

91

Data Documentation

Deutsches Institut für Wirtschaftsforschung

2017

The Early Childhood Education and Care Quality in the Socio-Economic Panel (SOEP-ECEC Quality) Study – K²ID-SOEP Data

Pia S. Schober, C. Katharina Spieß, Juliane F. Stahl, Gundula Zoch and Georg F. Camehl

IMPRESSUM

© DIW Berlin, 2017

DIW Berlin

Deutsches Institut für Wirtschaftsforschung

Mohrenstr. 58

10117 Berlin

Tel. +49 (30) 897 89-0

Fax +49 (30) 897 89-200

www.diw.de

ISSN 1861-1532

All rights reserved.

Reproduction and distribution

in any form, also in parts,

requires the express written

permission of DIW Berlin.

Data Documentation 91

Pia S. Schober^{1,2}, C. Katharina Spieß^{1,3}, Juliane F. Stahl^{1,4}, Gundula Zoch⁵ and Georg F. Camehl^{1,3}

The Early Childhood Education and Care Quality in the Socio-Economic Panel (SOEP-ECEC Quality) Study – K²ID-SOEP Data

Forschungsprojekt im Auftrag der Jacobs Foundation (Projektnummer 2013-1063).

Berlin, 29 June 2017

¹German Institute for Economic Research (DIW Berlin)

²University of Tübingen

³Freie Universität Berlin

⁴International Max Planck Research School on the Life Course

⁵University of Bamberg

Table of contents

- 1 Objectives of the study..... 1**
- 2 Sample and survey design 2**
 - 2.1 Data basis and sample 2
 - 2.2 Interview methodology and survey instruments..... 7
 - 2.2.1 Survey instruments at the household level..... 7
 - 2.2.2 Survey instruments at the institution level..... 13
 - 2.2.2.1 Director questionnaire 14
 - 2.2.2.2 (Group) Educator questionnaire 18
 - 2.2.2.3 Compressed questionnaire 20
- 3 Survey preparation and implementation 21**
- 4 Pretest results and response rates 26**
 - 4.1 Pretest results 26
 - 4.2 Response rates of the main study 27
- 5 Data structure and weighting 34**
 - 5.1 Variables..... 34
 - 5.2 Data sets..... 36
 - 5.3 Weighting..... 39
 - 5.3.1 Weights for the Parent Survey 39
 - 5.3.2 Weights for the Institution Survey 40
- 6 Survey instruments 41**
- Literature 42**
- Appendix..... 44**

List of tables

Table 1 Net sample of households in SOEP-ECEC Quality divided by subsamples	5
Table 2 Sample and results of Parent Survey by wave	29
Table 3 Sample and results of Institution Survey by wave	30
Table 4 Questionnaire versions and levels of information by wave (Institution Survey).....	31
Table 5 Net sample of children in SOEP-ECEC Quality by wave.....	33
Table 6 Overview of questions in Parent Survey, their sources and modifications if applicable	44
Table 7 Overview of questions in Institution Survey, their sources and modifications if applicable	47

List of figures

Figure 1 Gross sample of SOEP-ECEC Quality divided by subsamples 4
Figure 2 Locations of institutions attended by children in SOEP-ECEC Quality 6
Figure 3 Timeline of data collection for SOEP-ECEC Quality 22

Acknowledgments

The authors thank the Jacobs Foundation for funding the “Early Childhood Education and Care Quality in the Socio-Economic Panel (K²ID-SOEP, Project number 2013-1063)” project, which allowed to create the data set described in this study.

1 Objectives of the study

This survey report summarizes main information on the data collection process within the study ‘Early childhood education and care quality in the Socio-Economic Panel’ (SOEP-ECEC Quality).¹ The project ran from September 2013 to April 2017 and was funded by the Jacobs Foundation (project number 2013-1063).

In the context of the recent expansion of childcare places in early childhood education and care (ECEC) institutions for children below three years of age as well as increasing availability of full-day care also for children aged three years and older in Germany, the project constructed a unique data set on provision and access to ECEC quality. This was accomplished by carrying out surveys which can be linked to the Socio-Economic Panel (SOEP), the largest and longest running multidisciplinary household panel in Germany (Wagner et al. 2007).²

The K²ID-SOEP extension study (short: K²ID) was realized between 2013 and 2015. K²ID is an acronym for ‘Kinder und Kitas In Deutschland’ (‘children and daycare centers in Germany’), the official name used in communications with survey respondents. K²ID encompasses two surveys, the Parent Survey and the Institution Survey. In a first step, the project surveyed parents of all SOEP children below school age to investigate the parental decision making process and subjective evaluations with respect to ECEC quality of the institutions attended by their children. In a second step, data on the ECEC institutions attended by these children were gathered from directors and group educators of the ECEC institutions. These data include various quality indicators, primarily indicators of structural and orientation quality.

By combining institutional information on the educational context with individual and household data collected in the SOEP, the data set allows to examine associations with parental decision making and parental outcomes and children’s development first during early childhood and school years and subsequently until adulthood. Researchers may analyze the transition into quality by socio-economic status and the effects of ECEC quality on parental

¹ The authors would like to thank all student research assistants and interns who were involved in the project: Nittaya Fuchs, Clara Höltermann, Josefine Koebe, Oktay Tuncer, Matthias Weierer and Marian Weigt. Gundula Zoch contributed to this report while she was a researcher at the Department of Education and Family at the DIW Berlin.

² This data set also contains all waves of the ‘Families in Germany’ study (Schröder et al. 2013). For further information see Box 1 in chapter 2.

employment and wellbeing as well as on the socio-emotional development of children. Hence, the study provides information on possible direct consequences of ECEC quality on children's short, medium and long term educational and labor market outcomes and inter-generational mobility, as well as indirect effects on child wellbeing through parental employment and wellbeing. Moreover, it enables to investigate both the socio-economic selectivity in parental choices of ECEC quality and information asymmetries between mothers and ECEC providers. A summary of the project, the data collection and related publications can be found on www.k2id.de. The data were collected in collaboration with the institute 'TNS Infratest'.³

Please note that this report refers to the core SOEP-ECEC Quality study. The Institution Survey, however, was also conducted among ECEC centers attended by 5-year old twins who participated in the SOEP-Related Study 'TwinLife'.⁴ This is a representative behavior genetic study on pairs of twins and their families. Regarding information that is specific to the TwinLife sample, there will be an additional report available.

2 Sample and survey design

2.1 Data basis and sample

The sampling system of SOEP-ECEC Quality was based on all households with children below school age in the German Socio-Economic Panel (SOEP). More precisely, target households comprised those with a child born on 1 September 2007 or later and which participated in the 2013 wave and did not refuse survey participation for 2014 from the outset. Therefore, at the beginning of the field work in October 2013, target children were between zero and six years of age.

The SOEP is a representative longitudinal survey of private households and persons in Germany that was started in 1984 (Wagner et al. 2007). The SOEP consists of numerous subsamples that either existed from the beginning or were added during the course of the study (see Box 1 for more information on the subsamples from which the SOEP-ECEC Quality sam-

³ In September 2016, the institute was renamed 'Kantar TNS' (see <http://www.tns-infratest.com/>).

⁴ For additional information on *TwinLife*, please consult www.twin-life.de.

ple was drawn). In 2013, 19,406 persons from 11,447 households of the main sample (A-H, J-K) were interviewed. In addition, 4,964 interviews with individuals from 2,723 households of sample M were realized (Gerstorf and Schupp 2014; for an overview of the longitudinal development of the SOEP we refer to Kroh et al. 2015).

The SOEP has been augmented by several supplementary studies, most importantly the ‘Families in Germany’ study (FiD, “Familien in Deutschland”) (Schröder et al. 2013). FiD is a data set on households with young children and households with special needs (low income, lone parents, and large families, see Box 1), which were first surveyed in 2010 and were integrated into the usual SOEP in 2014. The FiD data contain information from about 4,500 households with a total of about 7,500 respondents and over 8000 children. Since 2014, the FiD sample has been part of the SOEP (for further details see <http://www.diw.de/fid-soep>). SOEP and FiD data can be analyzed jointly using sampling weights.

For the SOEP-ECEC Quality study, all main SOEP and FiD subsamples questioned in 2013 were taken into account (‘main sample’, wave 1 in 2013/2014). As an exception, we postponed sampling and data collection by one year with respect to sample M, a newly added subsample consisting of households with a migration background (wave 2 in 2014/2015). All parents in households meeting the aforementioned criteria were asked to participate in the Parent Survey.

Box 1: Subsamples in SOEP-ECEC Quality

Within the SOEP, 18 subsamples are included, covering households in both East and West Germany but also households with at least one member of foreign origin or high income. All non-migration samples of the SOEP are multi-stage random samples which are all regionally clustered, and all households were selected by random-walk (SOEP Group 2005: p. 19-21). In 2013, for the third time in the history of the SOEP, a subsample was added that is composed exclusively of migrant households. Sample M represents younger generations of migrants, including immigrants since 1995 and second generation migrants, which were underrepresented for the decade before (for further Information see Gerstorff and Schupp 2014).

The FiD data set includes seven subsamples, four of which are representative cohort samples ('Cohort samples') based on information from the local residents' registration office. For the remaining subsamples ('Screening samples'), which represent single parents, low income families, and large families with at least three children, no particular sampling frame exists in Germany. Therefore, sampling was conducted using a screening process (for further information see Schröder et al. 2013 and Jänsch et al. 2011). For more details on the composition of the SOEP-ECEC Quality sample, see Figure 1.

Figure 1
Gross sample of SOEP-ECEC Quality divided by subsamples

SOEP		FiD	
Subsamples A-H, J-K	Migration Sample M	Cohort Samples Birth cohorts 2007-2010	Screening Samples Low income, lone parents, large families
906 Households	527 Households	1096 Households	536 Households
 <p>3065 HOUSEHOLDS WITH AT LEAST ONE CHILD BORN ON 1 SEPTEMBER 2007 OR LATER</p>			
<p>GROSS SAMPLE = 3916 CHILDREN</p>			

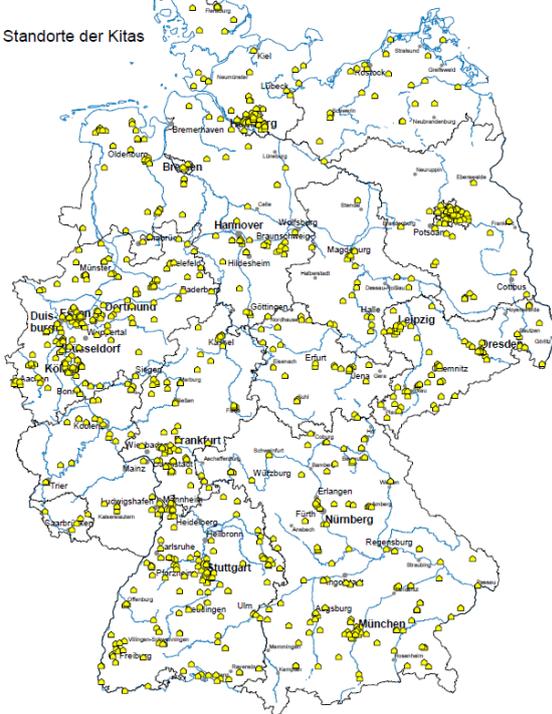
Table 1
Net sample of households in SOEP-ECEC Quality divided by subsamples

	Wave 1 (2013/2014)								Wave 2 (2014/2015)	
	SOEP A-K		FiD Cohorts		FiD Screening		Total		SOEP M	
	N	%	N	%	N	%	N	%	N	%
Gross sample	906	100.0	1096	100.0	536	100.0	2538	100.0	527	100.0
Non-response	115	12.7	328	29.9	225	42.0	668	26.3	170	32.3
Completed	791	87.3	768	70.1	311	58.0	1870	73.7	357	67.7
Fully	788	87.0	762	69.5	310	57.8	1860	73.3	350	66.4
Partly	3	0.3	6	0.5	1	0.2	10	0.4	7	1.3

Starting from a gross sample of 3,916 children in 3,065 households (Figure 1), the net sample of children for whom information was provided amounts to 2,841 children in 2,227 SOEP households. The overall response rate thus equals 73 percent but varies across subsamples and waves (Table 1). In the course of data collection, 1,528 valid addresses of ECEC institutions attended by SOEP and FiD children could be retrieved for a total of 2,074 children. The centers were spread across the entire country (Figure 2).

