

**DIRECTORATE FOR EDUCATION AND SKILLS
EDUCATION POLICY COMMITTEE**

Network on Early Childhood Education and Care

**MEASUREMENT OF PRESCHOOL QUALITY WITHIN THE NATIONAL EDUCATIONAL PANEL
STUDY**

RESULTS OF A METHODOLOGICAL STUDY.

SUPPLEMENTARY DOCUMENT to EDU/EDPC/ECEC(2013)14

**10-11 DECEMBER 2013
Amora Hotel
Wellington, New Zealand**

Miho Taguma, Project Leader; Tel: + 33 1 45 24 92 65; Email: miho.taguma@oecd.org

JT03350324

Complete document available on OLIS in its original format

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Measurement of Preschool Quality Within the National Educational Panel Study – Results of a Methodological Study¹

Thomas Bäumer², Hans-Günther Roßbach³

Abstract

It has been argued that for the assessment of educational quality—especially in preschool—observational methods is the silver bullet. But in large-scale assessments like the National Educational Panel Study (NEPS) observations can hardly be conducted. Against this background we carried out a study organized around the question as to what extent preschool quality can be assessed using teacher questionnaires. Therefore, a standardized observation of 60 preschool groups, using the German versions of the Early Childhood Environment Rating Scale (ECERS; KES-R and KES-R-E)—a well-established rating instrument for (process) quality in preschool environments—was conducted. Also, teachers filled out a questionnaire on preschool quality from pilot studies of the NEPS. In this paper, we present main results from the comparative analyses of observation and survey of preschool quality using regression analytical methods. It can be shown that on a global level preschool quality can be reproduced quite well by means of questionnaire data. Conclusions concerning the questionnaire design of the NEPS main study on the Starting Cohort 2–Kindergarten are drawn. Finally, some cautionary notes on the use of single indicators of preschool quality and on causal inferences are given.

¹ We would like to express our gratitude and appreciation to Carina Pömp who collected the data in the context of her diploma thesis. This paper discusses the main results of the study. In a forthcoming NEPS Working Paper we will deal with the rich data resource of this study in more detail.

² National Educational Panel Study (NEPS), University of Bamberg.

³ National Educational Panel Study (NEPS), University of Bamberg.

Measurement of Preschool Quality Within the National Educational Panel Study – Results of a Methodological Study

1 INTRODUCTION

The measurement of the quality of learning environments is one of the most prominent but also one of the most difficult issues in educational research. It has been argued that a comprehensive view on educational quality necessitates a triangulation of the perspectives of teachers, students and external observers. Every perspective has its own advantages as well as disadvantages for different aspects of quality and also in relation to outcomes that may depend on educational quality. In panel studies like the NEPS, external observations are often too costly and impose additional challenges to data dissemination. Therefore, we have to rely on reports from respondents, usually teachers and students. Yet in preschool only the teachers' perspective can be surveyed because children are too young to respond adequately to a questionnaire or interview. Especially in the case of preschool quality, one has to account for respondent bias because it is very likely that teachers will perceive the survey as an evaluation of their work. In the following, we will describe a methodological study that compares the observation of preschool quality to reports made by preschool teachers. First, we will give a few notes on the concept of educational quality that is used within the NEPS.

2 THEORETICAL BACKGROUND

At least since the “PISA-Shock” in Germany, early education institutions were brought to public attention. There is ample evidence that preschool attendance has long-term effects on cognitive and social development of the child (cf. Anders, in press). Furthermore, it can be shown that these effects are moderated by the quality of preschool (cf. Roßbach, 2004, 2005; Roßbach, Kluczniok, & Kuger, 2008; Sylva, Melhuish, Sammons, & Taggart, 2004). Therefore, the assessment of quality of the preschool groups attended by the children (target persons) of NEPS Starting Cohort 2–Kindergarten is given primary attention.

As has been argued by Clausen (2002) in educational settings different perspectives in the evaluation of quality should be accounted for. Teachers as well as students can perceive different aspects with different accurateness. An external observer might be the most reliable source of information but has only access to the educational situation for a limited time and observation might additionally alter the situation that is observed. In preschool accounts by children cannot be used because they are too young to give reliable assessments. Therefore, observation is the method of choice in preschool settings. But in large-scale assessments like NEPS with a huge amount of target persons and educational settings observational studies are too costly and time-consuming that they can be applied on a regular basis. Here we have to rely on teachers' reports on the quality of their preschool groups. The assumption that teachers' perspectives might be distorted because they perceive the survey setting as an evaluation of their own work within the preschool group motivated the present study.

