



OPEN ACCESS

EDITED BY

Susanne Garvis,
Griffith University, Australia

REVIEWED BY

Ellie Christoffina (Christa) Van Aswegen,
Griffith University, Australia

Bin Wu,

Swinburne University of Technology, Australia

*CORRESPONDENCE

Antonia Elisabeth Enikoe Baumeister

✉ antonia.baumeister@
psychologie.tu-chemnitz.de

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 19 January 2023

ACCEPTED 04 July 2023

PUBLISHED 19 July 2023

CITATION

Baumeister AEE, Jacobsen J and
Rindermann H (2023) The educational and
parenting test for home-based childcare: a
socially valid self-rating instrument.

Front. Educ. 8:1147954.

doi: 10.3389/feduc.2023.1147954

COPYRIGHT

© 2023 Baumeister, Jacobsen and
Rindermann. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

The educational and parenting test for home-based childcare: a socially valid self-rating instrument

Antonia Elisabeth Enikoe Baumeister^{1*†}, Julia Jacobsen^{2†} and Heiner Rindermann¹

¹Professorship of Educational and Developmental Psychology, Institute of Psychology, University of Technology Chemnitz, Chemnitz, Germany, ²Department of Psychiatry, Behavioural Medicine and Psychosomatics, Chemnitz Hospital, Chemnitz, Germany

Introduction: Home-based childcare is increasingly becoming the focus of research, policy and public interest. Self-assessment of quality can increase the social validity of quality improvement efforts among stakeholders. A new online self-assessment tool for parents and non-relative providers of home-based childcare is introduced that has been developed in Germany, the Educational and Parenting Test for Home-Based Childcare (EPT; in German: ‚Bildungs- und Erziehungstest für TagesElternBetreuung BET‘).

Methods: In two studies, the social validity of the EPT was investigated: a stakeholder study with 45 parents and 12 non-relative caregivers, and an expert study with nine experts of child pedagogy. The stakeholders rated the EPT survey ($N=57$) and the subsequent report of test results ($n=22$). The experts evaluated the survey and the feedback report based on vignettes of three fictitious test results (i.e., below average, average, and above average quality). Criteria included face validity, measurement quality, controllability (i.e., comprehensiveness), freedom of response, freedom of pressure, counseling quality, usefulness, control of bias, and privacy protection.

Results: Most aspects of social validity achieved good to very good ratings. All three samples graded the EPT survey as “good.” If the stakeholders felt that their educational quality was undervalued, they rated the report of test results worse ($r_s(20) = 0.52, p = 0.02$). Five of seven experts would recommend the EPT to others.

Discussion: Based on participants’ comments, the instrument was thoroughly revised. The EPT is a socially valid instrument for assessing and developing quality in home-based childcare.

KEYWORDS

home-based childcare, self-assessment, stakeholders, quality, parenting, social validity, acceptance, experts

1. Introduction

There is a broad and multidisciplinary consensus on the lifelong importance of the early years in a person’s life. Three general aims of research on early childhood education and care (ECEC) are to offer children from all social backgrounds a good start in their lives and educational careers, to support parenting as well as parents’ work force and, thereby, to

strengthen the national economy over several generations. Both familial characteristics and features of the providers of childcare are relevant for young children's development (e.g., [Watamura et al., 2011](#)). Research both on regular external day care and on special state-wide programs for children at risk have corroborated the positive long-term effects of ECEC on both individual (e.g., child development) and systemic factors (e.g., a nation's economy; [Barnett and Hustedt, 2005](#); [Bailey et al., 2021](#)).

Research questions of ECEC studies address amongst others, which types of non-maternal childcare are used with which intensity (e.g., center-based childcare, home-based childcare), which effects do they have on children, families and the society, and which role does the quality of childcare play for individual and systemic outcomes (e.g., [NICHD ECCRN, 2002](#)). The present study is dedicated in particular to the following two forms of childcare: (1) home-based childcare provided by day care parents as non-relative caregivers, and (2) maternal or parental childcare. (Non-clinical) behavior problems of young children were sometimes associated with, for example, lower educational quality levels of the external childcare institution (e.g., [NICHD ECCRN, 2002](#); [Sylva et al., 2004](#); [Pianta et al., 2009](#)). Core features of childcare givers' quality are the educational activities subsumed as process quality and the existence of an educational concept with defined aims (e.g., [Rindermann and Baumeister, 2012](#); [Baumeister et al., 2014](#)). While there is an enormous bulk of studies on formal center-based childcare, hardly any studies on home-based childcare and the subcategories of it (i.e., family/relative/non-relative childcare) exist ([Adamson and Brennan, 2016](#); [Blasberg et al., 2019](#)).

1.1. Assessment and development of quality in ECEC

Since the 1980s, variants of basic scales for assessing all aspects of educational quality of different childcare institutions were developed by Thelma Harms, Deborah Cryer, Richard Clifford and colleagues in the United States and translated into several languages. This scale family, known as Environment Rating Scales (ERS), is also used in research and practice in German-speaking countries.

From a scientific view, the advantage of using basic scales worldwide clearly is the possibility of cross-country comparisons of ECEC quality (e.g., [Vermeer et al., 2016](#)). A disadvantage of basic scales could emerge if country-specific characteristics of childcare settings are not considered sufficiently. From the practitioners' view, specific childcare settings (e.g., home-base childcare) may require specific conceptual models for their quality assessment not being addressed by basic scales (e.g., the role of the neighborhood; [Blasberg et al., 2019](#)).

In Germany, day care parents have a maximum number of five children aged 0–3 years in their care, either in their (i.e., the provider's) private home or in apartments especially rented for this purpose, in contrast to the larger (publicly or privately funded) day-care centers (e.g., kindergartens). In 2021, 16% (i.e., $n = 129.406$) of all German children below age 3 in external childcare attended home-based childcare, a decrease of 4% compared to 2020 ([BMFSFJ, 2022](#)). As reasons of this decrease of home-based childcare within the same age group are discussed: demographic changes, a lack of childcare places, and organizational problems of the COVID-19 pandemic.

Since various terms exist in different countries for specific day care arrangements, the following distinctions are used in this article:

In Germany and, therefore, in this article, home-based childcare ("Kindertagespflege") is provided by day care parents ("Tageseltern") in their private homes, whereas family-based childcare is offered by nannies, relatives or au pairs who visit the children in the children's homes. In contrast, in the United States, home-based childcare is an umbrella term, subsuming family, relative and non-relative childcare ([Porter et al., 2010](#)), thus ranging from birth to age 12 ([Blasberg et al., 2019](#)).

