacher-directed approaches have clearly defined, specific aims and strategies. Thus, teacher-directed approaches may be less challenging for the professionals as they are easier to apply. Studies confirm that

early childhood professionals have a better understanding of their role in more predictable, teacher-led activities (Sproule et al. 2005; Walsh et al. 2010). As a consequence, the approach of early language education embedded into daily routines can also be considered a very challenging approach for the early childhood professional. Early language education embedded in daily routines is strongly shaped by the individual situation and children's spontaneous interests and thoughts. It requires specific professional qualifications, experiences and various professional competencies such as content knowledge, pedagogical content knowledge, general pedagogical knowledge, motivational prerequisites (e.g. enthusiasm in providing language-related learning opportunities) and certain pedagogical beliefs (e.g. a positive attitude towards using multiculturalism as a pedagogical resource) (Anders et al. 2015). These assumptions point to the need of comprehensive professional development programmes. In the next section, the concept of early language education embedded into daily routines will be linked to a broader concept of ECEC quality.

### **A language-specific concept of ECEC quality**

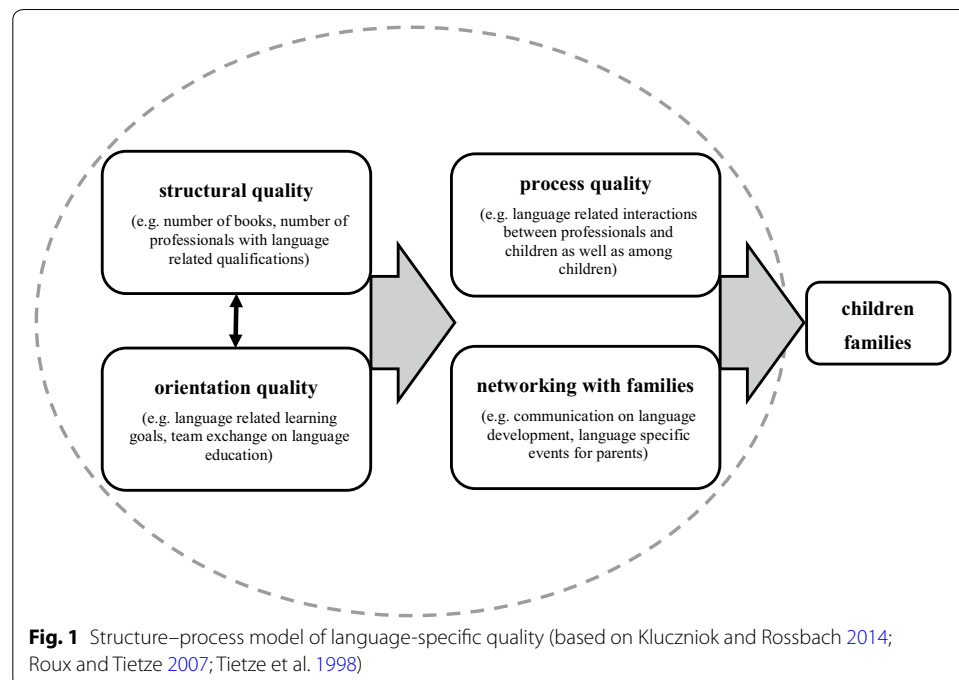
The quality of learning in ECEC is generally seen as a multidimensional concept covering structural characteristics, teachers' beliefs and orientations, and processes (NICHD ECCRN 2002a, b; Pianta et al. 2005). Structural quality is regarded as being subject to regulation by policy and funding. It covers characteristics such as class size, teacher-child ratio, formal staff qualification levels, provided materials and the size of the setting. Orientation quality refers to the pedagogical beliefs of the early childhood professionals (e.g. their definition of the professional role, educational values, attitudes with regard to the importance of different educational areas and learning goals). This quality dimension refers not only to the individual early childhood professional, but also to the setting, especially the pedagogical approach of the setting and its implementation. Process quality comprises the nature of the pedagogical interactions between early childhood professionals and children, the interactions among children, the interactions of children with space and materials and the quality of interactions between staff and parents (e.g. Lamb-Parker et al. 2001). It is commonly assumed that different dimensions of ECEC quality are interrelated. While the quality of pedagogical interactions should yield a direct impact on child development, structural quality and orientation quality are seen as prerequisites and determinants of process quality (Kluczniok and Rossbach 2014). Conceptualizations of ECEC quality make a difference between global characteristics and domain-specific aspects of the stimulation in learning areas such as literacy, emerging mathematics and science (Kluczniok and Rossbach 2014; Sylva et al. 2003). In doing so, they acknowledge the proven relevance of early domain-specific skills and abilities for later school success.

We assume that, as a consequence, language-specific components of ECEC quality need to be highlighted in the framework of early language education embedded into daily preschool routines (Anders et al. 2015). Structural aspects which are particularly relevant for early language education are, for example, the number of professionals with specific qualifications, the number of books or the existence of other language material. Looking at the quality dimension of orientation quality, epistemological beliefs regarding the development and learning of language and the support of the specific pedagogical

concept are important. At the level of the centre, the implementation of the pedagogical concept and team exchange on language education seem to be highly relevant. With regard to the process quality, it is the quality of the language-related interactions, which is most important. As process quality comprises the interactions between early childhood professionals and children, the overlaps and links with pedagogy and the pedagogical concept are obvious. Interactions between staff and parents can also be focused in a language-specific way. This relates to activities such as advice for parents in stimulating the language development of their child, language-specific events for parents or everyday communication on the language development of the child. Figure 1 illustrates the theoretical model of language-specific ECEC quality.