Tietze et al. (1998) define educational quality of preschool with the help of three elements that match the conceptualization of educational quality within the NEPS (cf. Bäumer, Preis, Roßbach, Stecher, & Klieme, 2011; Bäumer, & Roßbach, 2012): *Structural quality* includes the framing conditions of an educational setting, *orientational quality* reflects the beliefs and opinions concerning education of the actors within an educational settings, and *process quality* relates mainly to the interactions of the actors within an educational setting, for example of teachers and children of a preschool group. Structural and orientational quality serve as input conditions for process quality that directly influences outcomes, whereas the former are proposed to influence outcomes mainly indirectly.

3 METHOD

3.1 Sample

The sample consists of 60 groups from 42 preschools in Upper Franconia. Recruitment was conducted under the following two conditions: (1) age range of the attending children is between three and six years and (2) no special pedagogical conception, like Waldorf, is followed.

3.2 Instruments

Observational Instruments

The observation was conducted using the *KES-R* (“Kindergarten-Skala – Revidierte Fassung”; Tietze, Schuster, Grenner, & Roßbach, 2005) and the *KES-R-E* (“Kindergarten-Skala-Erweiterung”; Roßbach, & Tietze, in prep) that are the German versions of the ECERS-R (“Early Childhood Environment Rating Scale. Revised Edition”; Harms, Clifford, & Cryer, 1998), respectively ECERS-E (“Early Childhood Environment Rating Scale-Extension”; Sylva, Siraj-Blatchford, & Taggart, 2003). The *KES-R* assesses the global educational process quality of preschool groups. It consists of 43 items each with a seven point rating scale ranging from 1 (“deficient”) to 7 (“excellent”). As a general guideline scores of 1 to 3 denominate a poor quality and scores of 5 to 7 a good quality. Scores of 3 to 5 denominate a mediocre quality. The 43 items can be organized into seven subscales: *Space and Furnishings*, *Personal Care Routines*, *Language-Reasoning*, *Activities*, *Interactions*, *Program Structure*, *Parents and Staff*. The last subscale was not used in this study thus using only 37 items for the observation. The *KES-R-E* adds to the global assessment more domain-specific aspects. It consists of four subscales: *Literacy*, *Mathematics*, *Science and Environment* as well as *Diversity*. The 18 items are rated using the same scale as with the *KES-R*.

Teacher Questionnaire

The questionnaire used in this study was compiled of questions used in the first pilot study of NEPS Starting Cohort 2—Kindergarten. The questions can be subdivided into four broad domains: (1) *structural characteristics*: opening hours, closing times, group size, room count and space and personnel (full time equivalents), (2) *compositional characteristics*: handicapped children, children with development disorders, children with migration background, age groups of children, children by gender, amount of child care per child, (3) *materials and activities* and (4) *teachers characteristics*: qualification, working hours and work schedule, further education and supervision.

3.3 Procedure

The study was undertaken from November 2009 to February 2010. The observations took place during a three to four hour session in the morning and were conducted by one or two trained observers. For half of the sample the questionnaires were sent to the teachers previously to the observation, for the other half the questionnaire was handled over after the observation to control for the sequencing of the survey methods. The data of the ratings form and the questionnaires were manually keyed in twice by different typists to avoid input data error. Analyses were conducted using IBM SPSS Statistics 20.0.0.

4 RESULTS

First the descriptives of the observational and the questionnaire data are shown. After that, regression of the observed quality on teacher reports is presented.

4.1 Descriptives

KES-R Observational Data

In Table 1 descriptives of the scales and subscales of the KES-R and KES-R-E are summarized. All scales are computed as means of the corresponding items and scale values range accordingly from 1 to 7. Most scale mean scores show a mediocre quality with the exception of *Language Reasoning* which is on average of good quality. Two scales show on average even poor quality of the 60 examined preschool groups: *Diversity* and *Personal Care Routines*. For diversity with a maximum score of 4.00 no single preschool group reaches good quality. For the three general scales (*KES-R*, *KES-R-E* and *KES-R/R-E Composite*) the preschool groups score on average a mediocre quality. Correlations of the six KES-R subscales range from $r = .23$ and $r = .71$. Associations of the four KES-R-E subscales are generally higher, ranging from $r = .63$ to $r = .96$. Scale-subscale correlations range from $r = .64$ to $r = .81$ for the *KES-R* respectively from $r = .82$ to $r = .96$ for the *KES-R-E*. Finally, *KES-R* and *KES-R-E* scales are correlated by $r = .71$ thus forming a close association. Therefore, the *KES-R/R-E Composite* score is used for the regression analyses (see below). Correlations of the *KES-R/R-E Composite* to *KES-R* equals $r = .90$ and to *KES-R-E* $r = .95$. Associations with the 10 subscales range from $r = .52$ to $r = .92$ so that most information in the data is retained. Using the *KES-R/R-E Composite* score, most preschool groups show a mediocre quality ($n = 45$, 75 %) and only four groups (7 %) score above 5 thus showing a good quality. For 11 preschool groups (18 %) the quality has to be termed poor.