Figure 2
Locations of institutions attended by children in SOEP-ECEC Quality
Kinder und Kitas

Standorte der Kitas



Note: This map is for illustrative purposes only and shows the locations of an incomplete list of centers. Source: TNS Infratest Forschung GmbH (December 2013)

Subsequently, we gathered information on the quality of these ECEC institutions from the perspective of ECEC directors and educators. Surveys were conducted via paper and pencil or telephone. 854 out of 1,528 institutions from all over Germany provided valid information. This implies a response rate of 56 percent at center level. More information on response rates can be found in chapter 4.2.

Child minders or family daycare homes were not surveyed, as they usually care for a small number of children, meaning that identification of individual children could be possible if only few child minders worked in a municipality. Within the SOEP, only a small number of children is in this type of care, which is in accordance with administrative data indicating that only about four percent of children below three years of age are cared for by registered child minders in Germany (Statistisches Bundesamt 2012).

2.2 Interview methodology and survey instruments

As part of the SOEP-ECEC Quality study, the two K²ID surveys were set up by the project team.⁵ This involved developing a set of pretested questionnaires for the main caregiver ('Hauptbetreuungsperson') in each household as well as for the director and at least one educator within each of the ECEC institutions attended by the children of the respective households. Main caregivers and ECEC institutions were interviewed by self-administered paper questionnaires first. To ensure high response rates, households received several reminders by phone and from interviewers as part of the regular SOEP survey. In case of non-response, institutions were contacted via telephone. Directors had the possibility to answer a highly compressed version combining questions from both the director and educator questionnaires in order to increase sample sizes (see below).

The applied methods were favored over personal interviews since they presumably interfered less with daily routines of ECEC staff, parents and children. Survey-based methods of collecting ECEC quality data appeared particularly advantageous in the context of the SOEP. Observations in classrooms of ECEC centers attended by SOEP children would have been very costly and inefficient, given that only very few children attended the same center. For a discussion on the collection of quality information via surveys see also McCabe and Ackerman (2007). Note that households and institutions that participated in the pretest were not supposed to take part in the main study anymore.

2.2.1 Survey instruments at the household level

The questionnaire for the main caregiver within each household ('Parent Survey') was developed primarily by drawing on instruments from various other surveys and instruments developed by the research team. Particular questions and items of these questionnaires were adapted to ensure their applicability to the wider age ranges of the SOEP target children in the ECEC institutions. Furthermore, some questions were adapted or refined to improve specific questions where the wording or answer scales might have led to missing re-

⁵ The authors would like to thank Yvonne Anders for her contributions to all questionnaires. Moreover, Yvonne Anders gave us very useful feedback on earlier versions of the questionnaire.

sponses or skewed distributions of the variables. For details on the origin and modifications of specific questions please see Table A-1 in the appendix.

To ensure the parent questionnaires' validity, a pretest including 23 questions on 8 pages was conducted in 58 out of 100 intended households taken from the representative cohort samples of the FiD study (Fall 2013). For households to be included in the pretest, they had to meet several criteria, such as participation in FiD 2013, no previous refusal to participate in 2014 and at least one child attending an ECEC institution. Households were chosen with equal proportions from East and West Germany, and large families were disproportionately represented. To incentivize parents' cooperation, small presents for their children were included in the pretest mailing. Furthermore, a reminder was sent after a couple of months, followed by a phone call to increase participation rates. The latter provided the opportunity to reveal any kind of problems with the survey instrument or other reasons for non-participation. For the majority there were no particular problems with the questionnaire.

In addition to the results of the pretest, advice was given by Hans-Günther Roßbach, University of Bamberg, and Wolfgang Tietze, Freie Universität Berlin, both of whom are national experts on early childhood education and care. In addition, Edward Melhuish, University of Oxford and Birkbeck, University of London, and Pam Sammons, University of Oxford, contributed to the survey instruments from an international perspective.⁶

On the basis of the results of the pretest and expert advice, a longer and a shorter version of the final instrument were developed and implemented in the field. While the former instrument was used for FiD households, the latter questionnaire was used for all subsamples of the SOEP, which were supposed to have a greater risk of non-response. In case of non-response, FiD households were allowed to answer the shorter version as well. Note that some of the questions are also part of the Institution Survey (see also Table A-1). In the following, the questions will be grouped and briefly described, including the respective question numbers in brackets.

Attitudes towards early education and care (Question number - Long: 1 to 2 / Short: 1)

⁶ The authors are grateful for Professor Yvonne Anders' assistance in recruiting these experts as well as several interviewers who were students in the field of Early Childhood Research for the project.

To investigate the parental decision making process with regard to use of ECEC, a first block of questions on parental attitudes on early education and care was included. The first question (Long/Short: 1) deals with general aspects of non-parental care with regard to respondents' own family and children. The respondent is asked to rate the importance of each aspect by using a 4-point-scale which ranges from 'very important' to 'not at all important'. The question and items were inspired by a SOEP question on the importance of different life domains. After rating the aspects, in the long version respondents are furthermore asked to pick the most important one.

Question 2 (Long: 2 / Short: -) includes various statements on early childhood education and care. Respondents are asked to state their opinion by using a scale ranging from 1 'I completely disagree' to 7 'I agree completely'. Both the question and items were taken from ESKOM-V (wave 1; see also Table A-1 in the appendix). Whereas the question is included in the FID samples, it is not part of the shorter version for all SOEP samples.

Attendance of ECEC center (Long: 3 to 6 / Short: 2 to 5)

Screening question 3 (Long: 3 / Short: 2) serves to identify households with children attending an ECEC center. The question was taken from the SOEP household questionnaire 2012. However, it does not refer to after school centers. Whereas households without a child in ECEC have completed the questionnaire, households with a child in an ECEC center are asked about the usual days and daily hours of attendance (Long: 4 / Short 3), as well as the years and month when their child attended the center for the first time (Long: 5 / Short: 4). Whereas the former question was taken from the 'Kommunale Bedarfserhebung' on behalf of the DJI/TU Dortmund, the latter question originally stems from ESKOM-V. Question 6 (Short: 5) investigates whether any other child in the household is attending the same establishment and, if so, whether a questionnaire has already been completed for the respective child. If this is the case, respondents can skip questions 7 to 12 on features of the institution (Short: 6 to 9). Otherwise, respondents are navigated to the next question.

Name, address and opening hours of ECEC institution (Long: 7 and 8 / Short: 6)

Both the name and address of the child's ECEC institution are asked in question 7 (Short: 6). This information is used to identify and survey the respective establishment. The question is

identical to the one used in the SOEP Pretest 2007, which showed a high response rate (see TNS Infratest Sozialforschung 2011, p. 18).

Question 8 continues with the daily opening hours of the center on most days of the week. The question is only included in the longer questionnaire for FID samples but not in the shorter version for SOEP samples. Note that the daily opening hours are also inquired in the questionnaire for the director.

Information asymmetries between mothers and ECEC providers (Long: 9 to 12 / Short: 8 to 9)

A third block of questions is included to examine quality aspects as well as information asymmetries on the respective features between parents and ECEC providers. Question 9 (Short: 8) asks if the ECEC institution has a pedagogical concept or profile and if so, whether this concept is accessible to parents via posting, homepage or handover (long questionnaire only). The question was retrieved from the NUBBEK questionnaire for pedagogic directors. The question is included in the questionnaire for directors of ECEC institutions as well.

Besides the general pedagogical work, the establishment might have one or several focus areas, such as language, music, etc. Question 10 asks about these points of focus. Notably, institutions were defined as having a focus when a specific training or concept is a relevant part of daily routine and the respective members of staff are trained. Both question and answer categories were taken from the NEPS SC2 director questionnaire 2011/2012. The question is only included in the longer version of the parent questionnaire, as well as in the survey for directors.

Question 11 revolves around opportunities for parental involvement. Opportunities include communication at drop-off or pick-up and several specific activities for parents or families. Furthermore, there is an open category for activities not listed ('other'). For each item, respondents can indicate the respective frequency which ranges from 'several times a week' to 'once a year' or 'not offered'. Questionnaires from both ESKOM-V and BIKS-3-8 served as inspirations for the question. It is included in the questionnaire for directors but not in the shorter version of the parent questionnaire.

Question 12 (Short: 9) asks to what extent parents' suggestions and wishes with regard to different matters such as opening times, meals, equipment and activities are taken into consideration by the center. Respondents can select values on a scale ranging from 0 'not at all' to 10 'to a large extent'. To investigate any information asymmetries, the question is included in the questionnaire for the director of the center, too. Also, parents can choose "don't know".

Individual criteria for choosing ECEC center (Long: 13 / Short: -)

With regard to the decision process of choosing an ECEC institution, question 13 asks for the respondents' five most important aspects influencing the decision back at the time. Individual criteria should be ranked in order of importance: 1 for the most important reason, up to a maximum of 5. Respondents who did not have a choice, e.g. due to limited availability of ECEC places, and therefore cannot not rank important aspects, are supposed to mark an additional box. The question is only part of the longer questionnaire for FiD samples.

Satisfaction with educational support and care in center (Long: 14 / Short: 10)

Question 14 (Short: 10) comprises an extensive item battery on parental satisfaction with diverse aspects of education and care of the child in the actual institution (organization and structure, personnel, facilities and equipment, care provided to children, domain-specific activities, cooperation with parents). Respondents can choose values on a scale from 0 'completely dissatisfied' to 10 'completely satisfied'. Question and items are based on a FiD parent questionnaire and an ESKOM-V questionnaire (wave 2). In addition to the main caregiver, both the director and the pedagogical staff of the ECEC establishment receive the same question to analyze potential information asymmetries between parents and staff.

Participation in center activities (Long: 15 / Short: 11)

The frequency with which the child participates in a range of activities in the center is inquired in question 15 (Short: 11). Activities include offers such as early musical education, early language support, outdoor activities and excursions. Respondents can also indicate how often (i.e., never, sometimes or always) participation in the respective activity involves additional costs. To examine potential information asymmetries, similar questions are included in the questionnaires for both the director and the group educator.