### The federal programme “core daycare centres language & integration”

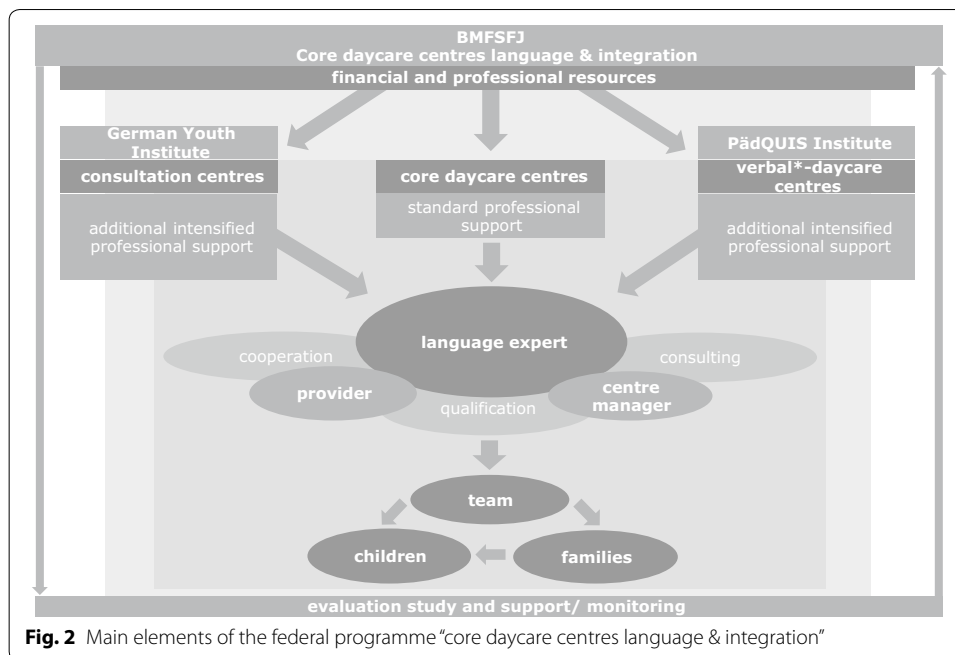
The federal programme “core daycare centres language & integration” was set up by the BMFSFJ in the year 2011. Nationwide, 4000 daycare centres were funded and supported to become core daycare centres for language and integration (in the following core daycare centres). Initially, the programme was set up to last until the end of 2014, but it was extended for a further year and ended by the end of 2015. The federal programme aimed at making a contribution to raise the level of language-related quality in the daycare centres and impact children’s language development. The eligible settings were located in socioeconomically disadvantaged communities; as a consequence, the number of children with an immigration background and the number of children who grow up in families with low socioeconomic or educational status in these centres were higher than average. The programme aimed at promoting early language education embedded into daily routines in the centres with a special focus on children who grow up in less stimulating families or who do not speak German at home. Furthermore, the initiative



wanted to accomplish other language initiatives in putting children under the age of 3 years in the focus of the federal programme. Each core daycare centre received funding for an additional part-time professional (0.5 = approximately 18.5 h per week), further training and learning materials. This language expert was required to be an ECEC professional and have additional training and experience in language education or the work with children under the age of 3 years. The central resource of the programme was the language expert who was meant to fulfil up to three tasks: (1) consulting, coaching and professional support of the team of the centre in early language education embedded into daily routines, (2) consulting, coaching and professional support of the team of the centre in establishing effective partnerships with the families and (3) exemplary language-related pedagogical work with children, particularly with children under the age of 3 years. Thus, the main focus of the language expert was on working with the team of the centre, and the assumed effect of the language expert on the children was an indirect effect. The language expert was meant to give support to the team, and this would raise the language-related process quality in the different groups of the setting and impact the children's development.

The settings and language experts received different forms of professional support to implement early language education embedded into daily routines into the pedagogical concept of the centre. Practice materials and a handbook (Jampert et al. 2011) developed by the German Youth Institute, one of Germany's largest social science institutes, were made available to all core daycare centres. Starting in September 2011, conferences organized by the steering group were held in all federal states to facilitate the programme implementation and the development of networks. In addition, an online platform was developed and new materials were fed into the platform consistently. Two hundred and fifty centres received additional intensified professional support as they were developed and certified by the German Youth Institute to become consultation centres. Further 500 daycare centres had the opportunity to take part in the professional support programme "verbal" which was developed and conducted by the PädQUIS Institute, Berlin. Centres who took part in the "verbal" programme were linked to regional networks of 10–15 centres, and they met regularly. On the meetings, professional input was given, and different topics related to the federal programme and the implementations of the pedagogical concept were discussed. A subsample of 40 centres received in addition video-based coaching. Professionals of these centres agreed to be videotaped while working with the children. The videos were used to reflect on and improve the learning opportunities and pedagogical interactions. Further professional support was regionally organized by the federal states and some of the childcare providers. Figure 2 provides an illustration of the main elements of the federal programme "core daycare centres language & integration".

The BMFSFJ defined the overarching themes and aims of the initiative, provided different forms of professional support and set certain requirements for the centres and professionals. But the core daycare centres and the language experts still had a lot of autonomy in adjusting the federal programme and the implementation of early language education to their individual needs. For example, the centre and the language expert were free in how to balance the different tasks. It could be expected that different realizations of the programme would develop. This variability was a main challenge to handle when planning the design of the evaluation study.



### The evaluation study

The evaluation was designed and conducted by the University of Bamberg (Prof. Dr. Hans-Günther Roßbach), Freie Universität Berlin (Prof. Dr. Yvonne Anders), and the PädQUIS Institute (Prof. Dr. Wolfgang Tietze). The three groups worked in close coordination with the steering group of the programme “core daycare centres language & integration”. As described above, the PädQUIS Institute also developed and conducted the professional support programme “verbal”. However, within the institute, the roles of the individuals were clearly split, meaning that those persons who were involved in the professional support programme were not involved in the evaluation study and vice versa. The funding for the study was provided by the BMFSFJ.