Table 1

Descriptives of the KES-R and KES-R-E scales

Scale	<i>M</i>	<i>SD</i>	<i>min</i>	<i>max</i>
1. Space and Furnishings	3.59	0.84	2.13	5.88
2. Personal Care Routines	2.82	1.06	1.50	6.00
3. Language-Reasoning	5.17	1.07	3.00	7.00
4. Activities	3.96	0.79	2.50	6.11
5. Interactions	4.59	1.01	2.00	7.00
6. Program Structure	3.71	0.84	2.33	7.00
7. Literacy	3.89	0.95	1.67	6.17
8. Mathematics,	3.43	1.08	1.22	5.67
9. Science and Environment	3.53	1.08	1.14	5.71
10. Diversity	2.39	0.97	1.00	4.00
11. KES-R	3.97	0.69	2.71	5.79
12. KES-R-E	3.31	0.93	1.72	5.30
13. KES-R/R-E Composite	3.64	0.75	2.53	5.46

NEPS Questionnaire Data

Description of the questionnaire data is ordered concerning the subdivision shown in section 3.2. Due to space limits, we only show variables that are used in the subsequent regression analyses. Selection was done by examining the bivariate correlations between all variables and due to content-related considerations. Also, the exact mathematical derivation of the variables from the items will not be discussed in detail.

The following *structural characteristics* are considered:

1. *Opening Hours* (Monday to Friday): This is simply the sum of the daily opening hours. The score varies from 33 to 63 hours per week with a mean of $M = 44.53$ ($SD = 4.50$; $n = 59$).

2. *Weekend Opening*: This is a dichotomous variable. The teachers of 4 of the 60 preschool groups (6.7 %) state that their preschool opens at weekends.
3. *Vacation Closure*: Here the weeks per year the preschool is closed is specified. This variable ranges from 0 to 6 weeks with a mean of $M = 3.95$ ($SD = 1.23$; $n = 58$).
4. *Group Size*: One teacher stated the group size to be 49 children. Because this clearly point to open group work, this group was not used in the regression analyses. The variable varies between 12 and 28 children per group with a mean of $M = 23.53$ ($SD = 3.04$; $n = 59$).
5. *Room Number*: Most of the groups use one room ($n = 17$, 28.3 %) or two rooms ($n = 20$, 33.3 %), only 12 groups use more rooms (20.0 %). There is no information available for 11 groups (18.3 %).
6. *Room Size*: Size of the available rooms vary from 28 to 431 sq. m with a mean of $M = 78.93$ ($SD = 63.43$; $n = 45$). Please note the large standard deviation and number of missings.
7. *Working Time*: This variable consists of the full time equivalent of the first and second teacher in the group. It varies from 86 % to 200 % with a mean of $M = 167.10$ ($SD = 34.27$; $n = 58$).

Next, we display the *compositional characteristics*. In general, a bit more computation was necessary for these variables. Because the variable *Count of Children with Migration Background* did not vary sufficiently in the sample and is highly skewed ($M = 4.18$; $Md = 1$; $SD = 7.69$; $n = 60$), it was skipped from the analyses. 60 % of the preschool groups are attended by no or only one child with migration background.

1. *Count of Children with Disabilities*: Here all children in a group with disabilities or developmental disorders were added. The variable is a composite of several items of the questionnaire. Most groups are not attended by these children ($n = 47$, 78.3%). In the remaining groups the count varies between 1 and 12. Therefore, the mean is relatively small with $M = 0.58$ ($SD = 1.80$; $n = 60$).
2. *Average Age*: The average age of the children varies between groups from 43.5 months to 61.68 months with a mean of $M = 55.11$ ($SD = 3.43$; $n = 54$).
3. *Age Variability*: Here, simply the standard deviation of the age in month of the children in the group was computed. It varies between 6 and 16 months with a mean of $M = 11.88$ ($SD = 2.21$; $n = 54$). Taken average age and age variability together, it is indicated that the preschool groups are aged-mixed
4. *Average Hours of Care*: This variable was computed by summing up the hours of care for every child in a group divided by group size. The score ranges between 4.69 and 7.74 hours per day with a mean of $M = 6.28$ ($SD = 0.82$; $n = 46$).
5. *Balanced Gender Ratio*: This variable is derived to reflect an equal count of boys and girls in the group, because scatterplot of boy-girl-ratio with *KES-R/R-E Composite* showed an inverted U-shaped curve with a maximum at 1 (corresponding to an equal count of boys and girls). The variable is computed by the absolute value of 1 minus count of boys divided by count of girls. Thus the larger the

value the more unbalanced is the gender ratio in the group. In contrast, a 0 reflects an equal count of boys and girls. Thus, a negative correlation coefficient reflects a positive association.