Child's group within the institution (Long 16 to 19 / Short 12 to 15)

Question 16 (Short: 12) helps identifying the specific group in the ECEC center the child attends. Respondents are asked to state the exact name of the group, and to contact the center in case of uncertainty. Alternatively, parents can indicate that children are not allocated to different groups. Furthermore, screening question 17 (Short: 13) investigates whether any other child in the household attends the same group. If this is the case, and if respondents have already completed the questionnaire for the respective child, they may skip questions 18 to 20 regarding features of the group (Short: 14 to 16).

The number of children in the group aged below three years and between three and school age is inquired in question 18 (Short: 14). Moreover, in question 19 (Short: 15), respondents can estimate the number of children who also speak a language other than German at home. For these questions, surveys from the NEPS, SOEP and FiD served as starting points.

Caregivers in institution (Long: 20 to 21 / Short: 16)

Question 20 (Short: 16) inquires the general number of educators responsible for the group. The question further distinguishes between the number of educators who are usually present at the same time, and the number of usually present educators who have not completed their training (yet), including trainees, interns and volunteers. In question 21, respondents are asked to indicate the frequency of staff turnover (i.e., recruiting or withdrawal) involving at least one educator in the child's group. The question is only included in the longer version of the questionnaire. Both questions are also part of the educator questionnaire to allow for analyzing any information asymmetries between parents and ECEC providers in this regard.

Activities in group (Long: 22 / Short: -)

The frequency of specific group activities is the subject of question 22. For each item, respondents indicate whether the respective activity is performed daily, a few times a week, once a week, every two weeks, once a month, less often, or never. Question and answer categories were taken from the AKFRA (Sprache Brandenburg) questionnaire. The question is incorporated in the long parent and the educator questionnaire.

Parental perceptions of ECEC institution and staff (Long: 23 to 25 / Short: 17)

Question 23 requires respondents to rate their perceptions concerning the enthusiasm of their child's group educator or director at work. In addition, question 24 asks about respondents' impression as to whether children's interests are taken into consideration in specific pedagogical activities. For both questions, response scales range from 0 'not at all' to 10 'very much'. They are not part of the shorter version of the questionnaire.

Question 25 (Short: 17) asks about parents' general satisfaction with their child's ECEC institution. Once again, an 11-point response scale from 0 to 10 is used. Both question and scale were taken from the FiD parent questionnaire 2010. Comparable questions are included in the questionnaires for directors and group educators.

2.2.2 Survey instruments at the institution level

The two questionnaires for ECEC institutions, one for directors and one for the pedagogic staff of the ECEC centers, are predominantly based on survey instruments developed by the National Education Panel Study (NEPS), more precisely on the Starting Cohort 2 director and educator questionnaires (wave 1), even though the questions were often modified to varying degrees. Some questions originate from other surveys or were developed by members of the research team. Again, many questions needed to be adapted or refined in terms of wording, response scales, etc. in order to avoid partial nonresponse or skewed distributions of the variables. Table A-2 provides more information on the sources and modifications of questions included in the Institution Survey.

Both instruments were tested in a pilot sample of 20 out of 56 intended ECEC centers, consisting of the ECEC centers whose addresses had been reported in the pretest of the Parent Survey. Moreover, pedagogical experts gave advice once again. Pam Sammons, Edward Melhuish, Hans-Günther Roßbach and Wolfgang Tietze provided detailed comments and suggestions for developing the questionnaire-based survey instruments for measuring ECEC quality in German ECEC centers. Based on these inputs, questionnaires were finalized.

The two questionnaires designed for ECEC institutions partly contain overlapping questions for the directors and educators (see Table A-2). The questionnaire for directors focuses on aspects of structural quality, pedagogic orientations, and satisfaction with various education and care aspects. The questionnaire for the educators of the specific groups attended by

SOEP children concentrates on surveying the frequency of various activities performed with the children, in addition to pedagogic orientations of the staff and satisfaction with the quality of education and care provided. Again, the questions will be categorized and summarized in the next section.

Note that in order to further ensure the validity of the applied self-reported measures of ECEC quality, a small-scale pedagogical study was conducted by Yvonne Anders and Axinja Hachfeld. The study compared the survey-based measures of ECEC quality developed in the SOEP-ECEC Quality study with assessments of professional observers. To this end, 29 ECEC centers outside the SOEP-ECEC Quality sample were assessed using the KES (Tietze et al. 2007). A separate report on this validation study is available (Hachfeld and Anders 2017).

2.2.2.1 Director questionnaire

a) GENERAL QUESTIONS ABOUT CENTER

Organizational aspects of center (1 to 3)

Questions 1 to 3 are designed to gather organizational information about the ECEC institution. Question 1 captures whether the provider is the municipality, a church, a private initiative or of another form. Question 2 captures the daily opening hours and question 3 indicates whether the ECEC institution provides lunch to children.

Attending children (4 to 9)

This section is about the composition of attending children as well as the selection process. Question 4 gives the overall capacity of the ECEC institution. Question 5 indicates the number of attending children overall as well as the decomposed numbers of attending girls and boys, children with non-German language of origin and children with handicaps. Question 6 gathers disaggregated information about the age groups of children in the ECEC institution and differentiates by daily care hours. In question 7 directors can indicate whether discounts or exempt from parental fees for low income parents are possible and how many children benefit from this. This question is newly developed and is supposed to hint at the socio-economic composition of children in terms of family income.

Questions 8 and 9 address future expectations of attendance numbers and selection criteria. In question 8, the number of available places for the next year as well as the number of chil-

dren on the waiting list (if available) is inquired as a measure of demand. In question 9 directors are asked to rank the criteria (from 1 to 9) based on which they admit children in case the demand exceeds available places. Possible criteria include, among others, age, older siblings in the institution, place on the waiting list and a personal interview with parents and children. This question was supposed to shed light on supply-side decision-making and to give an indication how centers choose children.

Group structure (10 to 11)

These two questions briefly sum up the work structure of the ECEC institution. Question 10 inquires whether children are assigned to fixed groups and whether there is any interaction between groups. If there are indeed different groups, question 11 captures the different kinds of groups (referring to age span) and their respective numbers. Otherwise, directors may skip this question.

Parental involvement and quality development (12 to 14)

The last questions of this section explore how parents are included in the work of the ECEC institution and whether quality is evaluated and developed with external actors. Question 12 asks for possibility and frequency of conversation with parents or parent consultation as well as other forms of interaction. Question 13 informs about how far the wishes and needs of parents regarding opening hours, pedagogical concept and others are incorporated within the ECEC institution. Question 14 reports whether the institution has taken part in any form of evaluation and development measure and if yes, whether this measure was an internal one or conducted by external providers.

b) QUESTIONS ABOUT BUILDING AND PEDAGOGICAL WORK

Facilities and local conditions (15 to 18)

The questions in this section inform about the size and the conditions of the center. Question 15 states the overall size of interior space of the ECEC institution in square meters. Question 16 asks whether there is a garden included and how big this is in square meters. Question 17 asks about specific rooms for promotion of child learning and development. Question 18 then explores perceived pressures to enhance quality due to local competition

among local ECEC institutions, difficulties in finding qualified staff and possible threats to the center's survival owing to low attendance numbers.

Pedagogical work (19 to 27)

Questions 19 to 27 provide details about the pedagogical work of the ECEC institution. Question 19 asks about the general pedagogical approach. Question 20 asks whether a pedagogical concept is incorporated in the mailing, whether and where it is available online, whether it will be sent later on or whether no pedagogical concept exists at all. If a pedagogical concept exists, in question 21 directors are supposed to indicate who participated in its development. Question 22 explores whether the institution has any pedagogical focus beyond the general pedagogical concept, for instance a focus on music, language development, mathematics, religious education and so forth.

Question 23 asks if there are any further written plans for the pedagogical work and question 24 captures how often there are regular meetings for the pedagogical staff. The content of these meetings is further explored in question 25, which asks how often certain aspects of the pedagogical work, such as general questions, the concept of the ECEC institutions, and excursions were on the agenda of the team meetings in the last six months.

Question 26 lists diverse activities a center might offer to children of varying age groups in order to ask whether such activities are offered and if so, how often they require additional financial contributions by parents. Question 27 then asks if there are special courses to support staff members working with children and parents whose language of origin is not German.

c) ASSESSMENTS OF CENTER AND PEDAGOGICAL WORK

Satisfaction with different aspects of center (28 to 29)

These two questions directly address the satisfaction of the ECEC director. In question 28, respondents can indicate on a scale from 0 to 10 their satisfaction with different aspects of organization and structure of the institution, caregivers, facilities and equipment, care provided to children, domain-specific activities and cooperation with parents. The question is designed to be comparable with replies provided by the group educators and parents. Finally, question 29 captures the overall satisfaction with the institution.

Perceived responsibilities and satisfaction with other aspects (30 to 31)

Question 30 inquires respondents' opinion about the responsibilities of the institution relative to the family. On a 7-point response scale, directors can indicate whether they think that different educational tasks should be fulfilled by the ECEC institutions, the children's families, or both. Such tasks relate for instance to the behavior within the group, motivation for sports, nature or words and poems. Question 31 then asks about the directors' personal satisfaction with health, work and income.

d) QUESTIONS ABOUT STAFF MEMBERS

Personnel resources (32 to 33)

The first two questions require detailed figures about the employed staff. Question 32 asks about the decomposed numbers in terms of all staff members' qualifications, part- or full-time work and weekly hours. Question 33 then captures the overall number of pedagogical staff and how many of them are of non-German language origin. Question 33 helps to distinguish staff members who work with children from those who do not (e.g., cooks).

e) QUESTIONS ABOUT DIRECTOR AND DIRECTOR'S WORK

Personal questions (34 to 45)

The questions in this section capture the director's personal information like gender, birth year, schooling, type of professional qualification and specialization, work and leadership experience, weekly working hours (in contract and actual hours) and whether he or she participated in further training during the last year. Question 45 then asks about the respondent's overall life satisfaction on a scale from 0 to 10, just as in person interviews of the regular SOEP.

Checks (46 to 48)

The last three questions are intended as checks. Question 46 asks about the exact position of the person who filled out the survey. Question 47 captures the date and question 48 asks whether the respondent would like to get informed about first results of the study, which was intended as an incentive to participate.

2.2.2.2 (Group) Educator questionnaire

a) QUESTIONS ABOUT GROUP COMPOSITION

Group structure and attending children (1 to 5)

Question 1 is designed as a filter. If there is no work in fixed groups, the next two questions are supposed to be left out.

Questions 2 and 3 explore the composition of the respective group. While question 2 differentiates between all children, girls, boys, children with a non-German language of origin and children with handicaps, question 3 focuses on the age structure and children's hours of attendance per day. This enables researchers for instance to calculate standardized child-teacher-ratios. Question 4 asks about the children's language development. Educators are supposed to estimate the number of children with average, above- and below-average language skills, differentiating between all children and those whose language of origin is not German. Question 5 captures the number of children who participate in special language training.