### Aims of the evaluation

The study was set up to establish the effects of the federal programme on (1) the language expert, (2) the daycare settings and the team of early childhood professionals, (3) the families and (4) the children. Due to the broad conceptualization of the federal programme, a comprehensive portfolio of potential effects was also likely to occur. The evaluation picked this up by considering various effect dimensions and variables. The overarching central goals were the identification of successful types of programme realizations and the identification of best practice approaches.

### Methods

#### Design and procedure

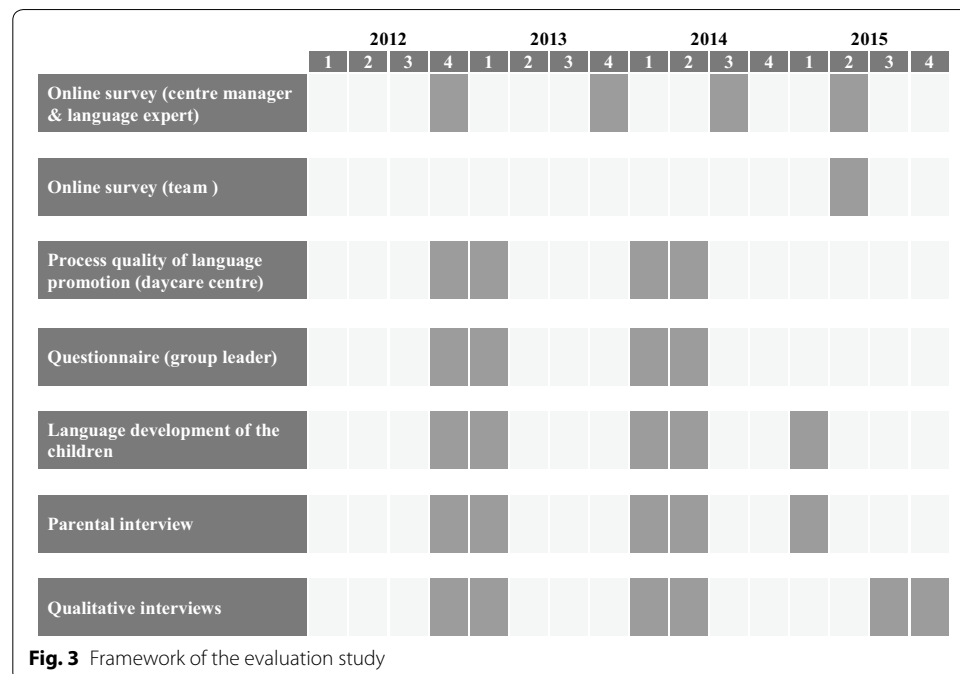
The evaluation was designed as a longitudinal study with four measurement points taking place between autumn/winter 2012 and summer 2016. To make effective use of variations in the quality of implementation of the programme, a quasi-experimental design

was chosen comparing four different groups of settings: (1) core daycare centres receiving standard professional support, (2) core daycare centres taking part in the qualification programme “consultation centres” offered by the German Youth Institute, (3) core daycare centres taking part in the qualification programme “verbal” offered by PädQUIS gGmbH and (4) daycare centres not participating in the federal programme serving as a comparison group. A mixed-method approach was chosen combining quantitative and qualitative research methods. Figure 3 illustrates the different methodological components and the longitudinal design.

Basic structural information on all core daycare centres was available through monitoring data collected by the steering group. Centre managers and the language experts of the settings included in the evaluation study filled in supplementary repeated online surveys. In addition, observations of the language-related process quality were carried out twice. The perspective of the parents was captured by half-standardized interviews. These data were collected in the homes of the families as part of a visit, and at the same time, the assessment of language skills of the children took place. Additional qualitative interviews with centre managers and language experts were conducted to get more insight into processes of implementation and factors of success.

**Sample**

The evaluation planned to include 80 settings per experimental group, 320 daycare centres altogether. This size was assumed to provide insight into the variations of implementation and developmental trajectories over the course of the programme. Core daycare centres of the three groups with different types of professional support were sampled with the aim of assuring representativity with regard to structural prerequisites (e.g. size of the centre, proportion of children with immigration background). Daycare centres of the comparison group were recruited in local proximity to the core daycare centres





included in the evaluation. In doing so, we aimed at achieving comparability of the comparison group with regard to the socioeconomic backgrounds of the families (mother tongue, income, educational level). The centres included in the study covered 15 out of 16 federal states, rural as well as urban areas and regions with varying social structures. The process quality of one group per centre was observed. When sampling the children, we aimed at including children as young as possible to test their language skills to be able to monitor their development. Furthermore, we tried to maximize the number of children who were cared for in the group, which was also chosen for quality observation. Based on these assumptions, we expected to be able to recruit at average four children and their families for participation in the study.

In total, 335 daycare centres and 1331 children and families were recruited for the study. Table 1 shows the number of participating settings and families in the different experimental groups over the course of the study. Thus, the recruitment was more successful than expected, and it was possible to include more settings, children and families in the study than initially planned; panel mortality was also low. While participation was compulsory for the core daycare centres, it was not for the families and the centres of the comparison group. Ten settings were chosen for the qualitative case studies based on the collected quantitative data representing interesting cases of programme implementation and quality management.