Concerning *materials and activities* only a few variables can be used. For activities an insufficient variability was realized in the study, leading to an adapted response scale in the NEPS main studies. In the following, activities are excluded. Only three variables will be treated:

1. *Materials*: The scale is a mean score of 14 items using a three-point response scale with values 1 (materials are available for some children) to 3 (materials are available for all children). An indication that materials are not available at all is scored 0. The variable varies between 0.53 and 3.00 with a mean of $M = 1.52$ ($SD = 0.40$; $n = 60$).
2. *Visits (Museum)*: Teachers had to indicate how often such visits are offered to the children. The response scale ranges from 1 (“never”) to 6 (“(nearly) daily”). In the sample only values 1 to 4 (“(nearly) monthly”) were used. The mean equals $M = 1.71$ ($SD = 0.80$; $n = 58$).
3. *Visits (Theatre, Cinema, Concert)*: Using the same scale and realizing the same range the mean of this variable equals $M = 2.39$ ($SD = 0.65$; $n = 57$).

Finally, the following *teacher characteristics* were considered:

1. *Years of Education*: 31 teachers (51.7 %) left school after Grade 10, 24 teachers (40.0 %) after Grade 12 and 5 teachers (8.3 %) after Grade 13.
2. *Work Experience*: The total count of years in work was specified. The variable varies between 2 and 41 years with a mean of $M = 17.65$ ($SD = 9.75$; $n = 57$).
3. *Contractual Working Hours*: The teachers reported contractual working hours between 19.5 and 40 hours a week. Mean equals $M = 34.59$ ($SD = 6.87$; $n = 60$).
4. *Overtime*: The difference between the factual and the contractual working hours was computed. It ranges from 0 to 11 hours per week with a mean of $M = 2.40$ ($SD = 3.29$; $n = 56$).
5. *Hours of Work with Children*: The direct work with the children in the group varies between 0 and 38 hours per week with a mean of $M = 24.85$ ($SD = 9.50$; $n = 56$).
6. *Hours of Work without Children*: Included in this variable is for example preparation time, team meetings and other work. This variable ranges from 1 to 37 hours per week with a mean of $M = 9.15$ ($SD = 6.66$; $n = 59$).
7. *Advanced Training*: This is simply a dichotomous variable. The item asks for additional certified training. 9 teachers (15 %) affirmed this question, 40 teachers (66.7 %) negate it. 11 answers (18.3 %) are missing.
8. *Number of Types of Further Education*: Teachers had to indicate how many courses of further education with different content they had attended in the last year. The answers vary between 0 and 11 with a mean of $M = 2.90$ ($SD = 2.05$; $n = 60$).

9. *Hours of Further Education*: Here, the total amount of hours, spend for further education in the last year was computed. This variable ranges from 0 to 171 hours a year with a mean of $M = 28.70$ ($SD = 29.24$; $n = 60$).
10. *Supervision*: This dichotomous variable indicates the availability of supervision for the teachers. In the sample only 7 teachers (11.7 %) receive supervision, the majority does not ($n = 53$; 88.3 %).

4.2 Regression Analyses

Because of the relatively small sample size, regression analyses are used in a more descriptive way. In the following, separate regression analyses of the *KES-R/R-E Composite* as criteria are performed for the four domains of questionnaire data. Starting with the complete bunch of predictors of one domain, a stepwise procedure is employed to reduce the set to the most predictive characteristics. Instead of using an algorithm that is implemented in statistical package, selection of variables is done by the following steps: (1) collinearity diagnostics (using condition index, CI), (2) selection of variable(s) to be excluded (beside the current criteria like tolerance and variance inflation factor, correlation of predictors and theoretical criteria are considered), (3) regression with reduced set till condition index falls under 30 (cf. Bühner, & Ziegler, 2009). After selecting the most predictive variables for each of the four domains an overall regression series is conducted using all selected variables. For all analyses only additive models are estimated, no interactions between predictors are added.

Regression of KES-R/R-E Composite on structural characteristics

Table 2 shows the stepwise regression on the *structural characteristics*. Outcome variable is the *KES-R/R-E Composite*.

Table 2

Regression of KES-R/R-E Composite on structural characteristics

	β_1	β_2	β_3	β_4	β_5	β_6
<i>Working Time</i>	.44	.26	.26	.29	.28	.30
<i>Opening Hours</i>	.21	.30	.30	.17	.20	.18
<i>Weekend Opening</i>	.19	.17	.17	.14	.13	
<i>Room Size</i>	-.49	-.13	-.13	-.09		
<i>Vacation Closure</i>	.21	.27	.26			
<i>Group Size</i>	.02	.01				
<i>Room Number</i>	.46					
<i>R</i>	.54	.47	.47	.42	.41	.39
<i>R²</i>	.29	.22	.22	.18	.18	.15
<i>R²_{corr}</i>	.15	.10	.12	.09	.12	.12
<i>CI</i>	60.7	56.4	47.5	35.2	28.9	28.2

In the final step, two variables remain that explain 12 % of the variation of the *KES-R/R-E Composite* (using the corrected explained variance, R^2_{corr}). *Working Time* i.e. the time spend by one or two teachers within the group and *Opening Hours* i.e. hours the preschool can be attended on weekdays.