Staff and space (6 to 9)

Question 6 asks about the overall number of educators in the group, the number of educators usually present at the same time and of those without completed training. In case there are fixed groups in the ECEC institution, question 7 inquires the working hours of the caregivers and question 8 about the available space reserved only for the group (number of rooms and square meters). Question 9 then asks all survey participants about the fluctuation of staff within their group.

b) QUESTIONS ABOUT EQUIPMENT AND ACTIVITIES

Materials (10)

Question 10 deals with the equipment of the group. For several materials like books, musical instruments, toys and so forth, the respondent is asked to indicate whether there is sufficient equipment to let some, half, or almost all of the children use it at the same time or whether this piece of equipment is not available. In addition, the educator can mark those materials which in their view are not available in sufficient numbers.

Activities (11 to 12)

Questions 11 and 12 ask about frequencies of certain activities. While question 11 tells us how often common activities like storytelling, singing, language and letter games actually take place, question 12 focuses on whether and how often more extraordinary and possibly fee-based activities (e.g., early musical education, foreign language training, excursions) are offered. This question also asks how often parents are required to pay extra fees for their offspring's participation.

c) EDUCATOR'S ATTITUDES

Orientation quality (13 to 15)

These three questions captures educators' attitudes towards the role of the institution versus the family, their personal role as well as educational goals. In question 13 educators should indicate whether they think that certain aspects of education should be covered primarily by the ECEC institutions or the family. This question is identical to question 30 in the director questionnaire. Question 14 explores to what extent the educator sees him- or herself as a friend, partner, teacher, advisor, authority, learner or expert towards the children. Question 15 asks how important different goals are in children's education. This question is based on a question from the SOEP mother and child questionnaire for 7 to 8-year-olds.

d) QUESTIONS ABOUT PEDAGOGICAL WORK

Perceptions about pedagogical work (16 to 24)

This set of questions is to reflect the role as an educator and interactions within the institution. Question 16 asks to what extent the educational plans of the federal states influence and assist in the daily education work. Question 17 includes five items from the Utrecht Work Engagement Scale. Educators are asked to indicate how often they perceive their work as useful, inspiring, challenging, etc., whereas question 18 captures perceived rush or time pressure.

In question 19, educators are asked to spontaneously report an activity or project from the previous working day. The next two questions are follow-ups and ask why this activity was chosen and whether it was documented. Question 22 asks if the educator enjoys integrating certain subjects into their work more than others.

Question 23 informs about the family language of the respondent. If he or she learned a language other than German as a child, question 24 asks how often this language is used when working with children in the group.

e) ASSESSMENTS OF GROUP WORK AND CENTER

Satisfaction with different aspects of center (25 to 26)

The next two questions explore the educator's satisfaction with the group work and the ECEC institution, which can be compared with the director's and the parent's replies. In question 25, the respondent can indicate on a scale from 0 to 10 the satisfaction with many aspects of organization, equipment, activities and so on. Question 26 evaluates the overall satisfaction with the ECEC institution and the specific group (if applicable).

f) QUESTIONS ABOUT EDUCATOR AND EDUCATOR'S WORK

Satisfaction with other aspects (27)

Question 27 asks about educators' satisfaction with health and personal income.

Personal questions (28 to 38)

Questions 28 to 38 inform about personal characteristics of the group educator such as gender, birth year, qualifications, work experience, contractual and actual weekly working hours and participation in further training in the past year. Question 38 addresses educators' overall life satisfaction.

Checks (39-40)

The last two questions are intended as checks. Question 39 asks about the professional position of the person who filled out the survey, and question 40 captures the month and year of the interview.

2.2.2.3 Compressed questionnaire

The compressed version was designed to promote participation of institutions whose staff did not have time to take part in the regular survey. The aim was to reduce the survey to four pages of questions for directors which can be answered on the phone. The selected questions were taken from both the questionnaire for directors and educators, covering diverse aspects of care in the institution (Table A-2 in the appendix shows which questions in

the compressed version correspond to questions in the director and educator questionnaire, respectively). The compressed version gathers information on structural quality and composition at center and group level, on admission procedures, directors' orientations and satisfaction, pedagogical work and offered group activities, groups' personnel resources, personal qualification, experience and professional role.

Questions 1 to 8 relate to the questionnaire for directors. Question 1 captures the number of children overall, including age structure, and with non-German language of origin in the center. In question 2, respondents can name the most important criterion in the admission process and add further criteria. The third question asks about the relative responsibility of the center versus the family in terms of fostering diverse competences and behaviors among children. Question 4 is a short version of the question about directors' satisfaction with diverse aspects in the institution. It is restricted to aspects of organization and structure as well as caregivers. Question 5 refers to the overall satisfaction with the ECEC institution. In question 6, respondents can indicate if a special pedagogical focus exists. Following this, question 7 asks whether or not the pedagogical concept is available online, or if such a concept does not exist at all.

Question 8 inquires whether there are fixed groups within the ECEC institution. If so, question 9 asks about the size and composition of the group, differentiating by age categories, daily hours of attendance and language background. Question 10 asks about the numbers of educators in the group, their presence and rough qualification. Question 11 collects information about the frequency of offered activities in the group.

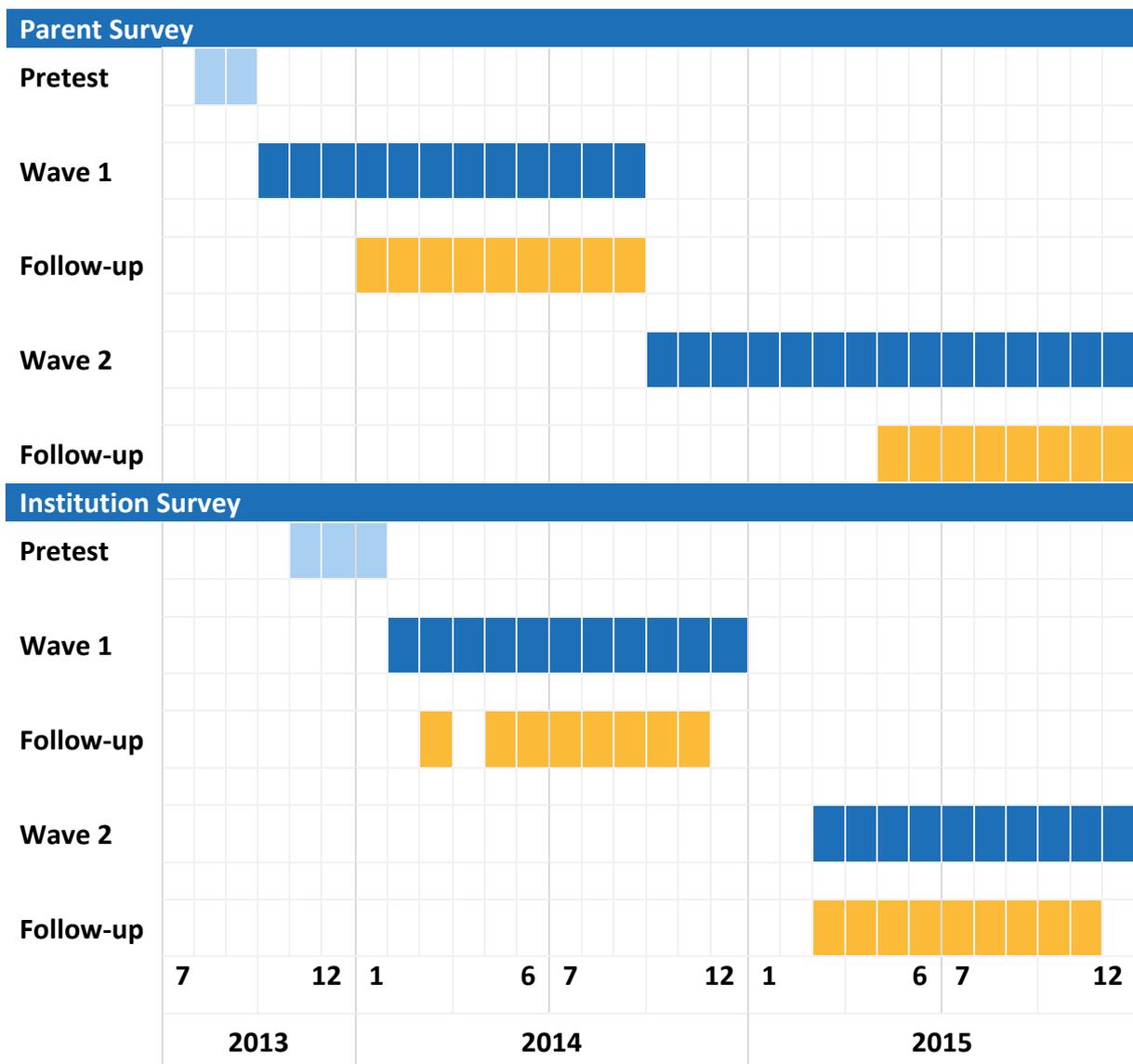
The last three questions are personal questions: Question 12 inquires the highest obtained professional degree, question 13 asks about the years of experience as a director and question 14 verifies the professional position of the respondent.

3 Survey preparation and implementation

The implementation of both surveys consisted of three core steps. First, the questionnaires for the Parent and Institution Survey were designed and tested. Second, the households and ECEC institutions were surveyed. Third, the periods of data collection were partially accom-

panied by follow-ups in order to raise participation rates. The entire period of data collection lasted from August 2013 until December 2015. For an overview of the timing of the different surveys and waves see Figure 3.

Figure 3
Timeline of data collection for SOEP-ECEC Quality



Parent Survey

Preparation of the Parent Survey started in summer 2013. After receiving expert advice and conducting the pretest in August and September 2013, the research team finalized the instruments. The period from end of October 2013 until January 2014 marks the main phase of wave 1 during which most parents (86 percent) answered the survey. The timing of the par-

ent survey was chosen so that both the Parent and Institution Survey could be completed for most children within one school year, that is, between September and June of the following year. The aim was to ascertain that information provided by parents and ECEC staff on the childcare setting was comparable, given that changes between institutions as well as in group compositions are more likely at the start of a new school year.

All households with children below school age received a separate mailing. The mailing included a self-administered paper questionnaire, asking the main caregivers, among other things, for the address of the ECEC institution and the name of the respective ECEC group their child attended, as well as for their consent to contact the institution.⁷ The mailing also contained a personalized letter, a leaflet with more detailed information on the study, a data protection sheet and a return envelope. To incentivize parents' participation in the study further, they received little toys for their children. Whereas FiD households received the longer version of the questionnaire, all SOEP subsamples were asked to fill in a questionnaire of only about half the length to diminish their rate of non-response. In addition, several follow-up measures helped to increase participation rates. In January and February 2014, TNS Infratest called FiD households and offered them to answer the most central questions on the phone and/or to send them the shorter questionnaire intended for SOEP members. 158 FiD households chose one of these options, with parents of 106 children only replying via CATI and parents of the remaining 102 children answering the short version. By contrast, SOEP members could answer the questionnaire as part of the regular SOEP interviews between February and September 2014.

All questionnaires returned up to April 2014 formed the basis for the subsequent ECEC Institution Survey. Those addresses reported in parent questionnaires which came in after April 2014 were processed in the course of wave 2 of the Institution Survey.