Twenty-five % of the centre managers and 27.5% of the language experts had a university or college degree. Thus, compared to the standard situation in German daycare centres, the sample includes a very high number of graduated early childhood professionals. It seems that the federal programme attracted centres staffed with professionals

**Table 1 Number of participating settings and families**

	Measurement points			
	1	2	3	4
Number of participating settings				
Consultation centres	n = 76 (22.7%)	n = 75 (22.9%)	n = 74 (23.3%)	n = 72 (24.8%)
Core daycare centres	n = 100 (29.9%)	n = 96 (29.3%)	n = 93 (29.3%)	n = 85 (29.3%)
Verbal*-daycare centres	n = 82 (24.5%)	n = 82 (25.0%)	n = 79 (24.9%)	n = 76 (26.2%)
Comparison group	n = 77 (23.0%)	n = 75 (22.9%)	n = 71 (22.4%)	n = 57 (19.7%)
Total sample	335	328	317	290
	<i>Dropout rate (%)</i>	2.1	5.4	13.4
Number of participating families				
Consultation centres	n = 325 (24.4%)	n = 275 (24.5%)	n = 224 (24.1%)	–
Core daycare centres	n = 413 (31.0%)	n = 358 (31.9%)	n = 288 (31.0%)	–
Verbal*-daycare centres	n = 309 (23.2%)	n = 251 (22.3%)	n = 210 (22.6%)	–
Comparison group	n = 284 (21.4%)	n = 239 (21.3%)	n = 206 (22.2%)	–
Total sample	1331	1123	928	–
	<i>Dropout rate (%)</i>	15.6	30.3	–

with high formal qualifications. The average age of the participating children at the first measurement point was 34 months. In total, 69% of the children grow up in families where both parents speak German with the child.

### **Instruments and indicators**

Established and validated instruments and indicators were accomplished with measures specifically developed for the purpose of the study. The evaluation used a broad and comprehensive measurement concept. Here we give insight into the central concepts at the different analytical dimensions to be able to discuss the methodological challenges.

### **Online surveys of the centre managers, language experts and the team**

The surveys of the centre managers and language experts asked for the motivation to participate in the federal programme, occurring problems as well as language-related activities for and with the parents and families and other possible partners. Furthermore, the survey asked for the existence and development of networks/supporting resources, the use of external and in-house trainings, courses and coaching and the stimulation of team exchange to implement the pedagogical concept of early language education embedded into daily routines. In addition, the survey sought to obtain information on the leadership concept of the daycare centre, especially with regard to the implementation of the programme.

The language expert<sup>1</sup> was asked to give information on her sociodemographic background, her formal qualifications and her professional career, how she understands and fills her role as language expert and what her language-related pedagogical beliefs and motivations were. The centres taking part in the federal programme cared for relatively large numbers of children with immigrant backgrounds. Therefore, the attitudes of the language expert towards multiculturalism and diversity were also included in the survey.

All respondents were also asked to rate their satisfaction with the programme and the different programme partners. Finally, centre managers and language experts were asked to elaborate on individual and centre strategies to realize early language education in the future once the federal programme will have expired.

### **Language-related process quality**

The evaluation used a set of established and validated instruments for the observation of language-related process quality: the German versions of the early childhood environmental rating scales ECERS-R (Tietze et al. 2007b) and ECERS-E (Rossbach and Tietze 2007), the German version of the infant/toddler environmental rating scales ITERS-R (Harms et al. 1998; Tietze et al. 2007a), the German translation of the Caregiver Interaction Scale CIS (Arnett 1989) and the Dortmund rating scale for the assessment of interactions relevant for the promotion of language skills DO-RESI (Fried and Briedigkeit 2008). Only those subscales relevant for language and diversity-related aspects of process quality were chosen.

---

<sup>1</sup> The vast majority of the language experts were female. To facilitate ease of reading, we use the female form to address to the language expert throughout the text, but the male form is always implied.

**Parental interview**

The parental interviews were conducted with the main caregiver of the child. In the interview, the parents gave information on structural background variables (e.g. marital status, immigration background, socioeconomic status, educational level), parental educational beliefs, experiences with the federal programme and the home learning environment of the child (e.g. the frequency of educational and language-related activities). German, Turkish, Russian and English versions of the interviews were developed. Parents could choose to answer the interview in the language they felt most confident in.

**Child development**

The testing of children's language skills took place three times over the course of the study in the homes of the families. Established standardized tests were used to assess two aspects of language skills: receptive vocabulary and comprehension of sentences. Only German language skills were tested. In addition, a short test for general cognitive skills was applied. Furthermore, parents and early childhood professionals were asked to rate different aspects of the development of their child.

**Collaboration of partners**

A governance board for the programme was set up and met once a month throughout the course of the project to discuss project matters. This board was initiated by the responsible department of the BMFSFJ and included members of the Federal Ministry, the coordinating office, the public relations office, members of the professional support teams (German Youth Institute and PädQUIS gGmbH) and the evaluation team. External guests were invited whenever needed. In addition, the steering group of the federal government and the federal state representatives met twice a year. All members of the governance board also participated in these meetings.

**Dissemination strategy**

The evaluation team reported regularly on the progress and findings of the study on the meetings of the governance board and the steering group of the federal government and the federal state representatives. These presentations were meant to serve formative needs, thus stimulating programme adjustments and improvements. As a consequence, the evaluation team started early and continued to analyse and report incoming data immediately. In addition, a final report for policy-makers and practitioners was planned, as well as a series of scientific paper publications. To inform scientific public, the evaluation team also reported results (in progress) on various conferences in Germany, Europe and the USA.

**Challenges of evaluating political initiatives in the area of ECEC**

The example of the evaluation of the federal programme "core daycare centres language & integration" in Germany will be used to discuss a number of challenges occurring when evaluating political initiatives in the area of ECEC.