Regression of KES-R/R-E Composite on compositional characteristics

Table 3 shows the stepwise regression on the *compositional characteristics*. Within three steps, three variables are selected that explain 22 % of the variance of the *KES-R/R-E Composite*. Whereas *Average Hours of Care* and *Balanced Gender Ratio* contribute positively to the overall quality of the preschool group, the more *Children with Disabilities* are taken care of the lower the *KES-R/R-E Composite*.

Table 3
Regression of KES-R/R-E Composite on compositional characteristics

	β_1	β_2	β_3
Average Hours of Care	.47	.47	.43
Balanced Gender Ratio	-.19	-.19	-.18
Count of Children with Disabilities	-.28	-.28	-.25
Average Age	-.14	-.14	
Age Variability	.00		
R	.54	.54	.52
R ²	.29	.29	.27
R ² _{corr}	.20	.22	.22
CI	52.5	46.4	18.9

Regression of KES-R/R-E Composite on materials and activities

Table 4 shows the stepwise regression on the *materials and activities* of the *KES-R/R-E Composite*. Using two of the three variables (*Materials* and *Visits (Museum)*), 4 % of the variance of the *KES-R/R-E Composite* can be explained.

Table 4
Regression of KES-R/R-E Composite on materials and activities

	β_1	β_2
Materials	.12	.17
Visits (Museum)	.16	.19
Visits (Theatre, Cinema, Concert)	.17	
R	.32	.27
R ²	.10	.08
R ² _{corr}	.05	.04
CI	11.1	9.4

Regression of KES-R/R-E Composite on teacher characteristics

Table 5 shows the stepwise regression on the *teacher characteristics* of the *KES-R/R-E Composite*. After stepwise reduction of the set of 10 predictors, 5 variables remain that predict 12 % of the variance of the *KES-R/R-E Composite*. These teacher characteristics include *Hours of Further Education*, *Advanced Training*, *Years of Education*, *Hours of Work without Children*, and *Supervision*.

Table 5
Regression of KES-R/R-E Composite on teacher characteristics

	β_1	β_2	β_3	β_4	β_5	β_6
Hours of Further Education	.27	.27	.21	.24	.26	.26
Advanced Training	.17	.17	.18	.18	.19	.18
Years of Education	.13	.13	.17	.19	.17	.18
Hours of Work without Children	.05	.04	.02	.14	.14	.13
Supervision	.19	.19	.16	.17	.16	.16
Work Experience	-.03	-.03	-.03	-.09	-.10	
Hours of Work with Children	.03	.03	-.00	.10		
Contractual Working Hours	.24	.25	.25			
Overtime	-.15	-.15				

<i>Number of Types of Further Education</i>	.03						
<i>R</i>	.53	.53	.51	.48	.47	.46	
<i>R</i> ²	.28	.28	.26	.23	.22	.21	
<i>R</i> ² _{corr}	.06	.09	.10	.09	.10	.12	
<i>CI</i>	29.4	27.8	26.5	21.9	18.3	16.1	

Overall regression of KES-R/R-E Composite

Finally, the *KES-R/R-E Composite* was regressed on the selected variables of the four domains. The same stepwise procedure as described above was applied to the 12 predictors. Table 6 shows the results.

Table 6

Overall regression of KES-R/R-E Composite

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
<i>Opening Hours</i>	.34	.34	.30	.30	.33	.31	.29	.24
<i>Average Hours of Care</i>	.26	.26	.29	.28	.32	.25	.26	.32
<i>Count of Children with Disabilities</i>	-.23	-.23	-.23	-.24	-.26	-.24	-.24	-.27
<i>Hours of Further Education</i>	.30	.30	.31	.32	.35	.31	.33	.32
<i>Supervision</i>	.21	.21	.22	.24	.27	.26	.26	.29
<i>Advanced Training</i>	.15	.15	.19	.21	.18	.15	.17	
<i>Materials</i>	.11	.11	.07	.08	.08	.08		
<i>Visits (Museum)</i>	-.15	-.15	-.18	-.20	-.18			
<i>Working Time</i>	.12	.12	.15	.17				
<i>Years of Education</i>	.11	.11	.13					
<i>Balanced Gender Ratio</i>	-.15	-.15						
<i>Hours of Work without Children</i>	.00							
<i>R</i>	.70	.70	.69	.68	.66	.64	.64	.62
<i>R</i> ²	.49	.49	.48	.46	.44	.41	.41	.39
<i>R</i> ² _{corr}	.25	.28	.28	.29	.28	.28	.30	.31
<i>CI</i>	60.0	57.5	55.1	50.5	46.9	42.8	35.3	33.3

In step 8, five variables remain as the most sufficient predictors explaining 31 % of the variance of the *KES-R/R-E Composite*. Included are structural characteristics (*Opening Hours*), compositional characteristics (*Average Hours of Care*, *Count of Children with Disabilities*) and teacher characteristics (*Hours of Further Education*, *Supervision*). Materials and activities do not reach a meaningful influence. All selected variables are of nearly equal importance (in relation to the standardized regression coefficient).