The main phase of wave 2 of the Parent Survey started in October 2014 and ended in February 2015. These were respondents from sample M (see above). Until then, 51 percent of surveys had been returned. No long questionnaires were sent out in wave 2 given that the sample comprised SOEP members only. The documents sent to sample M members in wave

⁷ A question regarding the name and address of the ECEC center had already been tested in a previous SOEP pretest in 2007 with a high response rate of about 90 percent (TNS Infratest Sozialforschung (2011)).

1 were very similar to those from wave 1. However, depending on the household, the mailing also included a translation of the letter to English, Turkish, or Russian. Moreover, vouchers were offered in case of participation. Again, households had the option to provide answers during the regular SOEP interviews between May and December 2015. Note that ECEC addresses reported in parent questionnaires coming in after September 2015 could not be processed anymore.

Institution Survey

The two questionnaires for ECEC institutions were developed in September and October 2013. Based on expert advice and the results from the ECEC institution pretest, the research team finalized the questionnaires in spring 2014. Once the center addresses were available and checked, TNS Infratest started the second phase of field work lasting from end of February until December 2014. The questionnaires were mailed to the directors of all ECEC institutions who were asked to complete one questionnaire and forward the other questionnaire(s) to an educator of each group serving a SOEP child, preferably the educator bearing the main responsibility for the class. In case the name of the child's group was unknown, the director could freely choose an educator. If the center did not assign children to different groups, the director was asked to hand the educator questionnaire over to an educator directly working with children. If this was true for the director, he or she could answer both questionnaires. If the institution was only willing to answer the compressed questionnaire but served several sample children in more than one group, the director was asked to refer to just one of these groups, which was randomly chosen by TNS Infratest.

Next to the questionnaires, the centers also received a letter, a leaflet with details on the study, a data protection sheet and a prepaid return envelope via mail. To ensure high response rates, all ECEC institutions received vouchers upon completion of each questionnaire, plus an additional voucher if all of the required questionnaires were returned. Also, directors could indicate if they would like to receive an information sheet with selected results at the end of the study. As part of the follow-up phase, TNS Infratest contacted ECEC institutions without response in March 2014 via telephone. Those who stated general interest in the survey when called but who did not send the questionnaires back received further reminders by phone starting in May. This time, we recruited pedagogy students with work experi-

ence in ECEC-related institutions or similar previous surveys from the Freie Universität Berlin to continuously follow up the responses of the institutions. The aim was to enable the student interviewers to build up a relationship with ECEC directors of their assigned institutions. The students were trained by the K²ID research Team and TNS Infratest. As of July 2014, a compressed version of the questionnaire was additionally deployed. First, its use was restricted to a postal non-response survey taking place between July and September among centers that had already refused participation. As of September, however, the student group could also use the compressed version in their telephone interviews with directors who were unwilling to answer the regular questionnaires. Only in few cases, the compressed questionnaire was answered on paper. Note that even at this late stage of data collection, responses to the regular, substantially longer version of the questionnaire came in very sporadically. As a further measure to boost participation, centers could ask for assistance with obtaining permission from their providers to take part in the survey. Both the research team and TNS Infratest provided telephone based support, and TNS Infratest contacted some providers directly.

The main phase of wave 2 of the Institution Survey spanned the period from March to December 2015. Data collection for wave 2 was conducted in a similar way as for wave 1. A major difference was that a call by a staff member was already announced in the letter sent to institutions together with the questionnaires. Correspondingly, one to two weeks after receiving the mailed documents, the student interviewers called each ECEC institution in order to clarify issues and discuss the course of the study. From the beginning, the students could offer directors support with obtaining permission from their providers and could make use of the compressed questionnaire. As a final remark on the development of response rates over time, a salient difference between the Parent and the Institution Survey became evident: Whereas the Parent Survey revealed clear peak months, institution questionnaires were returned fairly gradually over the entire field time (Bohlender and Glemser 2016).

After TNS Infratest had handed over the raw data in June and November 2014 (wave 1) as well as in January 2016 (wave 2), the research team processed, checked and documented the data in line with SOEP standards. This included renaming and labelling variables, creating further variables to improve handling of the data for users, and recoding missing values (see

chapter 5.1). Several data sets were created which can be merged (see chapter 5.2). The data were checked for inconsistencies as well as non-random missingness. In terms of the latter, statistical weights were created taking into account a range of variables predicting non-response of both households and ECEC centers (see chapter 5.3).

4 Pretest results and response rates

4.1 Pretest results

Parent Survey

The pretest of the household questionnaires was conducted in August 2013. The 8-page questionnaire with 23 questions, including the address of the ECEC institution, was answered for 117 children in 58 out of intended 100 FiD households. The gross sample was not drawn randomly, so that the share of families with two or more children was comparatively high. In addition, within each family at least one child was attending an ECEC institution.

Overall, respondents did not encounter severe difficulties when answering the pretest and most of them filled in the questionnaire correctly. Still, based on descriptive analyses of the pretest results multiple changes were made. In order to avoid misunderstandings, design and formulations of specific questions and attached instructions needed to be altered. Some questions were optimized by modifying their response scales. Questions lacking variation were deleted from the survey, making room for some additional questions asking about parental attitudes, children's days of attendance, centers' pedagogical concept, focus and activities as well as aspects of parental satisfaction.

Institution Survey

For the Institution Survey, the regular versions of the director and the educator questionnaire were pretested. The former consisted of up to 51 questions and the latter of up to 42 questions. Each questionnaire contained 24 pages. TNS Infratest contacted 56 ECEC institutions with 66 groups attended by 84 FiD children who were part of the pretest of the Parent Survey. Overall, 20 centers (i.e., 36 percent) participated, and most of them completed all of the questionnaires required. Furthermore, the study team carried out cognitive testing with three institution directors to explore potential issues with any of the questions.

No major problems occurred with regard to answering the pretest questions, most of which had already been included in previous studies. One aim was nonetheless to substantially shorten the page numbers of both questionnaires and to reduce the number of questions somewhat. Also, following the cognitive testing and an analysis of the pretest results several questions were further clarified, simplified (e.g., by making adjustments to design, wording or item lists), and in some cases multiple questions were combined to one. This served to increase comprehensibility and was supposed to decrease the time required to answer these questions. Some changes were moreover made to the order in which the questions occurred. In both the director and educator questionnaire questions were added concerning the date of completion, the respondent's role and a question as to whether the respondent's professional training had a focus on early education. Finally, two further questions included in the educator questionnaire referred to staff's educational goals and the frequency with which they felt rushed or under time pressure within the past four weeks.

4.2 Response rates of the main study

Sample sizes and response-rates

The net sample of the K²ID Parent Survey amounts to 2,841 children in 2,227 households. Starting from a gross sample of 3,916 children in 3,065 households, this equates to a response rate of about 73 percent (Table 2). Note that while the response rate was only 68 percent for children from sample M, it was nearly 74 percent for the other subsamples together. Focusing on differences within this latter group, the response rate was substantially higher for SOEP (87 percent) than FiD subsamples (cohort samples: 70 percent; screening samples: 58 percent). This is likely due to several reasons, including the fact that SOEP respondents were handed over a shorter questionnaire, and that participation was reinforced more strongly through a visit by an interviewer during their annual SOEP interview. Also, the FiD respondents entered the panel relatively recently, possibly resulting in lower commitment to the study. Overall, 42 percent of all answered parent questionnaires were the long version, 54 percent were the short version and nearly 4 percent of questionnaires were answered via CATI as part of the follow-up for FiD members (Table 2).

Parents were willing to reveal the address for 2,074 of 2,233 children attending an ECEC institution, which is 93 percent. Note that addresses for 131 children were provided after April 2014. Therefore, processing had to be postponed by one year (Table 2). Addresses of 45 children which were delivered late in wave 2 were not processed further. After correcting for duplicates and deleting addresses of schools and child minders, this translates to 1,528 unique center addresses included in the subsequent Institution survey (Table 2). Note that in some households several children visited one institution.

Out of these 1,528 unique ECEC institutions, 854 (56 percent) participated in the Institution Survey (Table 3). Response rates were somewhat higher for institutions in wave 2 of the Institution Survey (2015), which served children from sample M as well as from other sub-samples if parents answered too late for the institution address to be included in wave 1. In order to achieve these response rates, institutions' directors were repeatedly reminded of the study via phone. Moreover, directors had the option to answer a very compressed version of the questionnaires consisting of only four pages, usually on the phone. Indeed, 200 centers (23 percent) exclusively chose this option (Table 3). As opposed to this, all questionnaires were fully completed for 555 centers (65 percent).

Table 2
Sample and results of Parent Survey by wave

	Wave 1 Parent Survey		Wave 2 Parent Survey		Total	
	N	%	N	%	N	%
Households						
Gross sample	2538		527		3065	
At least one questionnaire answered	1870	73.7%	357	67.7%	2227	72.7%
Children						
Gross sample	3240		676		3916	
Questionnaire answered	2388	73.7%	453	67.0%	2841	72.5%
Long questionnaire	1201	50.3%	0	0.0%	1201	42.3%
Short questionnaire	1081	45.3%	453	100.0%	1534	54.0%
CATI	106	4.4%	0	0.0%	106	3.7%
Child visits ECEC institution	1924		309		2233	
Address reported	1788	92.9%	286	92.6%	2074	92.9%
Not processed (late reporting)	131		45		176	
Processed from previous wave	-		131		131	
Processed	1657		372		2029	
	Wave 1 Institution Survey		Wave 2 Institution Survey		Total	
Institutions						
Addresses reported on time	1275		292		1567	
School/ childminder	5		3		8	
Double / not eligible ('QNA')	26		5		31	
Adjusted gross sample	1244		284		1528	

Table 3

Sample and results of Institution Survey by wave

	Wave 1 Institution Survey		Wave 2 Institution Survey		Total	
	N	%	N	%	N	%
Institutions						
Gross sample	1244	100.0%	284	100.0%	1528	100.0%
Non-response	564	45.3%	110	38.7%	674	44.1%
Total response	680	54.7%	174	61.3%	854	55.9%
Fully realized	445	65.4%	110	63.2%	555	65.0%
Partially realized	73	10.7%	26	14.9%	99	11.6%
Compressed questionnaire	162	23.8%	38	21.8%	200	23.4%

Table 4

Questionnaire versions and levels of information by wave (Institution Survey)

	Wave 1 Institution Survey		Wave 2 Institution Survey		Total	
	N	%	N	%	N	%
Institution data from...	680	100.0%	174	100.0%	854	100.0%
Regular questionnaire	497	73.1%	128	73.6%	625	73.2%
Compressed questionnaire	162	23.8%	38	21.8%	200	23.4%
Missing	21	3.1%	8	4.6%	29	3.4%
Group data from...	788	100.0%	201	100.0%	989	100.0%
Regular questionnaire	526	66.8%	134	66.7%	660	66.7%
Compressed questionnaire	183	23.2%	38	18.9%	221	22.3%
Missing	79	10.0%	29	14.4%	108	10.9%
Combined data for all groups	788	100.0%	201	100.0%	989	100.0%
Institution and group	667	84.6%	162	80.6%	829	83.8%
Institution only	100	12.7%	29	14.4%	129	13.0%
Group only	21	2.7%	10	5.0%	31	3.1%

Going into greater detail, Table 4 shows which questionnaires were answered at both institution and group level. The regular director questionnaire was filled in for 625 institutions (73 percent). By contrast, for 200 institutions center data was provided using the compressed questionnaire (23 percent). For the remaining 29 institutions, directors refused to answer either questionnaire version.