By means of the national German study on early childhood education and care (NUBBEK), the quality of home-based and center-based ECEC is assessed for two-year-old ($n = 1,242$) and four-year-old ($n = 714$) children ([Leyendecker et al., 2014](#)). The comprehensive comparison of quality aspects is based on various data sources (observations, interviews, surveys, testing of children). The first data collection wave was conducted from 2010 till 2011 in eight federal lands of Germany, the second wave started in 2021 and is still ongoing. The majority (>80%) of non-relative caregivers achieved a moderate process quality. For home-based childcare provided by non-relative caregivers, process quality scored on average with 4.0 on a scale from 1 (insufficient) to 7 (excellent; [Tietze et al., 2012](#)). Interestingly, the other types of childcare achieved similar quality scores on average, although a large heterogeneity of their characteristics was noted by the NUBBEK research group. For example, home-based caregivers and caregivers of younger children reported of more wellbeing than center-based caregivers and caregivers of older children.

Self-assessments of caregivers increasingly become important for quality improvement efforts. In the past, self-assessments were used to complement external ratings in the context of inspections of day care providers. The acceptability and social validity of quality assessment and improvement methods, however, are contested among some stakeholders, and thus, the sustainability of these quality efforts may be in doubt. Therefore, self-assessment enables caregivers to participate in the process of quality assessment, and it increases the social validity of quality assessment among stakeholders. Social desirability of participants' answers, however, is a challenge for self-assessment instruments. Socially desirable answers can either be prevented, for example, by assuring the participants that the results are confidential. Also, questioning techniques such as the randomized response technique can be applied ([Warner, 1965](#); for limitations and alternatives, see [John et al., 2018](#)). Or socially desirable answers can be detected by a specific measure of socially desirable responding. The latter makes it possible, for example, to weight certain answers differently. In the context of national quality frameworks, instruments for self-assessment and improvement of quality are offered to providers (e.g., for Germany: [Tietze et al., 2017](#); for Australia: [Hadley et al., 2021](#)).

1.2. The educational and parenting test for home-based childcare EPT

The 'Educational and Parenting Test for Home-Based Childcare EPT' (original German name: 'Bildungs- und Erziehungstest für TagesElternBetreuung BET') is introduced as a new instrument for self-assessment and development of ECEC quality ([Baumeister and Rindermann, 2015](#)). The EPT assesses structure, process, orientation and contextual quality of home-based childcare with regard to a

target child aged 0–6 years (Baumeister and Rindermann, 2022). The theoretical quality model of the EPT is based on previous ECCE research (e.g., NICHD ECCRN, 2002). Structural quality aspects are assessed, for example, by asking how the childcare rooms are equipped with various materials for pedagogical activities (e.g., “how many books for children are available?”) and whether the safety and hygiene standards are met. Regarding contextual childcare quality, potential risk and protective factors of child development are asked (e.g., parental educational and income level; in which family constellation does the target child live: does it live with both biological parents?). Parenting styles (i.e., authoritative, authoritarian, permissive and negligent parenting) represent one aspect of orientation quality. In particular, the EPT focuses on the quality of the caregiving process, which is assessed (1) by a wide range of items for different daily routines (e.g., providing healthy food, bedtime rituals) and educational activities (e.g., excursions into nature, practice dressing on their own), and (2) by scales for observing the target child’s interest in some of these activities. Two versions of the EPT exist as online surveys, namely for non-relative caregivers (“Tageseltern”) and for the parents of the target child. The two versions are answered and analyzed independently of each other. It is possible to take the test several times for different target children each time. In its current form, each version of the online survey takes about 20 min, and the items differ according to age of the target child indicated by the caregiver. Consequently, some of the educational activities asked for very young children are different from those asked for 5-year-olds. Potential social desirability of answers is controlled for by using specific indicator items (e.g., “I never get loud when I am upset.”) for the response weighting techniques applied in data analysis. At the end of the survey, participants can express their interest in receiving detailed feedback on the quality of their childcare back. This short written report also includes graphic percentile rank scales for visualizing individual results (Groll, 2017).

In contrast to the aforementioned quality rating scales, the EPT has a stronger research focus; it assesses familial characteristics, for example, in more detail. A further difference between classic environment rating scales and the EPT lies in its self-report nature: Whereas classic scales are used by external raters who visit the children’s homes or day care centers for inspection, the EPT is a web survey which is answered directly by the parents and day-care nannies themselves. For this purpose, two versions of the EPT were developed addressing the two target groups specifically (Baumeister and Rindermann, 2022). The psychometric properties of the EPT are satisfactory (Baumeister and Rindermann, 2022): the average objectivity corresponds to $r=0.58$; the average reliability corresponds to Cronbach’s $\alpha=0.61$. Criterion validity was confirmed in the form of correlations of selected process quality items and scales with conceptually similar tasks of the developmental test ‘IDS-P’ (Grob et al., 2013: e.g., $r(19)=0.53$ for cognitive tasks; $r(35)=0.46$ for psychomotor tasks). Convergent validity was shown in the form of correlations of the parenting style items and scales with the conceptually similar questionnaire ‘EFB’ (Naumann et al., 2010: e.g., $\rho(42)=0.50$ for authoritarian behaviors; $\rho(42)=0.38$ for permissive behaviors). In addition, convergent validity was also confirmed with respect to associations between the EPT and the German version of the Leuven Involvement Scale for Young Children LES-K (Laevers et al., 1993). These two instruments assess child involvement (LES-K)

or child interest (EPT) in various educational activities. An average correlation of $r(62)=0.80$ ($p<0.01$) was obtained between the EPT and the LES-K for self-ratings and two foreign ratings across several activities (Gosmann, 2017). The psychometric properties of the EPT are constantly being investigated and improved in the context of bachelor’s and master’s theses recruiting independent samples of caregivers and parents.

It is an open research question how regulated childcare providers differ from parents regarding several aspects of their childcare quality (Porter et al., 2010; Baumeister et al., 2017a,b). The EPT aims at informing these two target groups directly about different aspects of the quality of their self-reported childcare features. Thus, a high level of participation in quality assessment and development of the stakeholders is realized. It can be assumed that participatory instruments and procedures will raise more acceptance of quality results and of the necessity to further develop quality compared to more ‘expertocratically’ applied procedures, that is, external ratings and requirements (Baumeister et al., 2017a). Acceptance of a self-assessment instrument by the stakeholders means that they regard this instrument, amongst others, as valid, reliable, and useful (Zimmerhofer, 2008). The term acceptance is often used as a synonym for social validity (Kersting, 2008). Kersting (2008), however, points out that the psychometric validity of an instrument is independent of its acceptance. A high acceptance is a prerequisite for using a self-assessment instrument frequently (Zimmerhofer, 2008). In addition, for regular application it is also important that the feedback provided by the instrument, that is, the report of the individual test results in this study, is accepted by the stakeholders. If the feedback includes a critique of the parenting behavior, for example, it is crucial to find out whether stakeholders feel threatened by this feedback (Landes and Laufer, 2013) or whether it helps them to improve their practices.