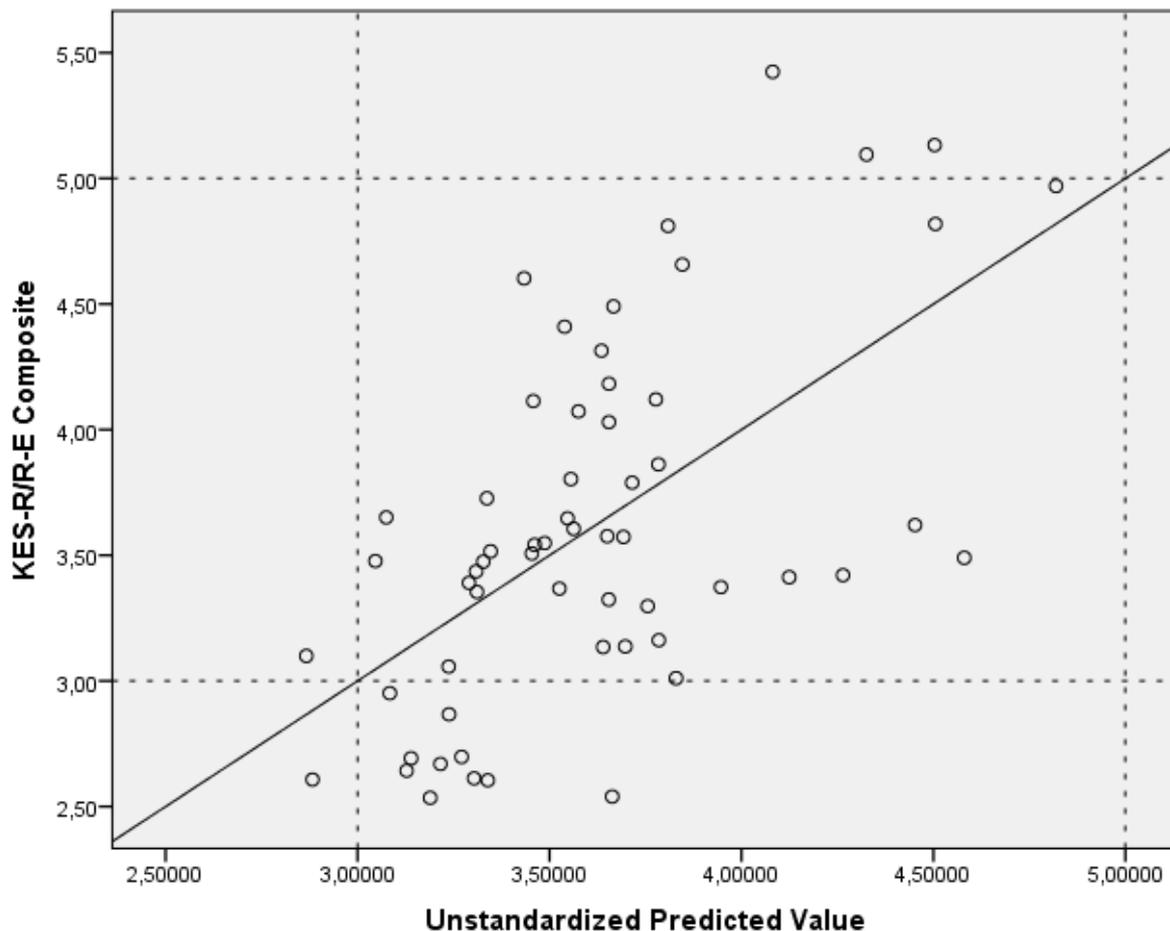


Figure 1. Scatterplot of the unstandardized predicted value of the overall regression with five predictors of *KES-R/R-E Composite*.

As shown in Figure 1 the overall quality of the preschool groups as measured via observation can be predicted quite well by the five selected variables of the teacher questionnaire. There are only few cases that are severely over- or underestimated.