Moreover, participating staff answered 660 regular educator questionnaires (67 percent). For 108 groups no group data is available at all. Although 221 groups belong to a center whose director answered the compressed questionnaire (Table 4), please note that valid group information is available for only 200 of the 221 ECEC groups in total. For the remaining groups, the available group information refers to another group. This applies to those centers which served several sample children attending *different* groups, since directors who answered the compressed questionnaire could report on one group only.

The K²ID data set provides information on both institution and group level for 829 out of 989 different ECEC groups (84 percent), while in most of the remaining cases only institutional data is available (13 percent) (Table 4).

Considering respondents' professional role, it becomes evident that in roughly 90 percent of cases the respondent to the director questionnaire was indeed the head of the center. If not, he or she was for the most part the deputy head. Out of all groups with valid data on the respondent's role from the educator questionnaire, about 75 percent of answers were provided by the head of the group. The remaining respondents were mostly the deputy, another staff member or the center's director. If only the compressed version of the questionnaire was used to provide group information, this was primarily done by the head of the center as well.

Data on 853 institutions could be matched with Parent Survey data referring to 1,082 sample children (Table 5).⁸ This equates to nearly 49 percent of all children attending ECEC institutions (N=2,233). Roughly half of them originate from the FiD study (53 percent). 79 percent of the institutions which could be matched with child data are attended by just one child, while 17 percent are attended by two children in the sample. The remaining centers serve between three and six children. The sample children can be assigned to 996 unique groups.⁹ Less than eight percent of these care for more than one child.

As displayed in Table 5, for 682 children (63 percent) the regular director as well as the regular educator questionnaires are available. For 115 only the former was answered, while merely the latter was handed in for 32 children. Data from the compressed questionnaire is available for 249 children (23 percent). The table also illustrates that for the vast majority of children, ECEC information was gathered during the intended wave. Of the 944 children whose parents participated in wave 1, ECEC information was provided for 869 children in the

⁸ One out of 854 participating institutions could not be matched.

⁹ This number slightly exceeds the number of groups in Table 4 because a few children could coincidentally be matched with data on their institutions via the institution identifier, even though their ECEC group was not originally part of the gross sample. This applied to children i) whose parents replied late in wave 1, but the center was nonetheless in the institution sample due to another child attending a different group in the same institution; ii) who were part of wave 2 of the parent survey, but their center (not group) had already been part of the Institution Survey 2014, so it was not contacted again. In addition, for two sample children neither the director nor the group educator participated. However, the educator questionnaire was filled out for another group in the same center.

course of 2014. Data collection had to be postponed in eight percent of cases due to late reporting of the center address. Regarding children from SOEP M, virtually all ECEC data were collected in 2015. Only three children attended centers which had already participated in the previous wave of the Institution Survey.

Table 5

Net sample of children in SOEP-ECEC Quality by wave

	Wave 1 Parent Survey		Wave 2 Parent Survey		Total	
	N	%	N	%	N	%
Households	808	87.3%	118	12.7%	926	100.0%
Children	944	87.2%	138	12.8%	1082	100.0%
Both questionnaires	601	63.7%	81	58.7%	682	63.0%
Director questionnaire only	91	9.6%	24	17.4%	115	10.6%
Compressed version only	221	23.4%	28	20.3%	249	23.0%
Educator questionnaire only	28	3.0%	4	2.9%	32	3.0%
Educator questionnaire and compressed version	1	0.1%	1	0.7%	2	0.2%
Information on other group	2	0.2%	0	0.0%	2	0.2%
Wave of Institution Survey						
Wave 1	869	92.1%	3	2.2%	872	80.6%
Wave 2	75	7.9%	135	97.8%	210	19.4%

Item non-response und filters

Generally, the quality of the data is high. However, some questions reveal elevated rates of item non-response. These include questions about compositions in all questionnaires, i.e., those questions requiring detailed information on numbers of children, staff members or numbers of groups with specific characteristics (e.g., age, hours of attendance, language skills, educational degrees). Users must decide whether and in which cases they want to interpret a missing value of -1 as zero, and how to deal with partial inconsistencies across answers. A substantial share of parents also skipped question 15 on the frequency of offered

activities, in particular the part on additional costs, while ECEC directors and educators seemed to have difficulties indicating the amount of space in m².

On the other hand, respondents sometimes answered questions they were supposed to skip. This is due to the fact that the surveys were mostly conducted via paper and pencil. Users should decide whether to overwrite or keep this information. Some of the filters were intended to avoid redundancies and to save respondents time. This mainly applied to parents with more than one child in the same ECEC center and group, respectively, and to centers that did not assign children to different groups. Also, if the educator questionnaire was answered by the director instead of an educator, he/she could skip questions on personal information and satisfaction which were part of the director questionnaire as well. The majority of respondents, however, complied with the filters, in which case values were set to -2. Users can replace these with data provided for another child or provided in the director questionnaire.

5 Data structure and weighting

The following chapter describes the data structure more carefully and offers guidance in terms of the procedures how the various data sets can be analyzed jointly. It ends with a description of the weights which have been created by the K²ID research team.

5.1 Variables

Unique identifiers: The K²ID data set includes several variables which allow researchers to distinguish between different units of analysis and to match different data sets correctly. These variables are listed and described in the following:

Unique identifier	Description	Relation to identifiers in SOEP/ FiD
hhnrakt	Current Household ID	Equivalent to hhnrakt in SOEP/ FiD
persnr	Person ID (child)	Equivalent to persnr in SOEP/ FiD
parid	Parent ID ('main caregiver')	Equivalent to persnr in SOEP/ FiD
kitaid	Institution ID	New identifier
kigrid	Group ID	New identifier

Variable names: All variables that originate from the different K²ID surveys start with the prefix 'kita'. Since participants were not asked the same questions in multiple years, there is no need to differentiate variables further by wave. In line with SOEP standards, the prefix `kita` is followed by a letter indicating the unit of analysis. The letter 'k' is used for all child data that were gathered in the Parent Survey, which corresponds to the 'kind files' in the SOEP. The letter 'd' marks all variables that are based on questions intended for directors, while all variables referring to questions intended for educators are assigned the letter 'e'. The last part of the variable name indicates the original questionnaire number. The numbers refer to the long parent questionnaire (wave 1), the regular director questionnaire (wave 2) and the regular educator questionnaire (wave 1). Variable names thus have at least seven digits (e.g., `kitak03`). If a question leads to more than one variable, the item number is added (e.g., `kitad1103`). Finally, letters are used if items are split into several responses, thus requiring multiple variables (e.g., `kitae1205b`).

Missing values: Survey variables can be missing due to various reasons. To distinguish item-non-response (i.e., no answer) from other reasons for missing values, the following codes are used:

- 1 No answer
- 2 Does not apply
- 3 After intensive checks a given value was found to be implausible and was finally deleted (to be interpreted like -1)
- 5 Question not included in version of questionnaire

Additional data: Please note that for the sake of data protection open answers and the first given names of the children and main caregivers as inquired in the parent questionnaire were excluded from the scientific use file (SUF). However, these variables are partly available upon request. Also, the open answers to two questions on the center's pedagogical focus (director questionnaire 22, 'other') and the criteria based on which centers admit children in case demand exceeds the number of places (compressed version question 2, two open answers) have been recoded by the research team, however only for wave 1. The director file contains the variables `kitadfoc2 - kitadfoc11` and `kitadcr1 - kitadcr9`, which provide these recodings without liability. The group names were not delivered by TNS Infratest.

In addition, users may contact the K²ID research team in case they are interested in the pedagogical concepts which were collected in the course of the Institution Survey.

Inconsistencies: As compared to wave 1, the director questionnaire for wave 2 was modified in two ways. First, in question 32 the category “Erzieherinnen und Erzieher (Berufsausbildung)” (‘educator (vocational training)’) which accidentally had been omitted in wave 1 was added. Second, as part of question 46, the respondent could indicate the year and month when he or she became the center’s director. Therefore, wave 2 of the Institution Survey has slightly more variables than wave 1.

Please also note that a few variables are based on questions which were only asked in the compressed questionnaire for directors.

5.2 Data sets

In the final data set, case numbers are slightly lower as compared to the tables in chapter 4.2. This is because a few households accidentally received the documents for the Parent Survey even though they had already been part of the pretest, an exclusion criterion for participation. Information on their ECEC settings thus had to be deleted as well. This leaves a total of 1,080 children from 925 households with any ECEC data in the final data set. They attend 851 different centers and 994 ECEC groups.

When just using data from the Institution survey, merging the director data with the educator data results in 826 observations (826 groups in 763 centers) with information from both data sets. For 61 centers, no group information is available, whereas information from 31 groups cannot be matched with information from the director survey.

The K²ID data set is delivered in four different files: a Parent file, a Director file, an Educator file and a Brutto file. These different files are described in the following.

The **Parent file** is at the child level (N= 2836) and includes the data gathered as part of the Parent Survey (i.e., all variables starting with ‘kitak’). The identifiers `hhnrakt`, `persnr` and `parid` can be used together with the variable `welle` to add additional SOEP / FiD data. `welle` indicates the survey year based on which the gross sample of households was constructed. This was 2013 for wave 1 of the Parent Survey and 2014 for wave 2, respective-

ly. The binary variable `kwave` also indicates if the household took part in wave 1 or 2 of the Parent Survey. Further variables in the Parent file indicate whether parents reported the institution address late, whether households belonged to SOEP or FiD, the questionnaire version answered for each child, the within-household child and parent IDs and information on the date of the parent interview. Finally, the identifiers `kitaid` and `kigrd` can be used to merge ECEC information.

The **Director file** is at the level of ECEC institutions (N= 824) and captures all variables derived from the regular director questionnaire. In addition, this file contains part of the data stemming from the compressed questionnaire, namely those questions intended for the director (i.e., all variables starting with `'kitad'`). The identifier `kitaid` serves to match information from children and/or ECEC groups. The variable `iwave` indicates if the Institution Survey was administered in wave 1 or wave 2. It is not completely identical with `kwave` due to 'late repliers', and due to some children from Sample M attending the same center as children in the previous wave. If the same institution address was reported in both waves of the Parent Survey, TNS Infratest did not contact it a second time. Other variables (metadata) provide information on the date of the interview, whether data were collected via phone or mail and if the information was provided using the regular or the compressed version of the questionnaire.

The **Educator file** is at the level of groups (N= 857) and contains the variables originating from the regular educator questionnaire, as well as from the section of the compressed questionnaire dedicated to a specific group (i.e., all variables starting with `'kitae'`). Child and center data can be added using the identifier `kigrd`. Another variable distinguishes different groups within the same center. As in the Director file, `iwave` indicates if the Institution Survey was administered in wave 1 or wave 2. Again, the interview date, mode of data collection and an indicator of the questionnaire version are also included.