Moreover, the EPT applies a state-of-the-art method to measure quality in home-based childcare, because the assessment is based on interactions between the adult and a focal child instead of global assessments (cf. Porter et al., 2010). Consequently, the EPT allows for more specific quality development strategies tailored to individual children’s needs. Two specific strengths of the EPT as an assessment instrument are, for example, that (1) parenting styles are identified, and that (2) a target child’s interest in different educational activities is observed. Thus, the EPT integrates three interdependent concepts (i.e., ECEC quality, parenting styles, child interests) in one instrument.

1.3. Research questions

The aim of this evaluation study was to investigate how (1) the EPT survey, (2) the resulting report of test results, and (3) the entire procedure consisting of online survey and digital report of test results are evaluated (1) by the participating parents and non-relative caregivers as stakeholders and (2) by independent experts of childcare. Thereby, the evaluation should consist of school grades and a measure of social validity. The first research question was whether the survey, the report of test results and the entire procedure will be evaluated as “good” in terms of school grades. This criterion was set, because the pilot version of the EPT survey achieved an average rating of $M=2.60$ ($SD=0.83$) by 15 stakeholders. Thus, the revised

version was supposed to score better. In the pilot study, the report of test results achieved a good grade by the stakeholders ($N=7$, $M=2.00$, $SD=0.58$; Baumeister et al., 2017a). The second research question explored whether the social validity of these components will be evaluated as good. The third research question addressed the problem of dropouts (i.e., part of a sample terminates study participation prematurely) by testing whether evaluations differ between persons who drop out and persons who complete this evaluation study. The following two time points were considered in the analysis of dropouts: (1) after completing the EPT online survey, that is, participants who were not interested in receiving the report of their individual test results, and (2) after receiving the report of individual test results, that is, participants who did neither evaluate the report nor the entire procedure. According to the fourth research question, it was examined whether stakeholders responded to poorer test scores with a lower rating both of the report and of the entire procedure.

2. Methods of stakeholder study

2.1. Participants

Fifty-seven persons filled out the EPT including its evaluation (mean age groups: caregivers 40–49 years old, parents 30–39 years old; only 2 persons with a migration background), whereby 12 participants indicated to be non-relative caregivers. With 45 female participants (including 10 female non-relative caregivers) and nine male participants (including one non-relative caregiver), males were underrepresented in the sample. The target children that were focused on in the quality survey ranged in age from 2.18 to 7.01 years ($M=4.09$, $SD=1.28$, $N=53$). Twenty-two persons (including seven non-relative caregivers) participated in the final evaluation after the test results were reported back.

2.2. Measures

Social validity was assessed by means of the “Akzept!-P” questionnaire by Kersting (2015). The “Akzept!-P” measures the acceptance of personality questionnaires. Two additional scales were used: (1) the scale “Counseling Quality” (Zimmerhofer, 2008), and (2) the newly designed scale “Usefulness.” The complete assessment of social validity comprised 54 items that were divided into the following three parts: (1) the evaluation of the EPT survey, (2) the evaluation of the report of test results, and (3) the final evaluation of the entire procedure (EPT survey and report of results). A few items appeared both in the parts (1.) and (2.) in case that some participants did not wish to receive a report of their test results. In this way, participants who did not complete the entire procedure could still evaluate the EPT survey at least.

The wording of the original questionnaire by Kersting (2015) was adapted to the study so that the evaluation questions referred either to “the questionnaire,” “the report of results” or to “the entire procedure.” The scales are briefly introduced in the following sections:

- *Controllability.* This scale explored whether participants understood the various instructions, questions and the report of

test results and whether they knew how to proceed during the test (sample item: “The questions were clearly understandable.”).

- *Freedom of pressure.* This scale indicated the extent to which participants were over- or under-challenged by the procedure or by individual components (sample item: “The report of test results lacks detail.”).
- *Face validity.* Face validity is high if the subjectively perceived intention of the measurement corresponds highly with the diagnostic question the participants were informed about (sample item: “The results reflect tasks that are required in early childhood education and care.”).
- *Freedom of response.* This scale examined whether participants could express their attitudes or behavior accurately by means of the given response alternatives (sample item: “Sometimes, I could not state the information that I wanted.”).
- *Measurement Quality.* This scale captured whether the participants thought that the test could accurately represent existing differences between individuals regarding their educational and parenting behavior (sample item: “The analyzed results can convey a correct impression of a person’s educational and parenting behavior.”).
- *Usefulness.* Did participants perceive the procedure as helpful and would they recommend it to other persons (sample item: “I would recommend this test to other parents/caregivers.”)?
- *Counseling quality.* This additional scale assessed how satisfied participants were with the feedback provided by the EPT as a tool for educational counseling (sample item: “I am still not clear on my strengths and weaknesses.”).
- *Privacy protection.* This scale addressed the extent to which participants felt that their privacy was violated by the questions asked (sample item: “I think that the topics addressed in the test are far too personal and intimate.”).
- *Control of bias.* Did participants present themselves to be better or worse than they actually were (sample item: “I presented myself to be better than I am.”)? Did the report of test results show participants in the wrong light?

All items were answered on seven-point Likert scales ranging from 1 (“strongly disagree”) to 7 (“strongly agree”). Of all intermediate scale points, only scale point four (“neutral”) has a verbal label. For the overall assessment of the questionnaire, the report of test results and the entire procedure, the participants were asked to provide school grades as used in German public elementary schools, ranging from “very good” (grade 1) to “insufficient” (grade 6). At the end of the acceptance survey, participants had the option to write free comments.