**Sampling**

Evaluations of initiatives such as the federal programme "Core daycare centres: language and integration" encounter a number of challenges. Often, a compromise needs to be

established between scientific standards and policy needs. Daycare centres that took part in the initiatives also had to take part in the evaluation study, if they were sampled. This was certainly a great advantage with regard to sample size. At the same time, a careful communication between the evaluation team and the sampled settings was necessary to make sure that the answers given in the survey were valid. Furthermore, the families did not have to take part in the study. To motivate parents, the early childhood professionals and the centre managers were needed. In the evaluation, we were very successful in recruiting families and children. This was not only due to the fact that the topic was appealing, but also due to careful communication and attractive incentives for families and practitioners. Attractive incentives were not so much vouchers or other goods, but the offer of further professional development courses for early childhood professionals or feedback regarding the language development of the child for the parents.

### **The lack of instruments in ECEC research**

Empirical educational research on ECEC effects involves the study of the development of very young children. Sophisticated instruments have been developed to monitor the development of young children in general cognitive, language-related and mathematical skills. Many other domains have not been subject to research so far. The possibility of testing the development of children's abilities and skills is quite often restricted to their own language skills and their ability to act naturally in playful testing situations. Due to these restrictions, testing of the vocabulary of the child cannot take place earlier than in the middle of the third year of life. But this in turn puts limitations in studying effects of educational initiatives aiming at younger children such as the federal programme "core daycare centres language & integration".

Compared to educational research on later school phases, empirical educational research on ECEC is a relatively young discipline, particularly in Germany (Anders and Rossbach 2013). Naturally, a lot of the research in place still is exploratory in nature and for many important constructs, reliable instruments still need to be developed and validated for the particular purposes and contexts. Empirical research on ECEC has a much longer tradition in the US, and reliable and valid instruments for constructs such as preschool quality have been developed. However, many of the instruments developed for the US context cannot be assumed to be valid in the German context, because the nature of the ECEC systems is very different in the US compared to Germany (see Kuger et al. 2012 for a discussion). As a consequence, instruments need to be developed and validated for the German context. But instrument validation is a research topic on its own and validation studies need time. But when policy-makers urgently need findings on the effectiveness of different policy means, time for validation studies is rare or not given at all. In these cases, researchers need to develop instruments for evaluation studies in very short time without being able to conduct careful pilot and validation studies. Naturally, these instruments will show the need of improvement and this will also be evident in the capacity of the instruments to capture policy effects.

For example, when the evaluation study of the programme "core daycare centres" was designed, no validated measure to capture the process quality of the specific pedagogical approach of early language education embedded into daily preschool routines existed. As has been described above, the evaluation team chose to use a combination

of subscales of validated instruments that tapped relevant aspects, but which were also not specifically designed to assess the quality of language education embedded into daily routines. Although the observations of language-related preschool quality generated important and fruitful information, not all relevant characteristics of the specific pedagogical approach were tapped. We also experienced a surprisingly low stability of the process observations over the course of the study. The correlations between two measurement points ranged between 0.11 and 0.18 for different scales. This result points either to rather unpredictable trajectories of quality development or the further need of improvement of the observational tool.

### **The nature of pedagogy in ECEC**

Pedagogy in ECEC in Germany follows a child- and situation-oriented tradition and so does the concept of early language education embedded into daily preschool routines. This approach does not only imply specific challenges for the early childhood professional but also for the researcher trying to grasp information on the quality of implementation of this approach. This is especially true for the quality of the interactions between early childhood professionals and children. Child- and situation-oriented pedagogy happens systematically, but is not tied to specific hours of the day, while formal schooling is organized according to a timetable, which can also be used to plan observations of pedagogical interactions. Appropriate situations to be extended for language education in ECEC may occur throughout the day. One day a conversation with the group at the water tray may happen to be meaningful, and on the other day the professional will have a good and deep talk with one child regarding putting on shoes for going outside. The early childhood professional needs not only to pick up interests and questions of the children systematically, but also spontaneously. Certain situations such as mealtime may be especially suitable to be used for language education, but in general the professional cannot foresee when suitable situations occur. Due to time and cost limitations of researchers as well as practitioners, the researcher can only be present in the setting for a limited time, so it is very likely that the researcher misses out on important interactions and it is arguable if a time sample of 3–4 h is representative for the practice in the group and the setting. Praetorius et al. (2014) recently showed for the observations of instructional quality of mathematics lessons in secondary school that the observation of nine lessons would be needed to generate reliable and valid conclusions on the level of cognitive activation. Given the nature of child- and situation-oriented pedagogy in ECEC, it may be hypothesized that observational intervals in ECEC would need to be even longer.

### **The autonomy given to the settings and early childhood professionals**

The autonomy which is given to the settings and early childhood professionals in Germany when designing learning environments for the children and when setting specific educational goals is also evident in the conceptualization of the federal programme “core daycare centres language & integration”. It has been explained that the possibility of various types of implementation also resulted in a broad and comprehensive measurement concept of the evaluation. The comprehensive approach evoked methodological challenges. First of all, it produced time and cost-intensive data collection for the evaluation

team, but especially for the participants. Evaluations should try to keep the burden for the participants to a minimum. A second challenge related to applying a broad concept of measurement is the reduced capacity of broad measures to detect effects compared to focused measures on focused programmes. The professionals of the core daycare centres chose very specific goals they wanted to achieve through their participation in the federal programme, dependent on the individual situation and qualifications of the settings and professionals. But these goals differed greatly between centres. Additionally, the evaluation team could not anticipate exactly how the settings were going to translate the programme into practice for their individual needs. Naturally, the broad measures applied were not able to detect all specific developments and fine-grained effects that occurred over the course of the programme. This needs to be considered when drawing conclusions from the evaluation. Otherwise, achieved benefits of the federal programme are underestimated.