5 DISCUSSION

Before discussing the results, some limitations of the study should be mentioned. Due to the small sample size, we did not use an algorithm exploiting significance of regression coefficients for selection of variables but rather reducing the variable set by collinearity diagnostics. For sake of clarity, regression analyses were done for the four domains of predictors separately. Also, interactions between predictors were not included. All this could lead to a suboptimal selection of variables. However, aim of this paper is mainly the comparison between observation and questioning of preschool quality. Therefore, selection of the best variables of the teacher questionnaire for prediction of observed quality is the main purpose. In a forthcoming NEPS Working Paper we will analyze the data in much more deeper way. Nevertheless, it should be kept in mind that the study is basically explorative and is in need for replication of the results. Another limitation relates to the selection of preschool groups, both concerning the regional context as well as pedagogical approach. Also, selection of two groups in some preschools may lead to dependencies in the data that were not accounted for in the analyses. Generalization of results should therefore be done very carefully. As for the observed quality (i.e. *KES-R/R-E Composite*) as the criteria of the regression analyses

it has to be noted that variability is limited. The upper as well as lower 1.5 points of the theoretical scale are missing and most preschools score in the middle of the distribution. This may lead to restricted regression coefficients and an underestimation of the real associations. Also, discrimination of the preschool groups by quality may be impaired due to the shortened variability.

Bearing these limitations in mind, we will discuss the results in the following, starting with regression on the *structural characteristics*. *Working time* and *Opening Hours* were the most predictive variables. Interpretation is straightforward: The more time for education and care of children on both levels of preschool as such and teachers working hours is available the more educational quality is actualized. There is one limitation to this interpretation: Computing the variable *Working Time*, only the first and second teacher was included although in the questionnaire the full time equivalent of up to four teachers was accounted for. This was due to the fact that the full time equivalent of the third teacher did not correlate with the *KES-R/R-E Composite* and full time equivalent of the fourth teacher even was negatively associated with preschool quality. We interpret this result with the psychological need of pre-school children for a definite attachment figure which is maybe missing if more than two teachers share the group. Another possibility is that the need for a third or fourth teacher gives hints that there are problems within the group that also reduce the quality as a whole. *Group Size* does not yield any effect, maybe surprisingly. Concerning *compositional characteristics*, three out of five variables contribute to the prediction of preschool quality. Whereas *Average Hours of Care* and *Balanced Gender Ratio* have a positive association with *KES-R/R-E Composite*, *Count of Children with Disabilities* correlates negatively. The positive effect of *Average Hours of Care* can be interpreted as before: the time of education and care, now on an individual level, for each child contributes to a positive educational quality. The positive effect of *Balanced Gender Ratio* is interpreted as follows: an appropriate amount of diversity (in this case concerning gender) is associated with a better quality. This is also true for *Count of Children with Migration Background*. Although this variable could not be used in the regression analyses because of its reduced variability, it shows a quadratic relation to the *KES-R/R-E Composite*. That is, a certain amount of these children leads to a better preschool quality than more or less children with migration background in the group. The maximum of the regression line is at 30 % of children with migration background which is in line with previous studies. Unfortunately, this interpretation does not hold for children with disabilities. The negative linear correlation shows that each additional child with disabilities in a group is followed by an impaired quality. This daunting result should in no ways be interpreted causally as a direct effect of presence of such children to the impairment of educational quality. Rather we propose the following interpretation that is in line with what we said before: A child with disabilities is certainly in need for more attentiveness by the teacher than a child without disabilities. This reduces the total amount of education and care time for the group as a whole as well as every (other) child, thus leading to an overall reduced evaluation of preschool quality. This could also be the case with children with migration background. Once their count exceeds an appropriate ratio within a group, more attention of teachers is needed and less time is available for the rest of the children and the whole group. Also, this may generate the need for more teachers caring for the group, contributing to the negative correlation of the fourth teacher to preschool quality (see above).

Looking at *activities and materials* assessment of these variables was not successful in this study and the NEPS pilot study, leading to an adapted response scale in the NEPS main study. As a consequence we cannot make any inferences on the significance of activities to (observed) preschool quality although their theoretical importance is beyond doubt. Nevertheless, availability of *Materials* and *Visits to Museums* could explain 4 % of the *KES-R/R-E Composite*. This result can be interpreted quite easily: The more stimulation (related to education) the children receive the higher the preschool quality.

Not surprisingly for an teacher questionnaire *teacher characteristics* provide the most variables. In the regression analyze five out of ten variables add to the explained variance of 12 %. Interpretation of the contribution of *Hours of Further Education*, *Advanced Training*, *Years of Education*, *Hours of Work without Children*, and *Supervision* is straightforward: the better the qualification of the teacher and the higher the effort for a better qualification by the teacher the higher the preschool quality. This

interpretation is in line with measures of professionalization of preschool teacher's training in Germany within the last years.

The overall regression does not change anything for the aforementioned interpretation except that it selects the most significant predictors. These predictors come from the different domains of the teacher questionnaire (except *activities and materials*, due to their impaired assessment) and are of nearly equal importance, thus point to evidence that there is no single indicator of preschool quality that could be asked for. Overall interpretation of the results refers to two major aspects: (1) hours of education and care a preschool group and every single child receives and (2) quality of teacher's education. The more time a well-trained teacher spends with the children he or she is responsible for the better the educational quality. Furthermore, there are hints that diversity in different aspects (for example gender distribution, children with migration background but also variety of materials and activities) could contribute to the educational quality of a preschool group. The results of the overall regression also show that the observed global educational quality can be reproduced with questionnaire data quite well. There is no massive over- or underestimation, so that the main purpose of this paper can be viewed as fulfilled.