2.3. Procedure

Participants were recruited on four ways: (1) *via* events organized by the Federal Association for Child Day Care, (2) *via* mail distribution lists of the various state associations for child day care, (3) *via* flyers and posters posted up in kindergartens, in educational and family counseling centers, at pediatricians, and (4) *via* advertising in social media. Completing this former version of the EPT questionnaire took about 30–45 min. Directly after the participants completed the EPT survey, they received the first

questions for evaluating the social validity of the survey. In the next step, the participants received a report of their individual test results consisting of about 16 pages. The individual test results regarding structure, process, orientation and contextual quality were classified on percentile rank scales (PRS). For this purpose, the report contained both verbal descriptions of the results (e.g., for PR above 84: “It is particularly important for you to foster your child’s education.”) and visual scales. The educational activities were evaluated, and the target child’s interest in specific educational activities was reported (e.g., “you motivate the child to solve mathematical tasks regularly,” and “according to your observation, the child shows a medium interest in mathematical tasks.”). As in the survey, the report of test results also always focused on one target child. Recommendations for and examples of specific educational activities were provided whenever the individual results fell in the below average to average range (i.e., PR below 16 to 84). Background reading tips were also included. Amongst others, the parenting style was described (i.e., which aspects were more or less prominent: authoritative/authoritarian/permissive/neglectful parenting). Examples of the individual familial risk and protective factors were listed (e.g., “protective factors of your family are that both biological parents live together, that the child received breast feeding, the school and professional education of the parents” etc.). The balance of individual protective and risk factors was shown on a visual scale with verbal labels (i.e., from -3 = “risk factors predominate” over 0 = “risk factors and protective factors are balanced” to $+3$ = “protective factors predominate”). At the end of the report of individual test results, parents and caregivers were asked to evaluate the social validity of the report of test results and of the entire procedure (EPT survey and report of results) by means of an online survey. Participation in this final online survey took about 15 min.

A major problem of online studies is the dropout rate of participants, especially if the dropout is selective (Zhou and Fishbach, 2016). To control for systematic associations between dropout or rejection of one’s report of test results and the participant’s rating of social validity, the evaluation of survey, report of results and the entire procedure was split into the parts described above. In this way, differences in the social validity ratings could be assessed between persons who completed the survey only and those who were interested in receiving the report on their results.

In addition, to also assess whether technical problems occurred (e.g., internet failure), at the end of each survey page, participants were asked whether they would like to continue the test. In case they did not wish to continue, participants were forwarded to the survey evaluation where they were asked to indicate the reasons for abandoning the survey (i.e., “technical problems,” “questions too personal,” “lack of time,” “questions seemed unsuitable or inappropriate for the purpose of the project,” “other”).

3. Methods of expert study

3.1. Participants

The social validity of the EPT procedure was also evaluated by experts of childcare. A total of 9 experts (1 male) participated (mean

age group: 35–45 years old). One person provided her written assessment but did not complete the questionnaire, thus, only her comments were included. For the present study, experts were defined as persons who worked both with children and with parents in their everyday professional lives and, if necessary, advised them. Also included in the group of experts were persons who were professionally involved in educational science and related occupational fields. The group of experts consisted of 3 counselors of an educational and family counseling service, 2 teachers, 2 scientists of educational sciences, and 2 workers in remedial professions with treatment of children and adolescents. Each expert received the EPT survey for inspection together with an exemplary report of test results for parents or non-relative caregivers. In order to keep effort for the experts as low as possible, vignettes of test results were presented, varying the reported educational quality on three levels (i.e., above average, average, below average; see Table 1). In these vignettes, the target child was 3.5 years old and was cared for by a female. Child gender varied.

3.2. Measures

Similarly to the stakeholder study, the experts’ ratings of social validity were also based on adapted items of the “Akzept!-P” questionnaire by Kersting (2015), supplemented by the scales “Counseling Quality” (Zimmerhofer, 2008) and “Usefulness.” Mostly the same items were used as those used in the evaluation done by parents and non-relative caregivers. Whenever necessary, items were adapted in their wording so that they were stated from the experts’ point of view. The questionnaire for the experts included 42 items, representing the same scales of face validity, controllability, quality of measurement, freedom of pressure, usefulness, freedom of response, counseling quality and privacy protection. Moreover, the experts were also asked to give school grades for evaluating the EPT survey, the report of test results and the entire procedure.

3.3. Procedure

The experts were invited to participate by e-mail or *via* direct conversation. Each participating expert received access to the online evaluation questionnaire, a sample EPT questionnaire for parents or daycare parents, and a sample report of test results for parents or non-relative caregivers, as described earlier. The experts could choose between electronic or printed documents. In addition to the standardized evaluation questionnaire, the experts were also asked to write comments directly in the documents they had received.

4. Results

Fifty seven persons completed the EPT including the evaluation of the instrument. Only 22 of the 57 participants (39%) continued to evaluate the EPT procedure by providing their feedback regarding the report of their individual EPT test results that they had obtained and read through. In the first analysis step,

TABLE 1 Vignettes of the report of test results used in the expert study (Excerpt).

Evaluator of the report of test results	Expert #1	Expert #2	Expert #3	Expert #4	Expert #5	Expert #6
Target person of the report of test results	Parent	Parent	Parent	Non-Relative Caregiver	Non-Relative Caregiver	Non-Relative Caregiver
Gender of target person	f	f	f	f	f	f
Gender of target child	m	f	m	f	m	f
Age of target child	3.5 years					
Care and nutrition	-∅	∅	+∅	-∅	∅	+∅
Promotion of education	+∅	-∅	∅	+∅	-∅	∅
Facilities	∅	+∅	-∅	∅	+∅	-∅
Explanations to the Child	-∅	∅	+∅	-∅	∅	+∅
Promotion of verbal development	+∅	-∅	∅	+∅	-∅	∅

Gender: f= female, m = male. Fictitious test results of the aspects of educational quality on percentile rank scales (PRS): +∅ = above average, ∅ = average, -∅ = below average.

TABLE 2 Distribution of German school grades in the stakeholder study and in the expert study.

Grade	EPT survey			Report of test results			Entire procedure		
	Parents (n = 44)	Non-relative caregivers (n = 11)	Experts (n = 8)	Parents (n = 15)	Non-relative caregivers (n = 7)	Experts (n = 8)	Parents (n = 15)	Non-relative caregivers (n = 7)	Experts (n = 7)
1 (very good)	5 (11%)	0 (0%)	1 (13%)	3 (20%)	0 (0%)	2 (25%)	3 (20%)	0 (0%)	1 (14%)
2 (good)	30 (68%)	9 (82%)	4 (50%)	6 (40%)	3 (43%)	4 (50%)	8 (53%)	2 (29%)	5 (71%)
3 (satisfying)	8 (18%)	2 (18%)	3 (38%)	5 (33%)	3 (43%)	1 (13%)	3 (20%)	4 (57%)	1 (14%)
4 (fair/pass)	1 (2%)	0 (0%)	0 (0%)	1 (7%)	1 (14%)	0 (0%)	0 (0%)	1 (14%)	0 (0%)
5 (deficient)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (13%)	1 (7%)	0 (0%)	0 (0%)
6 (fail)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Percentage within subgroup given in parentheses.

the representativeness of the participating parents and caregivers was estimated by comparing their features (e.g., highest professional degree) to those of the average German population according to the German federal office of statistics (Destatis, 2017). In this way, parental features (e.g., number of children, family status: single parent vs. married biological parents) were compared to average German families, and caregivers' features to average German providers of home-based childcare. The participating parents showed a higher educational level compared to average German parents. For example, 27% of the participating mothers owned a University degree, whereas only 9% did in the German population. Other familial features were similar to those of the German population. Similarly, the caregivers showed a higher educational level compared to average providers of home-based childcare in Germany. For example, 44% of the participating caregivers had graduated from vocational schools in contrast to only 30% of the population of German caregivers.