#### **The lack of control in national initiatives**

The programme “core daycare centres language & integration” picked up the topic of early language education which was of high interest to ECEC in Germany for a number of years already. Four thousand daycare centres in Germany, which is 5% of all daycare centres, received a meaningful structural resource. Providers of daycare used the initiative to invest into professional development of their professionals in this area, irrespective of the fact if these professionals worked in settings participating in the programme or not. The perceived movement of the field towards a more systematic and deepened approach of early language education may be seen as one important side effect of the federal programme. But for the evaluation, this evoked an additional challenge as it had impact on the power of the comparative group design. We found that daycare centres of the comparison group were also quite active with regard to the development of their pedagogical approach in the area of early language education. In addition to the general movement of the field, this pattern may reflect a selection bias, as these settings were taking part on a voluntary basis and it is unlikely that settings would agree to participate in such an extensive project if the core topic would not be important to them. As a consequence, differences between core daycare centres and the comparison group were smaller than expected, and this limitation also needs to be considered carefully when interpreting and communicating the findings.

#### **Effect sizes**

The discussed challenges of evaluating a broad and unspecific intervention also lead to the challenge of establishing and communicating effect sizes. In quasi-experimental or regression-type study designs, comparable effect sizes can be established which represent the size of the effect on a scale between 0 and 1. The most commonly used rule to evaluate the size of effects is the one introduced by Cohen (1992) who suggested to consider effects of 0.2 as small, effects of 0.5 as medium and 0.8 as large. At the first glance, these rules of interpretation are easy to apply and are nowadays also well known to many non-statisticians and policy-makers. The debate about the suitability of Cohen’s rule in empirical educational research is less well known. The effect sizes found in research are dependent not only on the size of the effect of the treatment, the capacity of the

instruments used to capture the effects, but also on the controllability of other influences. While Cohen's rules have been developed against the background of pure experimental settings, controlling carefully for all other influences despite of the intervention, effect sizes in empirical educational research are generated in much less well-controlled situations. Effect sizes in educational research are usually 0.2 or lower, and researchers have argued to consider them as meaningful, nevertheless especially in policy evaluation contexts (see Elliot and Sammons 2004 for a discussion).

With regard to the evaluation of the federal programme "core daycare centres language & integration", a number of aspects have to be considered to achieve realistic expectations regarding the effect sizes. First of all, language development has shown to depend heavily on characteristics of the child, the family and the home environment (e.g. Ebert et al. 2013); thus, a strong treatment is necessary to yield statistical significant effects. The conception of the programme implies an indirect transmission of the treatment on the families and children. Thus, effects on children may only be expected after the setting has undergone a significant change; frictional loss occurs naturally. Taken together, all this leads to the conclusion that the effect sizes on children's language development will have to be rather small in a statistical sense. But this does not mean that the impact of the federal programme is small. A rigid application of Cohen's effect size rules results in an underestimation of the impact of the national initiative.

### **Possibilities of evaluating political initiatives in the area of ECEC**

National initiatives such as the federal programme "core daycare centres language & integration" have the potential to stimulate a meaningful change of the field of ECEC in Germany as a whole. Individual early childhood professionals and centres gained experiences in how to raise the language-related quality in their settings. These experiences can be transferred to others in the field through reports, presentations, video material and other practice materials. These materials can further stimulate ECEC practice. Carefully planned and comprehensive evaluations of such programmes are highly needed for several reasons. First of all, there is a lack of evidence on the effects of ECEC characteristics and the effectiveness of different pedagogical approaches in the area of ECEC in Germany (Anders and Rossbach 2013). Thus, findings on these topics are needed to accumulate research evidence and to stimulate the scientific discourse. The framing conditions and the setting of evaluation studies on the effects of political initiatives in the area of ECEC in Germany put limit on the methodology and the interpretation of the findings. But on the other hand, these studies have a central advantage, which is the fact that the effectiveness of interventions and the factors of success are investigated under real life conditions. This makes a strong case for the ecological validity. Policy implementation studies provide valuable and important knowledge in addition to controlled experimental trials and studies on the effects of regular ECEC. US experiences showed that Head Start was an extremely effective approach when conducted under strongly controlled and costly conditions. The effects of Head Start after it was rolled out to become a national initiative are much less convincing (Puma et al. 2010).

From the perspective of policy-makers, evaluations of publicly funded programmes are necessary to justify their spending of public money. Continuous monitoring of programme implementation is also needed to adjust programmes if misguided

developments become evident. In the evaluation study, analyses of the collected data were undertaken as soon as the first data were collected. The findings were fed back continuously to the governance board and the steering group to facilitate the identification of necessary programme adjustments, to stimulate the practice and to facilitate the development of further professional support by the steering group. For example, according to the idea of the federal programme, the language expert should have the effect of a multiplier for the team, in sharing new knowledge, developing a concept of early language education for the centres with the other team members, coaching others and being a role model of good practice. But the survey data of the first measurement of the evaluation showed that some language experts experienced problems in defining their role as a multiplier for the team. Many language experts stated that they mainly work individually with children. Early analyses also showed that the working area “working with parents” was neglected in many settings when starting the initiative. The steering group picked this up and developed practice support material to guide practice into different directions. Early findings also highlighted the importance of team exchange and in-house trainings for the language-related process quality. These results were also transferred back to the settings and childcare providers. Furthermore, empirical evidence on the effectiveness of programmes is central to the idea of evidence-based policy. The findings of the evaluation of the programme “core daycare centres language & integration” were carefully analysed by the policy-makers and used to develop policy measures, practice materials and other programmes which have the potential to raise the quality of ECEC effectively. The Federal ministry launched a new programme “Language-preschools (Sprach-Kitas)”. To develop and conceptualize this programme, evidence of the evaluation study was used. A comprehensive system of professional support was established, because the evaluation showed that continuous professional support has impact on the implementation quality. Furthermore, the new programme focuses on parent–preschool partnership, because the evaluation had shown that this area of work needs to be more stressed by the professionals. The new evaluation study also used the experiences. It now has a stronger qualitative and formative part to be able to inform practice in a more procedural way.