The results of this study, together with the results of the NEPS pilot study, led to improvements of the questionnaire design of the NEPS main study. Response scales were adapted, question texts were altered and in general the most promising questions were selected. Therefore, we are hopeful that the NEPS Scientific Use File of Starting Cohort 2–Kindergarten will contain comprehensive data on preschool quality. The study clearly demonstrates that no single indicator of preschool quality can be expected from questionnaire data but that a whole bunch of variables should be considered that also have to be derived with more or less sophistication from the original questions. All variables in question are not indicators of quality per se but all contribute to a certain extent to quality. In this context we want to stress, that causal inferences of these variables to preschool quality or to further outcomes like children's competencies or competence development should be drawn very carefully. For example the negative relation of *Count of Children with Disabilities* to *KES-R/R-E Composite* does not mean that these children diminish the educational quality of the group or hinder other children in their development but rather show that there are special demands that are not (yet) met in the everyday work of the preschool group.

In a nutshell, it is possible to draw conclusions on educational quality using questionnaire data even from a single contributor who might be under suspicion to euphemize the facts. But it has to be kept in mind that all variables can only give hints to quality and should be derived and interpreted with caution. This at the back of one's mind, the NEPS provides a rich data resource for the analysis of educational quality of preschool institutions and will contribute to highly valuable scientific insights into this issue.

REFERENCES

- Anders, Y. (in press). Stichwort: Auswirkungen frühkindlicher, institutioneller Bildung und Betreuung. *Zeitschrift für Erziehungswissenschaft*.
- Bäumer, T., Preis, N., Roßbach, H.-G., Stecher, L., & Klieme, E. (2011). Education processes in life-course-specific learning environments. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)* (pp. 87-101). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bäumer, T., & Roßbach, H.-G. (2012). Die Familie macht's. Die Bedeutung der Familie für kindliche Bildungsprozesse. *DJI Impulse*, 100(4), 39-41.

- Bühner, M., & Ziegler, M. (2009). *Statistik für Psychologen und Sozialwissenschaftler*. München: Pearson Studium.
- Clausen, M. (2002). *Qualität von Unterricht – Eine Frage der Perspektive?* Münster: Waxmann.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environmental Rating Scale, Revised Edition (ECERS-R)*. New York: Teachers College Press.
- Roßbach, H. G. (2004). Kognitiv anregende Lernumwelten im Kindergarten. In D. Lenzen, J. Baumert, R. Watermann, & U. Trautwein (Eds.), *PISA und die Konsequenzen für die erziehungswissenschaftliche Forschung* (Zeitschrift für Erziehungswissenschaft: Beiheft 3, S. 9-24). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Roßbach, H. G. (2005). Effekte qualitativ guter Betreuung, Bildung und Erziehung im frühen Kindesalter auf Kinder und ihre Familien. In Sachverständigenkommission Zwölfter Kinder- und Jugendbericht (Ed.), *Bildung, Betreuung und Erziehung von Kindern unter sechs Jahren* (S. 55-174). München: Verlag Deutsches Jugendinstitut.
- Roßbach, H. G., Kluczniok, K., & Kuger, S. (2008). Auswirkungen eines Kindergartenbesuchs auf den kognitiv-leistungsbezogenen Entwicklungsstand von Kindern. In H. G. Roßbach, & H.-P. Blossfeld (Eds.), *Frühpädagogische Förderung in Institutionen* (Zeitschrift für Erziehungswissenschaft: Sonderheft 11, S. 139-158). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Roßbach, H.-G., & Tietze, W. (in prep). *Kindergarten-Skala Erweiterung (KES-E)*, unpublished manuscript.
- Sylva, K., Melhuish, E., Sammons, P., & Taggart, B. (2004). *The Effective Provision of Pre-School Education (EPPE) Project: Final Report. A longitudinal study funded by the DfES 1997-2004*. London: Institute of Education, University of London.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing Quality in the Early Years: Early Childhood Environment Rating Scale Extension (ECERS-E): Four Curricular Subscales*. Stoke on Trent, UK and Stirling, USA: Trentham Books.
- Tietze, W., Meischner, T., Gänsfuß, R., Grenner, K., Schuster, K.-M., Völkel, P., & Roßbach, H.-G. (1998). *Wie gut sind unsere Kindergärten? Eine Untersuchung zur pädagogischen Qualität in deutschen Kindergärten*. Neuwied: Luchterhand.
- Tietze, W., Schuster, K.-M., Grenner, K., & Roßbach, H.-G. (2005). *Kindergarten-Skala Revidierte Fassung (KES-R)* (3., überarbeitete Aufl.). Weinheim: Beltz.