Due to the small sample size, participants' responses to several of the items evaluating the EPT did not follow the normal distribution. Therefore, factor analyses were not conducted and instead, relatively robust measures like Spearman correlations, *t*-Tests or Welch-Tests (in case of no homogeneity of variance) were preferred (Sedlmeier and Renkewitz, 2008). All data were analyzed without exclusions of outliers. Missing data were not imputed. The distribution of German school grades is shown in Table 2. Supplementary Tables 1 (stakeholder study) and 2 (expert study) provide an overview of the descriptive statistics of the social validity items.

4.1. Properties of items and scales

Correlations in small samples (i.e., below *N*=250) usually are unstable (Schönbrodt and Perugini, 2013). Therefore, items were

TABLE 3 Test of grading of the EPT survey against criterion grade 2.

	N	M	SD	95% Confidence interval of mean value		T	df
				Lower bound	Upper bound		
Parents	44	2.11	0.62	1.93	2.30	1.22	43
Non-relative caregivers	11	2.18	0.41	1.91	2.45	1.49	10
Experts	8	2.25	0.71	1.66	2.84	1.00	7
Total	63	2.14	0.59	1.99	2.29	1.92	62

No statistically significant differences (two-tailed test).

TABLE 4 Test of grading of the report of test results against criterion grade 2.

	N	M	SD	95% Confidence interval of mean value		T	df
				Lower bound	Upper bound		
Parents	15	2.27	0.88	1.78	2.76	1.17	14
Non-Relative Caregivers	7	2.71	0.76	2.02	3.41	2.50*	6
Experts	8	2.25	1.28	1.18	3.32	0.55	7
Total	30	2.37	0.96	2.01	2.73	2.08*	29

*Statistically significant at $p < 0.05$ (two-tailed).

aggregated to scales according to their conceptual similarity, whenever the scales met both of the following criteria: (1) Cronbach's alpha should be equal or larger than 0.70, and (2) the discriminatory power of the items should be equal or larger than 0.50. [Supplementary Tables 3](#) (stakeholder study) and [4](#) (expert study) provide an overview of the properties of the social validity scales.

In the stakeholder study, none of the conceptually built scales for evaluating the EPT survey achieved a sufficiently high level of Cronbach's alpha, in contrast to the expert study. Therefore, only the results of homogeneous scales are reported in the following.

4.2. Evaluation of the EPT survey

Regarding the first research question, all subgroups and the total sample of this study graded the EPT survey on average with the grade 2, which is interpreted as "good" in the German educational system (see [Table 3](#)). Thus, this target criterion was fulfilled by the EPT survey. Further, no differences in grading were found between the three subgroups of parents, non-relative caregivers and experts, Welch-Test $F(2, 15.82) = 0.18$, ns.

According to the second research question, the ratings of the social validity (i.e., including the reversed items) on average should achieve values above 5 till 7 in order to be good or very good. Since the internal consistencies of the conceptually built scales were too low in the stakeholder study, the second research question could

not be answered on the scale level. Seven items of the 11 social validity items achieved high values in the subsamples of parents and non-relative caregivers, and four items showed a neutral position (see [Supplementary Table 1](#)). Therefore, a good to very good social validity was obtained among stakeholders for the majority of assessed aspects. As an example of a neutral position, participants were unsure whether a good self-assessment could be achieved.

The eight experts evaluated the "Face Validity" of the EPT survey to be good ($M = 5.03$, $SD = 1.15$), and its "Controllability" to be very good ($M = 6.38$, $SD = 0.90$). The rest of the internally consistent scales showed mediocre ratings, namely for "Freedom of Response" ($M = 4.83$, $SD = 1.46$), and for "Measurement Quality" ($M = 4.92$, $SD = 1.40$). In addition, the experts indicated a neutral position regarding 6 items (see [Supplementary Table 2](#)). They suspected that the survey responses could diverge from the person's actual behavior towards children. Moreover, the experts doubted that participants presented themselves to be better than they were in the survey.

In the following, only the most frequently stated strengths and weaknesses of the EPT are reported: 86% of parents and non-relative caregivers and all of the 8 experts agreed that the questions of the EPT survey reflect tasks that are required in early childhood education and care. 63% of parents and caregivers would recommend the EPT survey to others. 11% criticized that the survey contained too many questions. All subgroups criticized that some items were not appropriate for younger children, for example, fostering of reading with 3-year-olds.

4.3. Evaluation of the report of test results

The report of individual test results was seen more critically, as it failed to reach the criterion grade 2 both in the total sample and in the subsample of non-relative caregivers (see [Table 4](#)). Again, the three subsamples did not differ in their evaluations of the report of test results, Welch-Test $F(2, 13.59) = 0.79$, ns.

In the stakeholder study, the only internally consistent scale of "Controllability" achieved good ratings on average ($N = 22$, $M = 5.46$, $SD = 1.33$). Further, no differences emerged between the subgroups of stakeholders' ratings of "Controllability," Welch-Test $F(1, 11.86) = 0.28$, ns. 13 of the 18 items evaluating the report of test results achieved good to very good ratings of their social validity by the stakeholders. Neutral positions were obtained regarding four items. Both the parents and the non-relative caregivers, however, criticized that the report presented them worse than they actually were, $M = 3.57$ ($N = 21$, $SD = 1.66$).

All experts evaluated the following aspects to be good on average: "Freedom of Pressure" ($M = 5.37$, $SD = 1.24$), "Face Validity" (2 items, $M = 5.81$, $SD = 1.46$), and "Counseling Quality" ($M = 5.75$, $SD = 1.02$). Mediocre ratings were provided for "Control of Bias" on average ($N = 6$, $M = 3.08$, $SD = 1.28$). In contrast to the stakeholders, the experts feared more that the participants presented themselves to be better than they were in the survey. Moreover, the experts had a neutral position towards both negatively biased survey responses and negatively biased test results.