## **Conclusion**

Empirical evidence regarding the effectiveness of different means to raise the quality of ECEC in Germany is rare and urgently needed. Careful evaluations need comprehensive measurement approaches, a commitment to high methodological standards and its financial costs and the willingness of all stakeholders to communicate and collaborate in a constructive way. If these requirements are considered, evaluations of policy implementation can produce valuable and meaningful findings, in addition to other types of research such as controlled experimental trials and studies on the effects of regular ECEC.

### **Author's contribution**

The authors developed the design and the instruments of the study together. The fieldwork was carried out by the teams of University of Bamberg and PädQuis gGmbH. YA lead on the statistical analyses. She was the lead author of the manuscript. HGR and WT commented on the drafts. All authors read and approved the final manuscript.



**Author details**

<sup>1</sup> Early Childhood Education, Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 94, 14195 Berlin, Germany. <sup>2</sup> Leibniz-Institute for Educational Trajectories, Bamberg, Germany. <sup>3</sup> PädQUIS gGmbH, Berlin, Germany.

**Competing interests**

The authors declare that they have no competing interests.

Received: 5 April 2016 Accepted: 18 November 2016

Published online: 07 December 2016

**References**

- Anders, Y. (2013). Stichwort: Auswirkungen frühkindlicher, institutioneller Bildung und Betreuung. *Zeitschrift für Erziehungswissenschaft*, *16*, 237–275.
- Anders, Y., Ballaschk, I., Dietrichkeit, T., Flöter, M., Groeneveld, I., Lee, H. J., et al. (2015). Evaluation des Bundesprogramms „Schwerpunkt-Kitas Sprache & Integration“: Abschlussbericht. In H. G. Rossbach, Y. Anders, & W. Tietze (Eds.), *Evaluation des Bundesprogramms „Schwerpunkt-Kitas Sprache & Integration“*. Bamberg: Abschlussbericht.
- Anders, Y., & Rossbach, H. G. (2013). Blick in die frühkindliche Bildungsforschung in Deutschland. In M. Stamm & D. Edlmann (Eds.), *Handbuch Frühkindliche Bildungsforschung* (pp. 183–195). Wiesbaden: Springer Fachmedien. doi:10.1007/978-3-531-19066-2\_13.
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, *10*(4), 541–552.
- Barnett, W. S., & Hustedt, J. T. (2005). Head start's lasting benefits. *Infants & Young Children*, *18*, 16–24.
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Deutsches PISA-Konsortium (Ed.), *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 323–407). Leske: Opladen.
- Bock-Famulla, K., Lange, J., & Strunz, E. (2015). *Länderreport frühkindliche Bildungssysteme 2015: Transparenz schaffen—Governance stärken*. Gütersloh: Verlag Bertelsmann Stiftung.
- Bos, W., Lankes, E. M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (Eds.). (2003). *Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster: Waxmann.
- Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record*, *112*, 579–620.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- Cohen, N. J. (2006). *The impact of language development on the psychosocial and emotional development of young children. Encyclopedia of language and literacy development* (pp. 1–7). London: Canadian Language and Literacy Research Network. <http://www.literacyencyclopedia.ca/pdfs/topic.php?topid=4>. Accessed 21 Oct 2016.
- Ebert, S., Lockl, K., Weinert, S., Anders, Y., Klucznik, K., & Rossbach, H.-G. (2013). Internal and external influences on vocabulary development in preschool children. *School Effectiveness and School Improvement*, *24*, 138–154.
- Elliot, K., & Sammons, P. (2004). Exploring the use of effect sizes to evaluate the impact of different influences on child outcomes: Possibilities and limitations. In I. Schagen & K. Elliott (Eds.), *But what does it mean? The use of effect sizes in educational research* (pp. 6–24). Slough: National Foundation for Educational Research.
- Fried, L. (2013). Sprachförderung. In L. Fried & S. Roux (Eds.), *Handbuch Pädagogik der frühen Kindheit* (pp. 175–181). Berlin: Cornelsen Scriptor.
- Fried, L., & Briedigkeit, E. (2008). *Sprachförderkompetenz: Selbst- und Teamqualifizierung für Erzieherinnen, Fachberatungen und Ausbilder*. Berlin: Cornelsen Scriptor.
- Gasteiger-Klicpera, B., Knapp, W., & Kucharz, D. (2010). *Abschlussbericht der wissenschaftlichen Begleitung des Programms „Sag' mal was—Sprachförderung für Vorschulkinder“*. [http://www.ph-weingarten.de/zep/Projekte/Abschlussbericht\\_Sprachfoerderung\\_Landesstiftung\\_PH\\_Weingarten.pdf](http://www.ph-weingarten.de/zep/Projekte/Abschlussbericht_Sprachfoerderung_Landesstiftung_PH_Weingarten.pdf).
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale* (Rev ed.). New York: Teachers College Press.
- Jampert, K., Thanner, V., Schattel, D., Sens, A., Zehnbauser, A., Best, P., et al. (2011). Die Sprache der Jüngsten entdecken und begleiten: Überall steckt Sprache drin (2). Weimar, Berlin: verlag das netz.
- Kammermeyer, G., Roux, S., & Stuck, A. (2011). Additive Sprachförderung in Kindertagesstätten: Welche Sprachfördergruppen sind erfolgreich? *Empirische Pädagogik*, *25*, 439–461.
- Klucznik, K., & Rossbach, H.-G. (2014). Conceptions of educational quality for kindergartens. *Zeitschrift für Erziehungswissenschaft*, *17*, 145–158.
- Kuger, S., Sechtig, J., & Anders, Y. (2012). Kompensatorische (Sprach-) Förderung: Was lässt sich aus US-amerikanischen Projekten lernen? *Frühe Bildung*, *1*, 181–193.
- Lamb-Parker, F., Piotrkowski, C. S., Baker, A. J. L., Kessler-Sklar, S., Clark, B., & Peay, L. (2001). Understanding barriers to parent involvement in Head Start: A research-community partnership. *Early Childhood Research Quarterly*, *16*(1), 35–51.
- Lisker, A. (2011). *Additive Maßnahmen zur vorschulischen Sprachförderung in den Bundesländern: Expertise im Auftrag des Deutschen Jugendinstituts*. München: Deutsches Jugendinstitut.
- Müller, K., & Ehmke, T. (2013). Soziale Herkunft als Bedingung der Kompetenzentwicklung. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Eds.), *PISA 2012: Fortschritte und Herausforderungen in Deutschland* (pp. 245–274). Münster: Waxmann.
- NICHD ECCRN. (2002a). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, *39*, 133–164.