The **Brutto file** is at the child level (N= 4108) and includes all children who were identified as eligible for the study, including participants of the pretest. The file can be merged with information from all other K²ID files as well as with SOEP/FiD data using the identifiers `hhnrakt`, `persnr`, `kitaid`, `kigrd` and the variable `welle`. Some variables are the same as found in other K²ID files (e.g., `kwave`, `iwave`). Two variables provide information

on the households' subsample of origin. The file also provides the center's federal state (*ibula*), as well as two variables (*wpar*, *wkita*) containing the generated sampling weights (see also chapter 5.3). Another helpful variable is *ecec* as it details which combined information on ECEC from the different questionnaires (i.e., director and educator questionnaires, compressed questionnaire) is available for each of the 1,080 children in the data set. The remaining variables primarily refer to the data collection process. For example, they inform about which children and ECEC institutions were part of the pretest, whether a center address was reported, whether and why the interview was (not) realized, the processing method and when the questionnaires were sent out. Variables that refer to households and children start with the prefix 'k', whereas those referring to ECEC institutions start with the prefix 'i'.

Merging files: Below please find examples of STATA code which may help users to merge different K²ID files with each other or K²ID files with SOEP files. Regarding the latter, in many cases it is necessary to use the variable *welle* in addition to an individual or household identifier.

Merge K²ID-SOEP files

```
*open brutto data
  use "$directory\k2id_brutto.dta", clear

*merge parent data
  merge 1:1 persnr kwave using "$directory\k2id_parent.dta", update
  drop _merge //2836 matched

*merge director data
  merge m:1 kitaid using "$directory\k2id_director.dta", update
  ta _merge //1046 matched
  ren _merge m1

*merge educator data
  merge m:1 kigrd using "$directory\k2id_educator.dta", update
  ta ecec _merge //929 matched(32 with updated information
  ren _merge m2 //on date of institution survey)

*count number of children with matched ECEC data
  count if m1>2 | m2>2 //N=1078
```

Merge K²ID-SOEP files with SOEP data

```
*merge ppfad
  use "$directory\k2id_parent.dta", clear
  merge 1:1 persnr using "$soep\ppfad.dta", keepusing(sex gebjahr)
  drop if _merge==2 //2836 matched
  drop _merge

*merge bdkind using persnr and welle for participants of wave 1 of Parent Survey
```

```
merge 1:1 persnr welle using "$soep\bdkind.dta", keepusing(bdhhgr)
ta kwave _merge //2383 from wave 1 of Parent Survey matched
drop if _merge==2
drop _merge

*merge bekind using persnr and syear for participants of wave 2 of Parent Survey
gen syear=welle
merge 1:1 persnr syear using "$soep\bekind.dta", keepusing(behhgr)
ta kwave _merge //453 from wave 2 of Parent Survey matched
drop if _merge==2
drop _merge
```

5.3 Weighting

As detailed in chapter 4.2, non-response was an issue, a problem that is especially severe in case study participation is not random but correlated with unobserved characteristics of the sample units. A great advantage of the K²ID-SOEP study is that it is based on a panel survey, which means that one can draw on extensive information on both responding *and* non-responding families. Using available data on households and their members, the K²ID research team calculated statistical weights which correct for selective unit non-response, and combined these with the already existing design and survey weights for SOEP data. This procedure resulted in two weights, one for analyses of the Parent Survey only (*wpar*) and one for analyses including ECEC information from the Institution Survey (*wkita*). These weights are available for wave 1 of the K²ID study only, thus excluding Sample M.

5.3.1 Weights for the Parent Survey

The variable *wpar* was generated based on a logistic regression model estimating households' likelihood of participation in K²ID. A large set of predictors was taken into account, ranging from socio-demographic aspects such as mother's and child's age, household size, single parenthood and migration background to household income, maternal education and employment status. In addition, information on the participants' original study (SOEP or FiD), the federal state and size of the municipality in which the household is located was added. Note that if more than one sibling was eligible for survey participation in a household, child data was included for the youngest child. Refusal of the parents to participate in the survey can be considered independent of the willingness of the institution to participate in the

survey. Therefore, no information on the ECEC centers was needed for the generation of `wpar`. In the model, the following predictors of participation were statistically significant at the 10% level (with the direction in parentheses): mother's age (+), number of children in the household (-), mother single (-), migration background (-), household gross income (+), mother has no educational degree (-), participant originates from FiD (-). Based on the logistic regression model, households' inverse probability for participation was generated. It was then combined with the corresponding SOEP/FiD weight (`bdhhhrfak`). In fact, the largest part of the variation in the final weights originates from the SOEP. While the mean and standard deviation of these are 1120.72 and 1665.41, the mean and standard deviation of the combined weights are 1435.26 and 1937.23.

Questionnaire versions: All participants of the original SOEP samples were given the short questionnaire from the beginning; thus their assignment to the short questionnaire was independent of their behavior at this stage. On the other hand, FiD participants were given only the long version of the questionnaire. Only if they postponed the questionnaire repeatedly due to time constraints they could instead answer the short version of the questionnaire. Therefore, care has to be taken when considering items that appear only in the long version of the questionnaire as these could potentially be biased due to selective sampling into the short questionnaire by FiD parents. In terms of generation of the weights, selection into the different versions of the questionnaire is regarded as items non-response and is therefore not treated here. Weighting on the basis of different versions of the questionnaire or even items is theoretically feasible but does not seem efficient to the authors. Another way would be multiple imputation but we leave it to users of the data to implement this for their specific applications also due to efficiency reasons. The number of people opting into the CATI version of the questionnaire was relatively small and therefore this strand was not followed further in the weighting procedure.

5.3.2 Weights for the Institution Survey

Concerning the weight for ECEC settings as captured by `wkita`, some of the parental characteristics were used in the logistic regression model estimating ECEC centers' likelihood of participation in K²ID as well. Besides, even though the institutions were surveyed for the first

time, we still had some information also on nonresponding institutions thanks to answers provided by the parents, including the centers' location. Thus, we incorporated data on the institutions' providers reported as part of the SOEP/FiD interviews, as well as information on the institutions' federal state and municipality size. The average share of children in ECEC with migration background in the respective youth welfare office district was accounted for as well. Finally, institutions' inverse probability for participation was combined with w_{par} . While the available information on ECEC institutions is less extensive than data on households and individuals, it is still useful to build a nonresponse model. Our information is comparable to company surveys such as Eurofound's European Company Survey (Eurofound 2015) which only use country of origin, size of the company and industry sector. In the logistic regression model, the following characteristics were predictive of participation of the institution given that the parents had provided the address: mother has tertiary degree (+), mother is unemployed (not including those not in the labor force) (+), (direct) migration background (+), institution is privately run (+). These weights add a greater amount of variability (mean and standard deviation 2412.60 and 3420.93) to the baseline SOEP weights compared to the household weights meaning that certain observations can be relatively more influential in analyses using the institution weights.

6 Survey instruments

The survey instruments cover the following questionnaires: (1) Parent Questionnaire long: Wave 1, (2) Parent Questionnaire short: Wave 1 & 2, (3) Director Questionnaire: Wave 1 & 2, (4) Educator Questionnaire: Wave 1 & 2, and (5) Compressed Questionnaire: Wave 1 & 2.

They will be made available together with the data.

Further instruments include the questionnaires for the pretests and for the CATI follow-up, which can be made available upon request.

Literature

Anders, Yvonne and Itala Ballaschk, (2014): Studie zur Untersuchung der Reliabilität und Validität des Zertifizierungsverfahrens der Stiftung „Haus der kleinen Forscher“. In Stiftung Haus der kleinen Forscher (Ed.), *Wissenschaftliche Untersuchungen zur Arbeit der Stiftung „Haus der kleinen Forscher“*, Band 6. (pp.35-116). Schaffhausen: Schubi Lernmedien AG.

Anders, Yvonne and Hans-Günther Roßbach (2015): Preschool Teachers' Sensitivity to Mathematics in Children's Play: The Influence of Math-Related School Experiences, Emotional Attitudes, and Pedagogical Beliefs. *Journal of Research in Childhood Education*, 29(3), 305–322.

Becker, Birgit, Nicole Biedinger, Oliver Klein, and Franziska Schmidt (Forthcoming): Erwerb von sprachlichen und kulturellen Kompetenzen von Migrantenkindern (Technical Report). Mannheim: Mannheimer Zentrum für Europäische Sozialforschung.

Blossfeld, Hans-Peter, Hans-Günther Roßbach, and Jutta von Maurice (2011): Education as a Lifelong Process - The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*.

Bohlender, Axel and Anne Glemser (2016): Kinder und Kitas in Deutschland (K2ID). Methodenbericht. München: TNS Infratest Sozialforschung.

Eurofound (2015): Third European Company Survey – Overview report: Workplace practices – Patterns, performance and well-being. Luxembourg: Publications Office of the European Union.

Fuchs-Rechlin, Kirsten, Gudula Kaufhold, Mareike Thuilot, and Tanja Webs (2014): Der U3-Ausbau im Endspurt: Analysen zu kommunalen Betreuungsbedarfen und Betreuungswünschen von Eltern. Dortmund: Forschungsverbund DJI - TU Dortmund.

Gerstorff, Sandra and Jürgen Schupp (2014): SOEP Wave Report 2013. Berlin: DIW Berlin.

Goddard, Kate and Kate Groucutt (2011): Choosing quality: Supporting parents to access high quality childcare. Policy Insight Paper 4. London: Daycare Trust.

Hachfeld, Axinja and Yvonne Anders (2017): Dokumentation der Validierungsstudie im Rahmen des Projekts K2ID: Untersuchung der Validität der Fragebögen zur Messung von Kita-Qualität anhand eines Beobachtungsverfahrens. Berlin: Freie Universität Berlin.

Jänsch, Agnes, Simon Huber, Nico A. Siegel, Salma Stimmel, and Dirk Geue (2011): „Familien in Deutschland“ (FiD) 2010 Methodenbericht: Anlage und Ergebnisse der FiD-Stichproben. München: TNS Infratest Sozialforschung.

Kroh, Martin, Simon Kühne, and Rainer Siegers (2015): Documentation of sample sizes and panel attrition in the German Socio-Economic Panel (SOEP) (1984 until 2014). SOEP survey papers Series C - Data documentation: Vol. 297. Berlin: DIW Berlin.

Kurz, Karin, Jens Kratzmann, and Jutta von Maurice (2007): Die BiKS-Studie. Methodenbericht zur Stichprobenziehung. Retrieved April 4, 2017 from psy-dok.psycharchives.de/jspui/bitstream/20.500.11780/437/1/Methodenbericht_2007.pdf

McCabe, Lisa A. and Debra J. Ackerman (2007): Child Care Center Quality: Measurement Issues and Links to Child Developmental Outcomes. Data Documentation 25. Berlin: DIW Berlin.

Roßbach, Hans-Günther and Leal, T. B. (1993): Mütterfragebogen zu kindlichen Aktivitäten im Kontext des Familiensettings (AKFRA). Deutsche Fassung des Questionnaire on preschool-aged children's activities in the family. Unveröffentlichtes Manuskript.