91% of the parents and caregivers indicated that the length of the report of test results was okay for them. 27% criticized that reading and understanding the report was stressful. 41% indicated that the report of test results encouraged them in their parenting behavior. All

TABLE 5 Test of grading of the entire procedure (i.e., EPT survey and report of test results) against criterion grade 2.

	N	M	SD	95% Confidence interval of mean value		T	df
				Lower bound	Upper bound		
Parents	15	2.20	1.01	1.64	2.76	0.76	14
Non-Relative Caregivers	7	2.86	0.69	2.22	3.50	3.29*	6
Experts	7	2.00	0.58	1.47	2.53	0.00	6
Total	29	2.31	0.89	1.97	2.65	1.88	28

*Statistically significant at $p < 0.05$ (two-tailed).

subgroups criticized that the analysis of fostering reading and writing was not age appropriate for 3-year old children. In their opinion, the report was too strict regarding these areas of support.

4.4. Evaluation of entire procedure

The entire procedure, consisting of the EPT survey and the report of individual test results, failed to reach the criterion grade 2 in the subsample of non-relative caregivers (see Table 5). The average grades tended to differ between the three subsamples, Welch-Test $F(2, 15.15) = 3.16, p = 0.07$. Experts gave the highest ratings on average, and non-relative caregivers gave the lowest ratings (cf. Table 5).

Both subgroups of stakeholders took a neutral position towards “Face Validity” ($M = 4.62, SD = 1.29$; Welch-Test $F(1, 13.52) = 0.20, ns$), “Measurement Quality” ($M = 4.63, SD = 1.27$, Welch-Test $F(1, 12.33) = 0.42, ns$), and “Usefulness” of the entire procedure ($M = 4.74, SD = 1.32$, Welch-Test $F(1, 15.44) = 0.69, ns$; see Supplementary Table 1). In contrast to these aspects, non-relative caregivers evaluated “Freedom of Response” more critically ($N = 6, M = 3.44, SD = 1.60$) than parents ($N = 15, M = 5.24, SD = 1.43$), Welch-Test $F(1, 8.41) = 5.74, p = 0.04$.

The scale of “Usefulness” provided a summative evaluation of the entire procedure by the experts. On average, they rated “Usefulness” of the procedure to be good ($N = 8, M = 5.53, SD = 1.20$).

68% of the parents and caregivers regard the procedure as a good counseling instrument in questions of educational and parenting behavior towards young children. 55% of the participating stakeholders would further recommend this procedure. Seven of the eight experts believe that the procedure helps stakeholders to become clearer about their educational and parenting behavior and their goals. Five experts would further recommend the entire procedure.

4.5. Dropout control for the evaluation of the EPT survey

Only 8 parents (i.e., 14% of the participating stakeholders) were not interested in receiving the report of test results. These parents rated the EPT survey on average with the grade 2.0 ($SD = 0.54$), whereas the further participating stakeholders rated the EPT on average with the grade 2.14 ($N = 36, SD = 0.64$). This marks a small and statistically non-significant group difference of $d = 0.23, T(42) = 0.57$,

ns. It can be concluded that stakeholders that are slightly more critical continued this evaluation study.

4.6. Associations between test results and social validity

It was tested whether stakeholders whose test results were below average (i.e., PR scores of 0 till 15.9 coded by 1) would grade the report or the entire procedure worse compared to stakeholders with average or above average test results (i.e., PR scores of 16 till >84 coded by “0”). Only small and non-significant associations emerged both between the report of test results and its grading, $r_s(21) = -0.15$ (ns), and between the test results and grading the entire procedure, $r_s(21) = -0.19$ (ns). Moreover, no association was found between below average test results and agreeing with the item “I feel that the test results present me worse than I actually am.” Those who agreed with this item, however, more often gave worse grades to the report of test results, $r_s(20) = 0.52, p = 0.02$. Thus, the subjective feeling of being undervalued affected the ratings more than the percentile rank of test results achieved.

4.7. Dropout control for the evaluation of the report of test results

Again, it was tested whether those stakeholders who received their test results and who refused to evaluate the report and the entire procedure differed from those stakeholders who completed the entire evaluation study. Those who dropped out had achieved 5.30 below average test results on average ($N = 27, SD = 5.34$), whereas those who completed the last evaluation had achieved 4.68 below average test results on average ($N = 22, SD = 4.11$). This again marks a small and statistically non-significant group difference of $d = 0.13, T(47) = -0.44, ns$. Therefore, the evaluation of the report and of the entire procedure probably was not biased by below average test results.

5. Discussion

In this study, a new procedure for the self-assessment and development of ECEC quality, the ‘Educational and Parenting Test for Home-Based Childcare EPT’ and its report of test results, was evaluated by providers of home-based childcare and parents as stakeholders and by experts of child pedagogy. For this purpose, the experts received vignettes of three fictitious test results of quality (i.e., below average, average, and above average quality). The evaluation included grades and rating scales assessing the acceptance of the procedure (Kersting, 2005).

While the survey achieved a good grade on average, the report of test results failed to reach this criterion. The experts evaluated the entire procedure consisting of survey and report of results to be better than it was seen by the non-relative caregivers. Particularly if participants felt that their educational quality was undervalued in the report of test results, they rated the EPT worse ($r_s = 0.52, p = 0.02$). This pattern of findings can be explained by the psychological phenomenon that humans strive for self-confirmation and accept critique less in general (Ilgen et al., 1979). Beside this item of feeling

undervalued by the test results, only small associations between test results and stakeholders' evaluation of the report were found. Thus, stakeholders evaluate the instrument mostly independently of their own results. In addition, most aspects of measurement quality, controllability, freedom of response, freedom of pressure, counseling quality, and usefulness achieved good to very good ratings both by the stakeholders and by the experts. Therefore, it can be concluded that the EPT is a socially valid instrument that should be further optimized and applied for assessing and developing quality in home-based childcare.

Doubts were raised regarding the face validity, because stakeholders and experts deemed it possible that respondents' daily practices with children diverge from their self-assessments in the EPT. In addition, regarding the report of test results, experts provided mediocre ratings for "Control of Bias" because they feared that the participants presented themselves to be better than they were in the survey. In contrast to the experts, both the parents and the non-relative caregivers criticized that the report presented them worse than they actually were. The reported findings all represent plausible criticisms in face of a self-assessment instrument. Moreover, these findings illustrate how differently the procedure, consisting of online survey and subsequent report of individual test results, is perceived by the different subgroups of stakeholders. Consequently, several strategies were developed for addressing the different needs of stakeholders: Both the introduction to the online survey and to the report of test results were thoroughly revised to include motivating statements: That is, the participants are commended for actively and confidently dealing with sensitive and personal issues related to quality development in ECCE. In the survey, participants are assured that the results will be kept confidential. In the report of test results, arguments were added for why it is important and useful to repeatedly assess pedagogical quality in ECCE; for example, in order to become clearer about one's goals, strengths and weaknesses related to educational and parenting behavior. Further advice is offered in the report of test results and advice centers in Germany are named. In addition, a social desirability scale was included in all following versions of the EPT. This scale allows to identify and downgrade unrealistically positive answers.