- NICHD ECCRN. (2002b). Early child care research network: The interaction of child care and family risk in relation to child development at 24 and 36 months. *Applied Developmental Science*, 6, 144–156.
- Oertel, F. M. (1984). *Preschool social education curricula in West Germany: The situation-oriented approach*. Urbana: ERIC Clearinghouse on Elementary and Early Childhood Education, College of Education, University of Illinois.
- Pianta, R. C., Howes, C., Burchinal, M., Bryant, D., Clifford, D., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9, 144–159.
- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction*, 31, 2–12.
- Preissing, C. (2007). *Handreichungen für die Praxis zum Bildungsprogramm für saarländische Kindergärten*. Weimar: verlag das netz.
- Protz, R., & Preissing, C. (2006). *Bridging diversity: an early childhood curriculum*. Weimar: verlag das netz.
- Puma, M., et al. (2010). *Head Start impact study: Final report*. Washington, DC: Administration for Children and Families, U.S. Department of Health and Human Services.
- Reynolds, A. J., Temple, J. A., & Ou, S. (2010). Preschool education, educational attainment, and crime: Contributions of cognitive and non-cognitive skills. *Children and Youth Services Review*, 32, 1054–1063.
- Rossmach, H.-G., & Tietze, W. (2007). *Kindergarten-Skala: Erweiterung KES-E [German version of The Early Childhood Environment Rating Scale: Extension ECERS-E] (Unpublished instrument)*. Bamberg: University of Bamberg.
- Roux, S., & Tietze, W. (2007). Effekte und Sicherung von (Bildungs-) Qualität in Kindertageseinrichtungen. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 27(4), 367–384.
- Sachse, S., Budde, N., Rinker, T., & Groth, K. (2012). Evaluation einer Sprachfördermaßnahme für Vorschulkinder. *Frühe Bildung*, 1, 194–201.
- Siraj-Blatchford, I., Sylva, K., Muttock, S., Gilden, R., & Bell, D. (2002). *Researching effective pedagogy in the early years. (Research Report No. 356)*. London: Department for Education and Skills.
- Sproule, L., Rafferty, H., Trew, K., Walsh, G., McGuinness, C., Sheehy, N., & O'Neill, B. (2005). The early-years enriched curriculum evaluation project: Final report phase 1. Ireland: Report produced for Northern Ireland Council for Curriculum Assessment and Examinations (CCEA).
- Statistische Ämter des Bundes und der Länder. (2016). Kindertagesbetreuung. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Soziales/SozialeLeistungen/Kindertagesbetreuung/Kindertagesbetreuung.html;jsessionid=2B892B844B98D0D538CA0D3D34DCB06A.cae3>.
- Stevenson, J., Richman, N., & Graham, P. J. (1985). Behaviour problems and language abilities at three years and behavioural deviance at eight years. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 26(2), 215–230.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research*, 9, 109–124.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early childhood rating scale extensions (ECERS-E)*. Stoke On Trent: Trentham Books.
- Tietze, W., Bolz, M., Grenner, K., Schlecht, D., & Wellner, B. (2007a). *Krippen-Skala (KRIPS-R): Feststellung und Unterstützung pädagogischer Qualität in Krippen*. Berlin: Cornelsen.
- Tietze, W., Knobloch, J., & Gerszonowicz, E. (2007b). *Kindergarten-Skala (KES-R): Feststellung und Unterstützung pädagogischer Qualität in der Kindertagespflege*. Berlin: Cornelsen.
- Tietze, W., Meischner, T., Gänsfuß, R., Grenner, K., Schuster, K.-M., Völkel, P., et al. (1998). *Wie gut sind unsere Kindergärten? Eine Untersuchung zur pädagogischen Qualität in deutschen Kindergärten*. Neuwied: Luchterhand.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & Catts, H. (2000). The association of reading disability, behavioral disorders, and language impairment among second-grade children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(4), 473–482.
- Walsh, G., McGuinness, C., Sproule, L., & Trew, K. (2010). Implementing a play-based and developmentally appropriate curriculum in NI primary schools: What lessons have we learned? *Early Years. An International Journal of Research and Development*, 30(1), 53–66.
- Weinert, S. (2006). Sprachentwicklung [Language development]. In W. Schneider & B. Sodian (Eds.), *Enzyklopädie der Psychologie CN/2—Kognitive Entwicklung* (pp. 609–719). Göttingen: Hogrefe.
- Wolf, K. M., Felbrich, A., Stanat, P., & Wendt, W. (2011). Evaluation der kompensatorischen Sprachförderung in Brandenburger Kindertageseinrichtungen. *Empirische Pädagogik*, 25, 423–438.
- Wolf, K., Schroeders, U., & Kriegbaum, K. (2016). Metaanalyse zur Wirksamkeit einer Förderung der phonologischen Bewusstheit in der deutschen Sprache. *Zeitschrift für Pädagogische Psychologie*, 30, 9–33.