Schaufeli, Wilmar B., Anorid B. Bakker, and Marisa Salanova (2006): The Measurement of Work Engagement With a Short Questionnaire: A Cross-National Study. *Educational and Psychological Measurement*, 66(4), 701–716.

Schröder, Mathias, Rainer Siegers, and C. Katharina Spieß (2013): „Familien in Deutschland“ – FiD. Schmollers Jahrbuch, 133(4), 595-606.

SOEP Group (2005): Desktop Companion to the German Socio-Economic Panel (SOEP). Berlin: DIW Berlin.

Statistisches Bundesamt. (2012): Statistiken der Kinder- und Jugendhilfe. Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege: am 01.03.2012. Wiesbaden.

Tietze, Wolfgang, Fabienne Becker-Stoll, Joachim Bense, Gabriel Haug-Schnabel, Bernhard Kalicki, Heidi Keller, and Birgit Leyendecker (2013): NUBBEK. Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit. Weimar/Berlin: verlag das netz.

Tietze, Wolfgang, Käthe-Maria Schuster, Katja Grenner, and Hans-Günther Roßbach (2007): Frühe Kindheit : Qualitätssicherung. Kindergarten-Skala: (KES-R); Feststellung und Unterstützung pädagogischer Qualität in Kindergärten; deutsche Fassung der Early childhood environment rating scale - revised edition von Thelma Harms, Richard M. Clifford, Deborah Reid Cryer (Rev. Fassung, 3., überarb. Aufl., [Nachdr.]). Berlin: Cornelsen Scriptor.

TNS Infratest Sozialforschung (2011): SOEP 2007 – Methodenbericht CAPI-Innovationsbefragung zum Befragungsjahr 2007 (Welle 24) des Sozio-oekonomischen Panels – „Persönlichkeit und Gesundheit“. SOEP Survey Papers 63: Series B. Berlin: DIW Berlin /SOEP.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007): The German Socio-Economic Panel Study (SOEP) - Scope, Evolution, Enhancements. Schmollers Jahrbuch, 127(1), 139–169.

Appendix

Table 6

Overview of questions in Parent Survey, their sources and modifications if applicable

Question no.		Source	Comment
Long	Short		
01	01	Wagner et al. (2007), person questionnaire [2012]; inspired	Refers to the importance of aspects of non-parental care rather than different life domains
02		Becker et al. (forthcoming), parent questionnaire [wave 1]; modified	7-point response scale; reduced item list
03	02	Wagner et al. (2007), household questionnaire [2012]; slightly modified	After school centers are not included
04	03	Fuchs-Rechlin et al. 2014; adjusted	
05	04	Becker et al. (forthcoming), parent questionnaire [wave 1]; adjusted	
06	05	Own development	
07	06	Wagner et al. (2007), pretest [2007]	
08		Tietze et al. (2013), questionnaire for pedagogical leader; adjusted	'on most days of the week' instead of 'usually'
09¹	08	Tietze et al. (2013), questionnaire for pedagogical leader; adjusted	
10¹		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Several response categories added
11¹		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; Kurz et al. (2007), BiKS-3-8 educator questionnaire [wave 3]; modified	Altered item list and wording of question
12¹	09	Blossfeld et al. (2011), SC2 director	Altered item list and response scale

13		questionnaire [wave 1]; Kurz et al. (2007), BiKS-3-8 educator questionnaire [wave 3]; modified Becker et al. (forthcoming), parent questionnaire [wave 1]; Goddard, K., & Groucutt, K. (2011); strongly modified	Ranking of aspects required; modified response categories
14^{1;2}	10	Schröder et al. (2013), parent questionnaire (birth cohort 2007) [2010]; Becker et al. (forthcoming), parent questionnaire [wave 2]; strongly modified	Several aspects were added or modified
15^{1;2}	11	Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; Becker et al. (forthcoming), kindergarten questionnaire; Wagner et al. (2007), parent questionnaire (children 2 to 3 years) [2012]; strongly modified	Modified scales and response categories; refers to frequency with which child takes part in different activities
16	12	Own development	
17	13	Own development	
18	14	Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; strongly modified	Original question was split
19	15	Wagner et al. (2007), parent questionnaire (children 2 to 3 years) [2012]; Schröder et al. (2013), parent questionnaire (birth cohort 2007) [2010]; Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; inspired	Question refers to children in daycare rather than household members
20²	16	Tietze et al. (2013), questionnaire for	Sum of caregivers a) in group, b) usually

		educator; strongly adjusted	present at the same time, and who have not completed training (instead of listing the job titles, roles and working hours of all staff members)
21 ²		Own development	
22 ²		Roßbach und Leal (1993); Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	7-point response scale; shortened item list as compared to Roßbach und Leal (1993)
23		Own development	
24		Own development	
25 ^{1;2}	17	Schröder et al. (2013), parent questionnaire (birth cohort 2007) [2010]; adjusted	

¹ Question was posed to director as well; ² Question was posed to group educator as well.

Note: Adjusted = minor changes that are mostly related to wording; modified = more substantial changes that are further specified in the right column; inspired = only loosely related

Table 7

Overview of questions in Institution Survey, their sources and modifications if applicable

Question no.			Source	Comment
ECEC Institution				
Princi- pal	Educa- tor	Comp- ressed		
1			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; Tietze et al. (2013), questionnaire for pedagogical leader; modified	Response categories combined to five categories
	1	8	Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Refers to institution as a whole rather than work with four-year-olds
2			See Table A-1 (Long Q8)	
3			Becker et al. (forthcoming), kindergarten questionnaire; adjusted	
4			Tietze et al. (2013), questionnaire for pedagogical leader; adjusted	
5	2		Blossfeld et al. (2011), SC2 director / educator questionnaire [wave 1]; modified	Overall number of children with foreign language of origin (instead migration background) / with handicap ('integration status') but no distinction by birth year
		1/9		
6	3		Blossfeld et al. (2011), SC2	Distinction between age groups

			director / educator questionnaire [wave 1]; modified	instead of birth cohorts; in compressed version no question about number of boys and girls, children with handicap and no distinction by care hours for children in center
	4		Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Number of children with foreign language of origin (instead migration background)
	5		Becker et al. (forthcoming), kindergarten questionnaire; modified	Educator is not asked to indicate number of groups with language training
	6	10	See Table A-1 (Long Q20)	
	7		Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Working hours instead of employment percentages; trainees and interns explicitly included; more rows
7			Own development	
8			Becker et al. (forthcoming), kindergarten questionnaire; modified	Addition: number of children who will be admitted to facility next year
9		2	Becker et al. (forthcoming), kindergarten questionnaire; modified	Categories modified/added; in compressed version open question
	8		Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]	
	9		See Table A-1 (Long Q21)	

10		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; adjusted	
	10	Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Addition: educator can indicate if equipment is insufficient
11		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Six instead of five categories, distinction between groups serving children from 3 to school entry vs. from 2 to school entry
	11	Roßbach und Leal (1993); also see Table A-1 (Long 22); adjusted	
12		See Table A-1 (Long Q11)	
13		See Table A-1 (Long Q12)	
14		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; strongly modified	Only 2 response categories specifying if measures for quality development were internal / external
	14	Kurz et al. (2007), BiKS-3-8 educator questionnaire [wave 4]; modified	6-point response scale; 'role model' item dropped
15		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]	
	15	Wagner et al. (2007), mother and child questionnaire (7-8-year-olds) [2012]; modified	Changed order of items; 'obeys persons of authority ' instead of his/her parents

16			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	'No'-category included; refers to size of garden rather than outside area
	16		Blossfeld et al. (2011), SC2 director questionnaire [wave 1];	Refers to group, not institution as a whole; six-point response scale
17			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Refers to rooms offering a specific kind of learning provision rather than all existing rooms; other categories, including 'No'-category
	17		Schaufeli et al. 2006; modified	Strong reduction of items
18			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; strongly modified	Items combined / modified / replaced by items referring to quality and search for personnel
	18		Wagner et al. (2007), person questionnaire [2012]; modified	Item on time pressure kept, all others dropped
19			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]	
	19		Anders & Ballaschk 2014; modified	Restricted to one example of a game, activity, project or learning opportunity, not just experiments
20		7	Tietze et al. (2013), questionnaire for pedagogical leader; see also Table A-1 (Q9); modified	Respondents may indicate if and where pedagogical concept can be downloaded online; in compressed version only whether or

				not available online
21	20		Anders & Ballaschk 2014; inspired	
			Tietze et al. (2013), questionnaire for pedagogical leader; modified	Two questions combined, further response categories added
22	21		Anders & Ballaschk 2014; inspired	
		6	Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Response categories added; in compressed version open question
23	22		Anders & Roßbach 2015	
			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Yes/No question without further distinctions
24	23		Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Distinguishes between first and second language; educator is not asked to indicate the language
			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	5-point-scale ranging from 'at least once per week' to 'never'
25	24		Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Refers to language use with children in group, not their parents
			Frühe Chancen online questionnaire of director; modified	Response categories modified

26	12	11	See Table A-1 (Long Q15)	Educator / compressed version: refers to <i>frequency</i> with which different activities are offered; Director: refers to whether and for which <i>age groups</i> activities are offered; compressed version: question for additional costs dropped
27			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Speech promotion is explicitly included rather than excluded; refers to children with non-German language of origin rather than migration background
28	25	4	See Table A-1 (Long Q14)	Compressed version: 8 out of 21 items asked
29	26	5	See Table A-1 (Long Q25)	Educator: additional question on satisfaction with group
30	13	3	Kurz et al. (2007), BiKS-3-8 educator questionnaire [wave 4]; modified	Rating of importance of aspects removed; some items dropped; order changed
31	27		Wagner et al. (2007), person questionnaire [2012]; adjusted	Educator: no question on satisfaction with work
32			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Working hours instead of employment percentages; distinction between numbers of staff working full- vs. part-time; categories modified; order changed; director is not asked to indicate

33			Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; strongly modified	personal qualification Refers to institution as a whole in addition to staff with non-German language of origin (instead of migration background); trainees and interns explicitly included
34	28		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]	
35	29		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Only asks for year of birth, not specific month
36	30		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]	
37	31	12	Blossfeld et al. (2011), SC2 educator questionnaire [wave 1]; modified	Response categories combined / modified, e.g., to include persons with Bachelor and Master degrees; compressed version: open question
38	32		Own development	
39	33		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; adjusted	
40		13	Blossfeld et al. (2011), SC2 director questionnaire [wave 1]	

41	34		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]	
42	35		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Originally 2 questions; fewer response categories: overall, direct work with children, managerial tasks (director only)
43	36		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	Strong reduction of response categories
44	37		Blossfeld et al. (2011), SC2 director questionnaire [wave 1]; modified	No open questions, respondents can only check boxes
45	38		Wagner et al. (2007), person questionnaire [2012]	
46	39	14	Own development	Compressed version: fewer categories
47	40		Own development	
48			Own development	

Note: Adjusted = minor changes that are mostly related to wording; modified = more substantial changes that are further specified in the right column; inspired = only loosely related

September 12th, 2017

Corrigendum

In the parent file of the K²ID-SOEP data, the variable which captures the parent ID has been renamed from `parid` to `parnr`.