In addition, comments of all participants were collected in this evaluation study. A large consensus was found that the item on fostering reading and writing was not age appropriate for 3-year-old children. Subsequently, the instrument was thoroughly revised integrating all comments received. For example, in a further study, day-care nannies assigned items to specific age groups of children. In this way, six variants of the EPT were developed differing for age groups (i.e., for children aged 0–1, 1–2, 2–3, 3–4, 4–5, 5 and older). In the next evaluation study, it is tested whether these revisions result in improved ratings of the survey and its report of results by the stakeholders. In addition, a workshop for the professional development of home-based care providers was worked out that will accompany future EPT surveys to support sustained quality development (Hamm et al., 2005).

A limitation of both the stakeholder study and the expert study was the selective sample with its small size of 45 parents, 12 day-care nannies and 9 participating experts. Sufficiently large samples are needed, amongst others, for stable correlations between the social validity items (Schönbrodt and Perugini, 2013). Consequently, factorial analyses were impossible. Moreover, as a

reviewer stated, the small sample size was compounded by the number of dropouts. Dropouts are however to be expected in a study of this kind where, for example, parents avoid being confronted with critique.

To improve the comprehensibility of the EPT, plain language will be implemented in the next version of the survey and report on test results. In addition, as a reviewer noted, freedom of pressure of the report of test results needs to be reduced for the diverse backgrounds of families. This recommendation is supported by a supplementary analysis revealing that mothers with lower professional degrees rated the report to be more stressful than mothers with higher professional degrees, $r(10) = -0.62$, $p = 0.03$. Due to the highly selective sample (i.e., variance restrictions), no further correlations between education degrees and the social validity scales were found.

Overall, despite the limitations, the EPT provides a good starting point for assessing and developing quality in home-based childcare, a popular form of childcare. For example, the usefulness of the EPT procedure (i.e., survey and report of test results) was recognized by the majority of experts in this study. The majority of parents and day-care nannies recognized the counseling quality of the EPT procedure. All subgroups would further recommend the procedure to other parents or caregivers. As the reviewers of this contribution noted, further applications and development of the EPT should address, for example, the question whether the EPT is culturally sensitive to the diverse backgrounds and practices of different families and caregivers.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants.

Author contributions

AB and HR contributed to conception and design of the study. AB organized the materials and performed additional statistical analyses. JJ further developed the materials of the study and collected the data. JJ performed the initial statistical analyses and was supervised by AB and HR. All authors contributed to the article and approved the submitted version.

Funding

The publication of this article was funded by Chemnitz University of Technology and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 491193532.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

- Adamson, E., and Brennan, D. (2016). Return of the nanny: public policy towards in-home childcare in the UK, Canada and Australia. *Soc. Policy Adm.* 51, 1386–1405. doi: 10.1111/spol.12250
- Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: the long-run impacts of head start on human capital and economic self-sufficiency. *Am. Econ. Rev.* 111, 3963–4001. doi: 10.1257/aer.20181801
- Barnett, W. S., and Hustedt, J. T. (2005). Head Start's lasting benefits. *Infants Young Child.* 18, 16–24. doi: 10.1097/00001163-200501000-00003
- Baumeister, A., and Rindermann, H. (2015). Bildungs- und Erziehungstest für TagesElternBetreuungBET. Available at: <https://www.tu-chemnitz.de/bet>
- Baumeister, A., and Rindermann, H. (2022). "Der Bildungs- und Erziehungstest für TagesElternBetreuung (BET) in Forschung und Praxis [Abstract]" in *Kongress der Deutschen Gesellschaft für Psychologie (S. 118)*, in Hildesheim (Germany). eds. C. Bermeitiger and W. Greve, vol. 52 (Alfeld/Leine: Leineberglanddruck)
- Baumeister, A. E. E., Rindermann, H., and Barnett, W. S. (2014). Crèche attendance and children's intelligence and behavior development. *Learn. Individ. Differ.* 30, 1–10. doi: 10.1016/j.lindif.2013.11.002
- Baumeister, A., Schulze, H., and Rindermann, H. (2017a). "Qualitätsentwicklung sowie deren Akzeptanz bei Eltern und Tageseltern [Abstract]" in *Gemeinsame Tagung der Fachgruppen Entwicklungspsychologie und Pädagogische Psychologie PAEPSY 2017*. eds. R. Broomme, S. Dutke, M. Holodynski, R. Jucks, J. Kärtner and S. Pieschlet al. (Münster (Germany): Universität Münster, Institut für Psychologie, Institut für Psychologie in Bildung und Erziehung), S.342–S.343.
- Baumeister, A., Stock, S., and Rindermann, H. (2017b). Participatory educational quality development for parents and day care nannies [abstract]. In S. Branje and W. Koops (Hrsg.), *18th European Conference on Developmental Psychology ECDP 2017*. Utrecht, Netherlands: Utrecht University.
- Blasberg, A., Bromer, J., Nugent, C., Porter, T., Shivers, E. M., Toneyan, H., et al. (2019). "A conceptual model for quality in home-based child care. OPRE Report #2019–37" in *Office of Planning, Research and Evaluation, Administration For Children and Families* (Washington, DC: U.S. Department of Health and Human Services)
- BMFSFJ (2022). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2021*. Berlin: Federal Ministry for Family Affairs, Senior Citizens, Women and Youth.
- Destatis (2017). Familien und Familienmitglieder mit minderjährigen Kindern in der Familie. Available at: https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/HaushalteFamilien/Tabellen/2_6_Familien.html. Retrieved March 27, 2017.
- Gosmann, S. (2017). Vergleich der Interessantheitskala des Bildungs- und Erziehungstests für TagesElternBetreuung BET mit der Leuener-Engagiertheitsskala für Kinder LES-K (Comparison of the Interestingness Scale of the EPT with the LES-K). unpublished bachelor's thesis, Chemnitz University of Technology.
- Grob, A., Reimann, G., Gut, J., and Frischknecht, M.-C. (2013). "Intelligence and development scales" in *Preschool (IDS-P). Intelligenz- und Entwicklungsskalen für das Vorschulalter* (Bern: Verlag Hans Huber)
- Groll, J. (2017). Neuevaluierung des Bildungs- und Erziehungstests für TagesElternBetreuung (BET) (Re-evaluation of the EPT). Unpublished Master's thesis, Chemnitz University of Technology.
- Hadley, F., Harrison, L. J., Irvine, S., Barblett, L., Cartmel, J., and Bobongie-Harris, F. (2021). *Discussion paper: 2021 national quality framework approved learning frameworks update*. Sydney: Australian Children's Education and Care Quality Authority (ACECQA).
- Hamm, K., Gault, B., and Jones-DeWeever, A. (2005). *In our own backyards: Local and state strategies to improve quality in family child care*. Washington, DC: Institute for Women's Policy Research.
- Ilgel, D. R., Fisher, C. D., and Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *J. Appl. Psychol.* 64, 349–371. doi: 10.1037/0021-9010.64.4.349
- John, L. K., Loewenstein, G., Acquisti, A., and Vosgerau, J. (2018). When and why randomized response techniques (fail to) elicit the truth. *Organ. Behav. Hum. Decis. Process.* 148, 101–123. doi: 10.1016/j.obhdp.2018.07.004
- Kersting, M. (2008). Zur Akzeptanz von Intelligenz- und Leistungstests. *Rep. Psychol.* 33, 420–433.
- Kersting, M. (2005). *Akzeptanzfragebogen Akzept!-P*. [Acceptance questionnaire Akzept!-P]. <https://kersting-internet.de/testentwicklungen/akzept-fragebogen/>
- Kersting, M. (2015). *Akzeptanzfragebogen Akzept!-P*. Available at: <http://kersting-internet.de/testentwicklungen/akzept-fragebogen/>
- Landes, M., and Laufer, K. (2013). "Feedbackprozesse—Psychologische Aspekte und effektive Gestaltung" in *Psychologie der Wirtschaft*. eds. M. Landes and E. Steiner (Wiesbaden: Springer VS), 681–703.
- Laevers, F. (Hrsg.), Depondt, L., Nijsmans, I., and Stroobants, I. (1993). *Die Leuener Engagiertheits-Skala für Kinder LES-K: [German version of the Leuven Involvement Scale for Young Children] (2. revised German version; K. Schlömer, translation)*. Berufskolleg Erkelenz, Fachschule für Sozialpädagogik.
- Leyendecker, B., Agache, A., and Madsen, S. (2014). Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit (NUBBEK)—Design, Methodenüberblick, Datenzugang und das Potenzial zu Mehrebenenanalysen. *Zeitschrift für Familienforschung* 26, 244–258. doi: 10.3224/zff.v26i2.16528
- Naumann, S., Bertram, H., Kuschel, A., Heinrichs, N., Hahlweg, K., and Döpfner, M. (2010). Der Erziehungfragebogen (EFB). *Diagnostica* 56, 144–157. doi: 10.1026/0012-1924/a000018
- NICHD ECCRN (2002). Child-care structure → process → outcome: direct and indirect effects of child-care quality on young Children's development. *Psychol. Sci.* 13, 199–206. doi: 10.1111/1467-9280.00438
- Pianta, R. C., Barnett, W. S., Burchinal, M., and Thornburg, K. R. (2009). The effects of preschool education: what we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychol. Sci. Public Interest* 10, 49–88. doi: 10.1177/1529100610381908
- Porter, T., Paulsell, D., Del Grosso, P., Avellar, S., Hass, R., and Vuong, L. (2010). *A Review of the Literature on Home-based Child Care: Implications for Future Directions*. Princeton, NJ: Mathematica Policy Research.
- Rindermann, H., and Baumeister, A. E. E. (2012). Unterschiede zwischen Montessori- und Regelkindergärten in der Kindergartenqualität und ihre Effekte auf die kindliche Entwicklung. *Psychol. Erzieh. Unterr.* 59, 217–226. doi: 10.2378/peu2012.art17d
- Schönbrodt, F. D., and Perugini, M. (2013). At what sample size do correlations stabilize? *J. Res. Pers.* 47, 609–612. doi: 10.1016/j.jrp.2013.05.009
- Sedlmeier, P., and Renkewitz, F. (2008). *Forschungsmethoden und Statistik in der Psychologie*. München: Pearson Studium.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., Taggart, B., and Elliot, K. (2004). "The effective provision of pre-school education project – Zu den Auswirkungen vorschulischer Einrichtungen in England" in *Anschlussfähige Bildungsprozesse im Elementar- und Primarbereich*. eds. G. Faust, M. Götz, H. Hacker and H.-G. Roßbach (Bad Heilbrunn, Obb: Klinkhardt), 154–167.
- Tietze, W., Becker-Stoll, F., Bensel, J., Eckhardt, A. G., Haug-Schnabel, G., and Kalicki, B. et al. (Eds.) (2012). "NUBBEK: Nationale Untersuchung zur Bildung in Betreuung und Erziehung in der frühen Kindheit. Fragestellungen und Ergebnisse im Überblick (Weimar: verlag das netz)

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/feduc.2023.1147954/full#supplementary-material>

- Tietze, W., Viernickel, S., Dittrich, I., Grenner, K., Hanisch, A., Lasson, A., et al. (2017). *Pädagogische Qualität entwickeln: Praktische Anleitung und Methodenbausteine für die Arbeit mit dem Nationalen Kriterienkatalog*. Weimar: verlag das netz.
- Vermeer, H. J., van IJzendoorn, M. H., Cárcamo, R. A., and Harrison, L. J. (2016). Quality of child care using the environment rating scales: A meta-analysis of international studies. *Int. J. Early Childhood* 48, 33–60. doi: 10.1007/s13158-015-0154-9
- Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *J. Am. Stat. Assoc.* 60, 63–69. doi: 10.1080/01621459.1965.10480775
- Watanabe, S. E., Phillips, D. A., Morrissey, T. W., McCartney, K., and Bub, K. (2011). Double jeopardy: poorer social-emotional outcomes for children in the NICHD SECCYD experiencing home and child-care environments that confer risk. *Child Dev.* 82, 48–65. doi: 10.1111/j.1467-8624.2010.01540.x
- Zhou, H., and Fishbach, A. (2016). The pitfall of experimenting on the web: how unattended selective attrition leads to surprising (yet false) research conclusions. *J. Pers. Soc. Psychol.* 111, 493–504. doi: 10.1037/pspa0000056
- Zimmerhofer, A. (2008). Studienberatung im deutschen Hochschulsystem auf der Basis psychologischer Tests: Studienfachprofile, Vorhersagevalidität und Akzeptanz. dissertation, Rheinisch-Westfälische Technische Hochschule Aachen. Aachen: Publikationsserver der RWTH Aachen University.